FOREWORD

The research described in the report was performed by the Laboratory of Aviation Psychology, The Ohio State University, Columbus, Ohio, under Air Force Contract No. AF 33(616)-7269 during the period 1 May 1960 to 31 August 1962. This is one of several laboratory research activities being carried out on the subject contract on "Techniques for Promoting the Long-Term Retention of Learned Skills." Dr. George E. Briggs is the Principal Investigator.

This contractual work was performed under Project 1710, "Training, Personnel, and Psychological Stress Aspects of Bioastronautics," Task 171003, "Human Factors in the Design of Systems for Operator Training and Evaluation," with Dr. T. E. Cotterman as Task Scientist. The contract was initiated and monitored until June 1961 by Mr. Frederick H. Kresse of the Operator Training Section, Training Research Branch. Mr. Kresse provided valuable assistance in the development of the experimental task and encouragement in the data acquisition phase of the work. Dr. Cotterman provided most helpful comments on an initial draft of this report.

ABSTRACT

Three variables (amount of training, task organization, and length of the retention interval), each at two levels, were studied to evaluate their influence upon the long-term retention of skill. Subjects were assigned to each of the eight experimental conditions, and were given practice on the criterion task for 2 or 3 weeks. The criterion task was composed of two subtasks, a procedural task involving the learning of discrete responses to discrete stimuli and a tracking task which involved continuous control of a three-dimensional compensatory display. All subjects returned for a retention test either 1 or 4 weeks subsequent to the end of training. Amount of training had a significant influence upon the degree of measured retention loss, as did task organization under conditions of lesser training. This was found for both tracking performance and for the number of omissive errors committed in performing the procedural task. Amount of absolute retention was generally related to (a) amount of training, (b) task organization, and (c) the length of the retention interval.

PUBLICATION REVIEW

This technical documentary report has been reviewed and is approved.

Walter F. Grether

WALTER F. GRETHER
Technical Director
Behavioral Sciences Laboratory

TABLE OF CONTENTS

	Page
INTRODUCTION	1
METHOD	2
Apparatus	2
Experimental Design	5
Subjects and Procedure	5
RESULTS	7
Tracking Performance	7
Training	7
Retention	7
Procedural Task Performance	13
Omissive Errors	14
Commissive Errors	14
Response Time	15
DISCUSSION	15
SUMMARY	18
REFERENCES	19
APPENDIX	20

LIST OF ILLUSTRATIONS

Figure		Page
1	Subject Panel for Procedural Task	2
2	Block Schematic of Tracking System	4
3	Illustration of Subject Performing on Criterion Task	6
4	Tracking Performance during Training and Retention Testing of the Four Groups Receiving Two Weeks of Training	8
5	Tracking Performance during Training and Retention Testing of the Four Groups Receiving Three Weeks of Training	9
6	Performance at End of Training as a Function of Task Organization and Amount of Training	9
7	Performance during Retest as a Function of Task Organization and Amount of Training	12
8	Retention Performance as Measured by Average Omissive Errors on Procedural Task as a Function of Task Organization and Amount of Training	15

LIST OF TABLES

Table		Page
1	Experimental Conditions for the Various Retention Groups	4
2	Analysis of Variance of Tracking Performance on Last Block of Training Trials	8
3	Analyses of Variance of Difference Scores (Training-Retest) for All Performance Measures	10
4	Analyses of Variance of Absolute Retention Scores for All Performance Measures	11
5	Mean Performance per Subject for Each of the Eight Experimental Groups during the Last Block of Training Trials and the Retention Test for Each of the Three Procedural Task Metrics	13
6	Analyses of Variance of Performance during Last Block of Training Trials	14
7	Schedule of Stimulus Conditions for the Nine Light Pairs (Coded 1 through 9 from Top to Bottom of Panel) Used During Training	20

INTRODUCTION

During the past seventy-five years, research on retention of verbal information and of perceptual-motor skill has been concerned with relatively short retention intervals. Only recently has there been an emphasis on retention intervals longer than 24 to 48 hours. Thus, most of the available data and conclusions have been restricted to relatively short retention intervals. The advent of space explorations of several days to several months duration and the attendant requirements for extremely skilled performance necessitate research on retention in which much longer retention intervals are used. This is a report of the first in a series of studies which utilizes relatively long retention intervals.

Previous retention research is limited also in terms of generalization to space operations on two additional points: most of that research has utilized verbal rather than motor skill tasks, and those motor tasks employed in most instances have been very simple and of questionable fidelity to vehicular control tasks. Two notable exceptions to this criticism are the studies by Ammons et al. (ref. 1) and Menglekoch, Adams, and Gainer (ref. 8).

Despite the possible limitations of certain of the previous data and of the available principles of retention, it is important that a current research program build upon past work, and the present study was designed to explore variables shown in the past to influence retention in a systematic and important manner. Three such variables were utilized: amount of original training, length of retention interval, and task organization. The first two variables are self-explanatory; however, task organization requires some explication.

The concept of task organization is based on Koffka's (ref. 6) statements in developing a gestalt theory of memory. Essentially, this concept holds that memory traces of unstructured material (such as a serial list of unrelated words) will fade more quickly than will memory of structured or better organized material (such as sets of meaningful phrases). An obvious weakness with the concept is that Koffka provides little information on how one determines the degree of organization present in a set of material or in a task. It is this lack of clarity on operations to specify organization that has led to the ambiguity of results from experiments designed specifically to test the organization hypothesis. Thus, Siegel (ref. 12) reported data supporting the organization concept, while Saul and Osgood (ref. 11) failed to support it. In the earlier research the operation employed to provide an organized task was to insert a "unique" or "different" pair of items in an otherwise homogeneous list of verbal items, while a less well-organized task involved the same critical pair of items in a list of heterogeneous items. It can be argued that these studies have confounded task organization with item uniqueness, but since our purpose is not to improve on verbal learning methodology, we turn to possible dimensions by which one can manipulate task organization in tasks more representative of aerospace vehicle operation.

A procedural task was developed which represents an abstraction of similar tasks in aircraft and manned aerospace (Mercury Project) vehicle operations. Essentially, the subject (S) is provided with a panel of lights which indicate the occurrence of a sequence of nine critical events. If the events occur in a normal (automatic) fashion, S merely acknowledges the occurrence of each event by pressing an "OK" button; if, however, an "emergency" occurs, S must take over from the automatic mode and manually correct the situation by a particular sequence of button presses.

Task organization was manipulated by a spatial sequence in which the critical events occurred: for the highly organized task a simple serial (top to bottom) sequence of light occurrences was employed; for the low task-organization condition the spatial sequence was quite irregular. The purpose of the study, then, was to assess the task-organization concept as a retention variable and to determine if that variable interacts with other retention variables of known importance (amount of original training and length of retention interval).

METHOD

Apparatus: The S was required to perform two independent tasks: the procedural task (mentioned above) and a three-dimensional, compensatory tracking task. Both subtasks were developed to simulate in a general fashion the task requirements of the Mercury capsule. The procedural task was designed to simulate the essential characteristics of the event monitoring required of the operator, in which he must observe the outcome of a given number of specific events during the course of the mission. The action taken (response) to any event depends upon (a) what the outcome of that event was (successful or unsuccessful), and (b) the

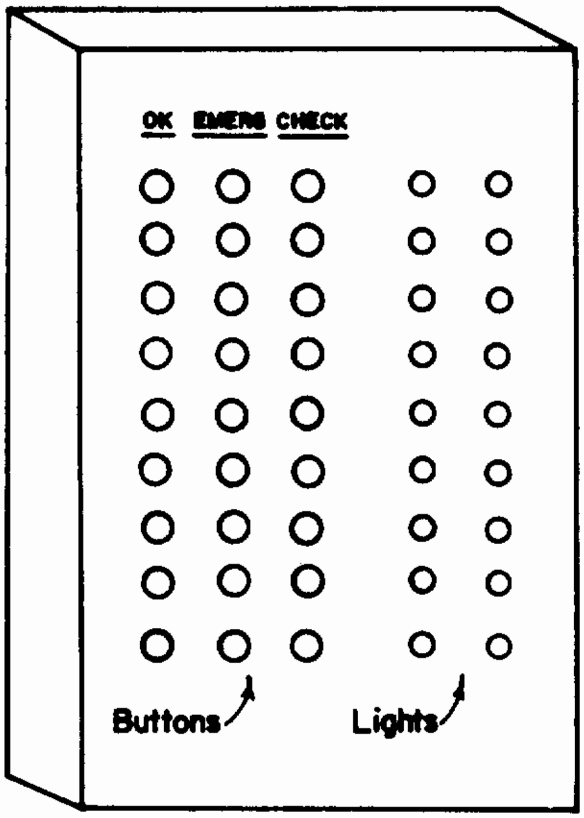

Figure 1. Subject Panel for Procedural Task.

time at which the event occurred (early, on time, or late). The task developed consisted of a panel of nine pairs of stimulus lights (one amber and one red) with 1-inch vertical separation between pairs (see figure 1). The left light of each pair was the amber light. To the left of each light pair are three response buttons labeled "Emergency," "OK," and "Check." If an amber light occurs, S presses the OK button; if a red light appears, he must respond by pressing the corresponding emergency and OK buttons in that sequence; and if no light at all occurs at the specified time, S must respond with the sequence of check-emergency-OK. A correct response or response sequence results in the light being locked into the amber or OK condition. Any failure to produce the appropriate response sequence results in the red light being locked in.

The procedural task panel was situated in front of S, approximately 2 feet from his left shoulder and 30° to the left of center of his frontal vision. The panel was rotated in such a manner that the plane area of the panel was at a maximum. Using a 1-second stepping switch control, it was possible to program (a) the spatial order of the nine stimulus events, (b) the duration of each stimulus event, and (c) the time between the onset of each stimulus event. In addition, E was able to program the condition for each stimulus event: (a) normal (amber light), (b) emergency (red light), or (c) possible emergency (no light at the proper time).

The procedural task organization was set in terms of the spatial contingency relationships of the stimulus event pairs. Two sequences of stimulus events were utilized: for the highly organized condition the spatial sequence was 1, 2, 3, 4, 5, 6, 7, 8, 9, where 1 refers to the top pair of lights, 2 to the next pair, and so on; for the low-organization condition the spatial sequence was 1, 5, 2, 9, 8, 3, 6, 7, 4. Since the two sequences represent sampling of light positions without replacement, an appropriate index is the information metric $\hat{H}$ calculated on the basis of relationships between changes in light positions rather than on the basis of light position itself. For the high-organization task $\hat{H} = 0.000$ bits, while for the low-organization task $\hat{H} = 2.808$ bits. Thus, the high-organization task condition provided more redundancy in terms of the progression of stimulus positions, and it follows that we have equated task organization with redundancy.

The duration of each stimulus event was constant at 4 seconds. The S was required to activate the appropriate response button or buttons within this time or the red light automatically locked in. There was a constant interval of 6 seconds between stimulus event onsets (4 seconds for a stimulus event plus 2 seconds delay before the next stimulus event).

The tracking task was three dimensional, providing simulation of the three attitude control dimensions of a vehicle in free flight (roll, pitch, and yaw): rate control dynamics were present in all three dimensions (see figure 2). The display panel was 19 × 10.5 inches and was situated directly in front of S to the right of the procedural task display. Three pairs of center-null-position meters, 4-5/8 inches wide × 4-1/2 inches high, were mounted in two rows of three each on the panel, with the upper dial of each pair providing S with attitude error and the lower dial providing rate error. The dials were labeled, from left to right: roll, pitch, and yaw. The input signal was a simple sine wave of 0.01 cycle per second as generated by a Hewlett-Packard Model 202A signal generator, and this signal was tracked in all three dimensions simultaneously. The S utilized a three-dimensional control stick for his tracking responses.

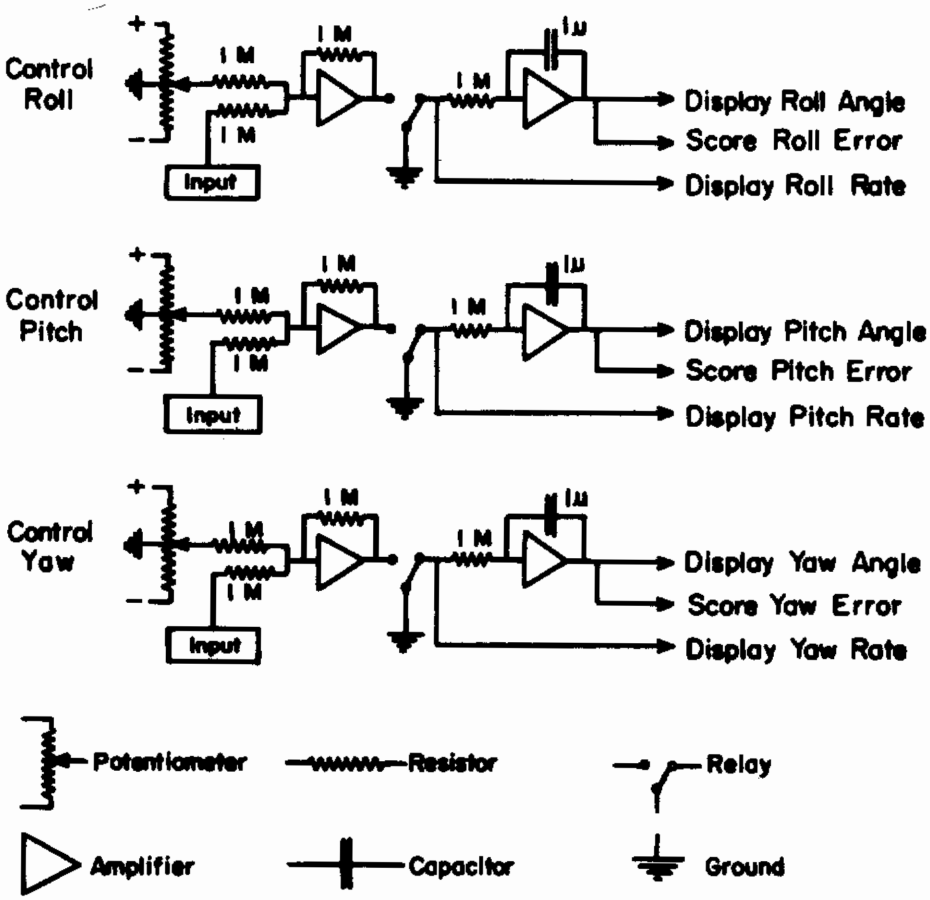

Figure 2. Block Schematic of Tracking System.

TABLE 1

EXPERIMENTAL CONDITIONS FOR THE VARIOUS RETENTION GROUPS

Group	Amount of Training (Weeks)	Task Organization	Retention Interval (Weeks)
1	2	High	1
2	2	High	4
3	2	Low	1
4	2	Low	4
5	3	High	1
6	3	High	4
7	3	Low	1
8	3	Low	4

Left-to-right stick movement controlled roll; front-to-back movement controlled pitch; and rotation (twisting) around the axis of the stick controlled yaw. The control stick was mounted to the floor approximately 1 foot in front of S. Height of the stick was 28 inches, and all control-display relationships were compatible, i.e., they conformed to population stereotypes. The displacement in degrees of the control for each of the control dimensions was as follows: roll, $\pm 60^{\circ}$; pitch, $\pm 60^{\circ}$; and yaw, $\pm 90^{\circ}$. The control-display gain was 0.025 inch/second of display pointer movement for each degree of arc of the control stick. The spring loading of the control was approximately 1 ounce per degree of arc.

Experimental Design: Table 1 shows the eight experimental groups and the associated experimental conditions. The three independent variables, each at two levels, were (a) training time (2 or 3 weeks), (b) retention interval (1 or 4 weeks), and (c) procedural task organization (high and low). The Ss were assigned to the various groups on the basis of their tracking performance on the sixth day of training in such a manner as to minimize random group differences. The mean of their combined tracking error was used as a matching criterion, and any S with a mean exceeding 40 volts was dropped from the experiment. There were 16 Ss per group by the end of the study.

Subjects and Procedure: A total of 128 undergraduate males served in the experiment. All were volunteers who received $1.00 per half-hour experimental session. All Ss were carefully instructed in the operation of both the procedural and the tracking task. Following the instructions in the first session, S received four 70-second trials on each of the two tasks separately, i.e., a part-task training schedule was employed with alternate tasks every trial. Part training continued through the fifth session with 12 trials per day (6 trials per task). Whole-task training began with session 6 and was continued through the end of the training period (session 10 or session 15, depending upon the group). Tracking performance was scored over the last 60 seconds of each trial and the onset of scoring coincided with the first stimulus event on the procedural task, i.e., S tracked alone for the first 10 seconds of each 70-second trial. In this way scored performance on the tracking task coincided with performance on the procedural task. The Ss were run in pairs on duplicate sets of equipment. Figure 3 illustrates the criterion task.

Integrated absolute error served as the performance metric for the tracking task and this score was taken for each of the three task dimensions separately. Performance on the procedural task was evaluated in terms of three metrics: total response time for the nine stimulus events of each trial, number of commissive errors (button presses in excess of the required number and/or button presses which were incorrect for the particular stimulus condition), and number of omissive errors (number of times S failed to respond at the appropriate time).

Although the spatial sequence and the time intervals between stimulus events were constant for S, the actual stimulus condition (amber light, red light, or no light) was programmed by E to change from trial to trial and from day to day. The schedule was as follows:

1. Sessions 1 and 2: No off-light conditions were used; S experienced four red- and five amber-light conditions on each trial with the order of conditions randomized.

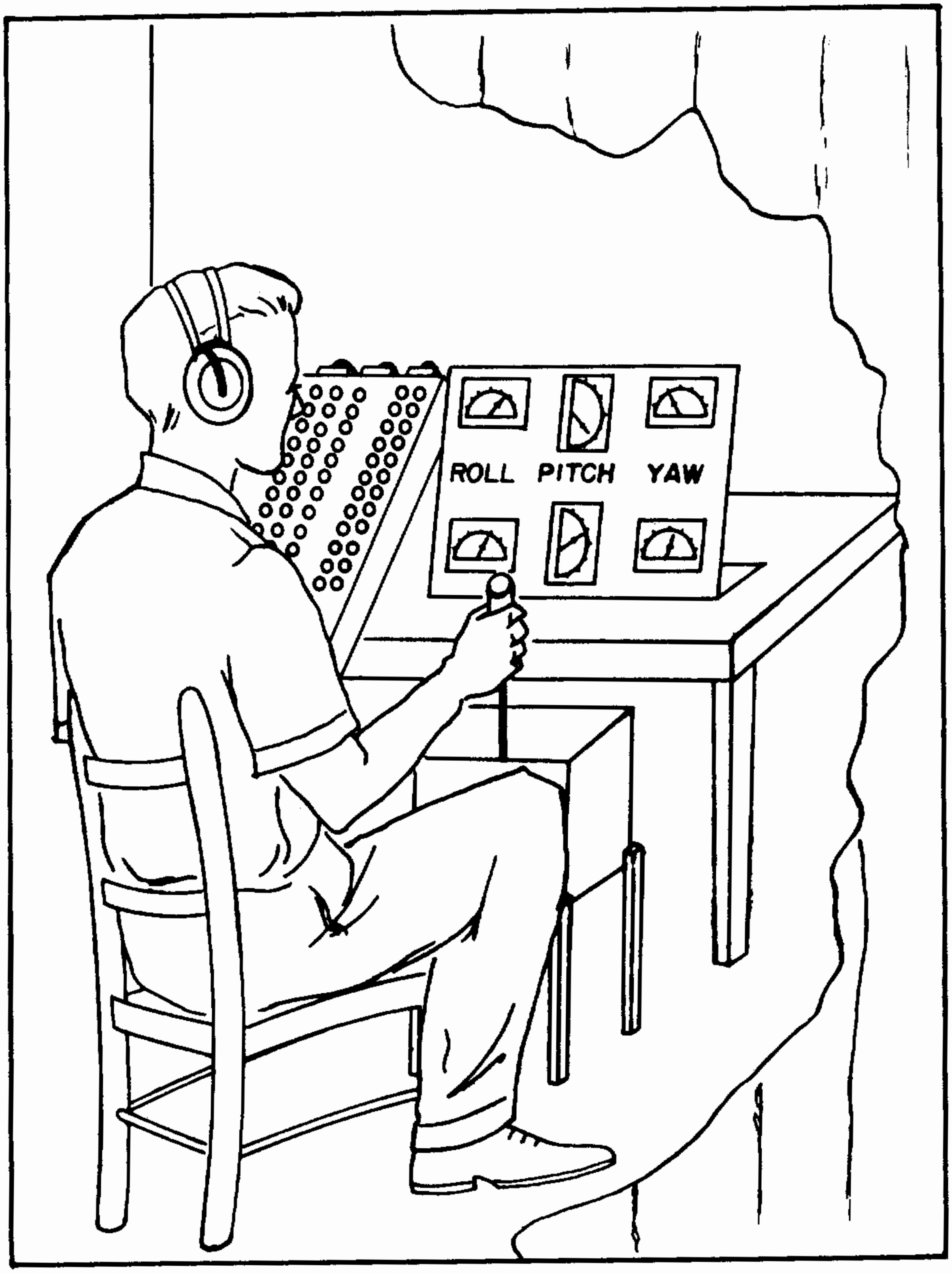

Figure 3. Illustration of Subject Performing on Criterion Task.

2. Session 3: One off-light condition occurred (in the No. 2 light position) during trials 2 and 3 and two off-light conditions occurred (in the No. 2 and No. 4 light positions) during trials 5 and 6.

3. Session 4: Three off-light conditions occurred (in light positions 2, 4, and 6) on trials 2 and 3, and four off-lights occurred (in positions 2, 4, 6, and 8) on trials 5 and 6.

4. Session 5 on: Four off-light conditions occurred (in positions 2, 4, 6, and 8) on trials 2 through 6 during the fifth session and on every trial for the remaining training sessions and for the retention test session.

Table 7 of the Appendix provides a detailed summary of these conditions for the procedural task.

RESULTS

Tracking Performance

On each trial there were three tracking scores generated by S—integrated absolute error for each of the three dimensions of the tracking task. These three scores were summed to provide a unitary score of tracking proficiency and group averages were obtained over the 12 trials of each session starting with session 6. These data are plotted in figure 4 (the 2-week training groups) and in figure 5 (the 3-week training groups). The data for the first five sessions are not included in these figures as (a) S was working on a part-training schedule, and (b) he was experiencing an increasing number of off-light (possible emergency) conditions on the procedural task which did not reach the "operational" level until session 6 (see table 7).

Training: In figures 4 and 5 one can observe the marked influence of task organization on performance during training: during session 10 the average of groups 1 and 2 is superior to that of groups 3 and 4, and during session 15 the average of groups 5 and 6 is superior to that of groups 7 and 8. These averages are plotted in figure 6 to illustrate this effect more clearly. It is apparent also that amount of training influenced performance: as expected, the average of groups 5-8 is superior to that of groups 1-4 at the end of their respective training periods (sessions 15 and 10, respectively).

These observations receive statistical support from an analysis of variance which was performed on the data of all groups from the last block of training trials. A summary of this analysis appears in table 2 where both amount of training and task organization are statistically significant at $P < .05$. The lack of statistical significance for the interaction of these two variables indicates that the slopes of the functions shown in figure 6 are essentially the same. Four t tests were conducted between the group pairs 1 vs. 2, 3 vs. 4, 5 vs. 6, and 7 vs. 8 on day 6 performance to check on matching procedure. None of the tests was significant.

Retention: Since the groups differed significantly by the end of training, analyses were conducted using (a) differences in retention performance which were corrected for the preretention differences, and (b) the absolute retention values

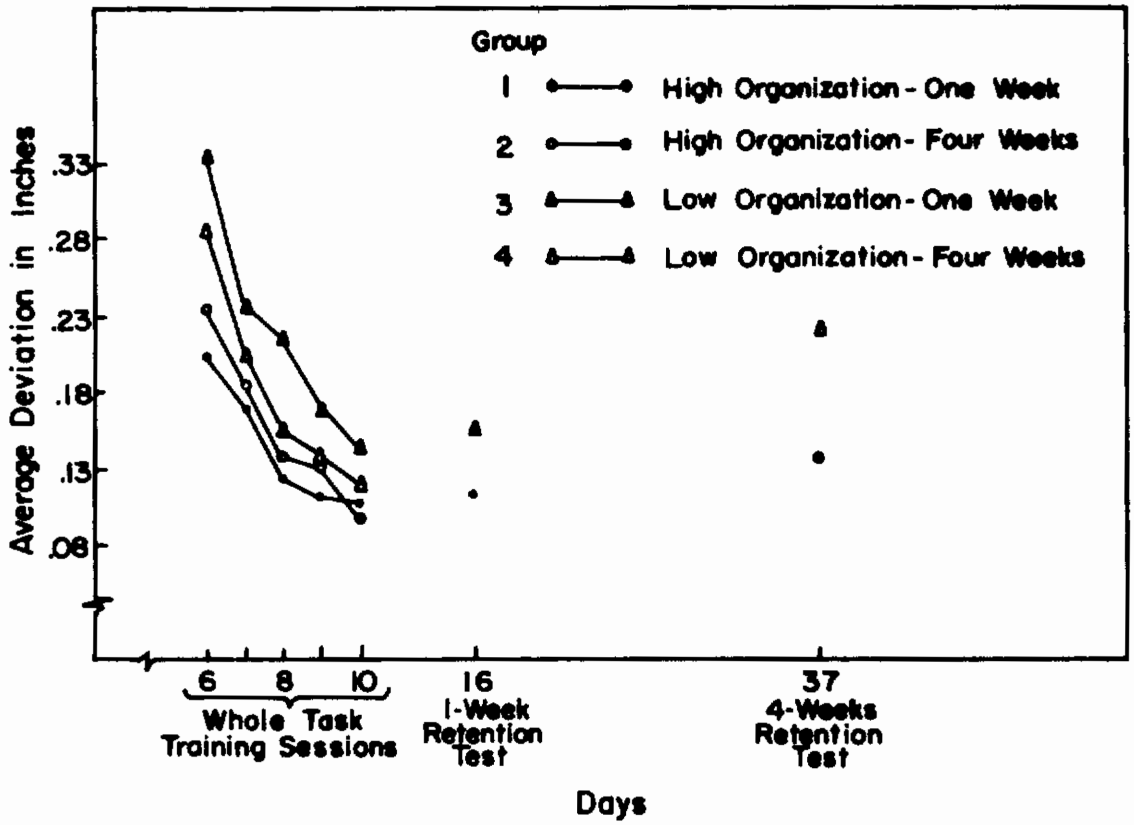

Figure 4. Tracking Performance during Training and Retention Testing of the Four Groups Receiving Two Weeks of Training. The first five training sessions consisted of progressive-part training and are therefore not plotted since performance during this period was not comparable to the data shown above.

TABLE 2

ANALYSIS OF VARIANCE OF TRACKING PERFORMANCE
ON LAST BLOCK OF TRAINING TRIALS

Source	df	MS	F
Amount of Training (T)	1	1550.55	5.92*
Task Organization (O)	1	1374.85	5.25*
T × O	1	0.56	0.00
Error	124	261.93	
Total	127		

* $\underline{P} < .05$

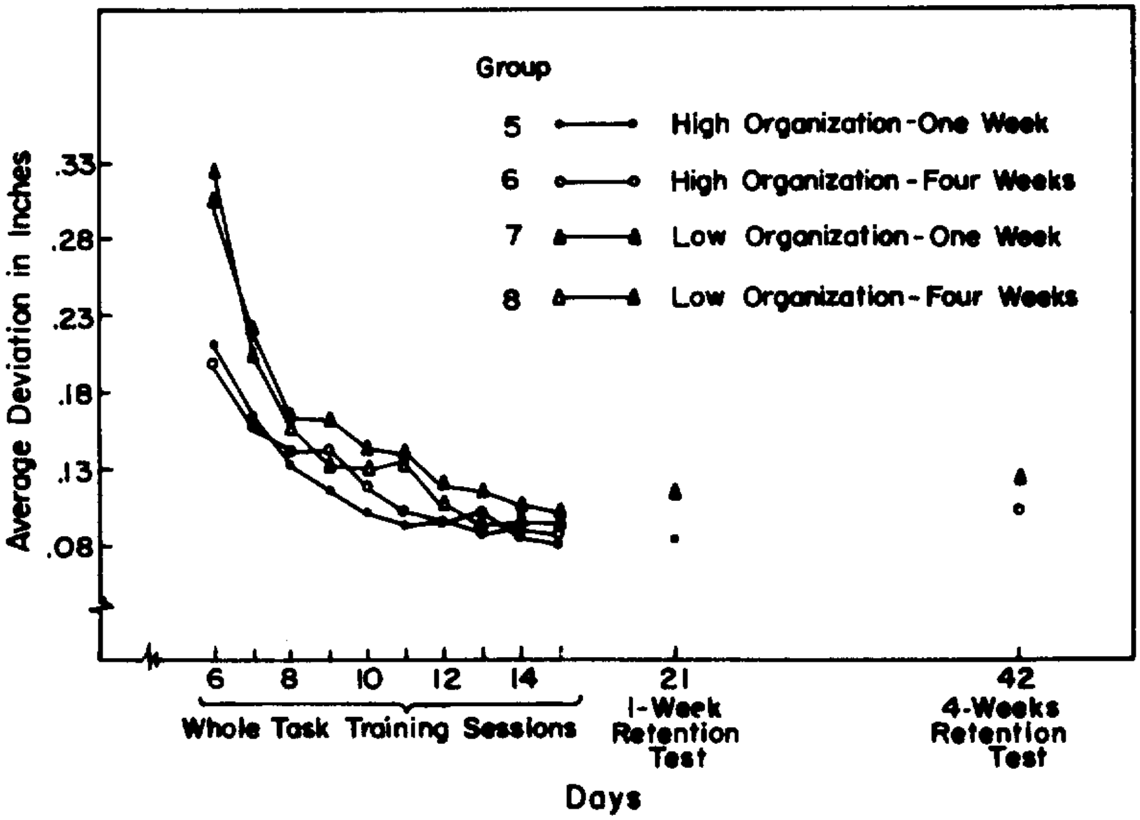

Figure 5. Tracking Performance during Training and Retention Testing of the Four Groups Receiving Three Weeks of Training. The first five training sessions consisted of progressive-part training and are therefore not plotted since performance during this period was not comparable to data shown above.

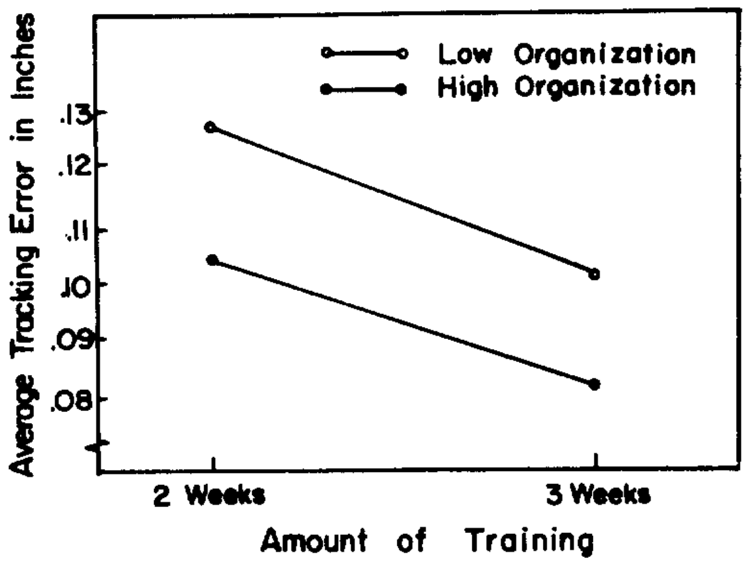

Figure 6. Performance at End of Training as a Function of Task Organization and Amount of Training.

TABLE 3

ANALYSES OF VARIANCE OF DIFFERENCE SCORES (TRAINING-RETEST) FOR ALL PERFORMANCE MEASURES

Source	df	Tracking		Omissive Errors		Commissive Errors		Response Time	
		MS	F	MS	F	MS	F	MS	F
Amount of Training (T)	1	12399.21	22.57**	2.0125	11.21**	2.1321	1.98	38.8967	3.28
Task Organization (O)	1	1660.34	3.02	0.2993	1.67	0.6786	0.63	0.0132	0.00
Retention Interval (I)	1	101.55	0.18	0.0037	0.02	13.1072	12.18**	0.6877	0.06
T × O	1	3012.80	5.48*	1.2940	7.21**	1.5269	1.42	14.0183	1.18
T × I	1	1664.63	3.03	0.0000	0.00	0.2794	0.26	1.5706	0.13
O × I	1	188.16	0.34	0.0001	0.00	0.5279	0.49	27.6208	2.33
T × O × I	1	1797.02	3.27	0.0469	0.26	0.7503	0.70	1.8865	0.16
Residual Error	120	549.41		0.1795		1.0764		11.8582	

* $p < .05$

** $p < .01$

TABLE 4

ANALYSES OF VARIANCE OF ABSOLUTE RETENTION SCORES
FOR ALL PERFORMANCE MEASURES

Source	df	Tracking		Omissive Errors		Commissive Errors		Response Time	
		MS	F	MS	F	MS	F	MS	F
Amount of Training (T)	1	9080.47	18.45**	1.9404	16.10**	5.2407	4.09*	36.7440	4.22*
Task Organization (O)	1	6419.86	13.05**	1.6608	13.78**	0.0069	—	11.2694	1.30
Retention Interval (I)	1	2263.62	4.62*	0.1891	1.57	8.8200	6.89**	1.6562	—
T × O	1	1546.37	3.16	1.1101	9.21**	0.9940	—	10.3285	1.19
T × I	1	380.53	—	0.0657	—	0.0443	—	0.0935	—
O × I	1	168.13	—	0.0741	—	0.8224	—	27.2137	3.13
T × O × I	1	421.59	—	0.0587	—	0.4705	—	19.5625	2.25
Residual Error	120	492.10		0.1205		1.2801		8.7001	

* $p < .05$

** $p < .01$

themselves. The correction was attained by using difference scores (last training session performance—retention test performance) for each S rather than the absolute retention scores themselves. The analyses of variance performed on these difference scores are summarized in table 3, and those conducted on the absolute retention scores are summarized in table 4.

From table 3 it is apparent that amount of training produced a significant effect on retention performance (as it did on training performance levels), but unlike the training results, task organization did not provide a significant influence on retention test performance. However, the amount of training by task organization interaction is statistically significant at $P < .05$. Figure 7 provides the data defining this interaction. From figure 7 it is apparent that while task organization did not differentially affect performance during retention following 3 weeks of training, there was a significantly greater level of retention of the high-organization task than for the low-organization task after only 2 weeks of training. It would appear, therefore, that task organization may influence retention but only if relatively little training has been received, and given enough training, retention of a less-organized task can be as good as that for a more highly organized task.

Stated another way, it is apparent from figure 7 that relatively little "additional" training is required to improve the retention of a less-organized task to that level obtained with a more highly organized task. Thus, by providing only 50% more training (3 weeks vs. 2 weeks), the retention of the tracking task performed under a low level of procedural task organization approaches closely the tracking level attained under a high level of procedural task organization. Also, with only 2 weeks of training, task organization clearly differentiated performance on the tracking task. This result, the finding of a

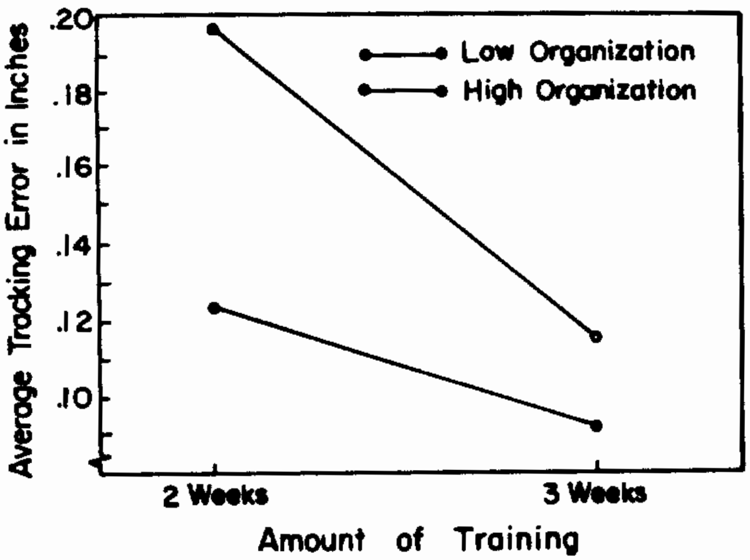

Figure 7. Performance during Retest as a Function of Task Organization and Amount of Training.

statistically significant interaction between task organization and amount of training, is a most important finding from a practical as well as theoretical point of view, and the implications of this are covered in a later section of this report.

Examination of the results obtained from the absolute retention values (table 4) for tracking were somewhat different from those obtained with the difference scores. With these data all three main effects (training, task organization, and retention interval) were significant. These differences are all clearly apparent when one examines only the retention points shown in figures 4 and 5. As was found with the difference score analysis, the more extensive the training, the greater the retention. In addition, the more organized task and the shorter retention interval also resulted in significantly greater retention. The interaction between task organization and training time was not significant in this analysis.

Procedural Task Performance

During each trial S's performance on the procedural task was scored in three ways: omissive errors—a failure to respond correctly and thus receive a red panel light at the end of a stimulus event; commissive errors—responding incorrectly and thereby generating extra responses; and response time—the time between onset of a stimulus event and the completion of the correct response or sequence of responses.

These measures of performance were not quite as sensitive to experimental conditions as were the tracking metrics, especially the analyses of difference scores. It is felt that any sensitivity for the omissive and commissive error

TABLE 5

MEAN PERFORMANCE PER SUBJECT FOR EACH OF THE EIGHT EXPERIMENTAL GROUPS DURING THE LAST BLOCK OF TRAINING TRIALS AND THE RETENTION TEST FOR EACH OF THE THREE PROCEDURAL TASK METRICS

Group	Omissive Errors		Commissive Errors		Response Time (in Seconds)	
	Training	Retention	Training	Retention	Training	Retention
1	0.18	0.14	1.02	0.92	7.61	6.91
2	0.25	0.17	0.71	1.36	8.24	8.89
3	0.27	0.46	0.79	1.07	8.74	9.77
4	0.30	0.68	0.60	1.59	7.87	8.35
5	0.08	0.08	0.67	0.77	7.98	7.24
6	0.05	0.11	0.69	1.05	6.74	7.55
7	0.08	0.12	0.58	0.34	7.04	7.41
8	0.13	0.16	0.60	1.17	8.41	7.44

scores is a function of their relative infrequency, especially after training had progressed into the second week.

Table 5 provides a summary of these data: The columns labeled "training" represent the average performance per S of the last block of training trials, while the columns labeled "retention" contain averages per S during the retention test period. Table 6 provides a summary of three analyses of variance which were applied to the training performance data, and table 3 gives a summary of three analyses of variance applied to the retention test data in the form of difference scores (see the earlier discussion of difference scores in regard to table 3). Table 4 presents the analyses performed on the absolute retention scores.

Omissive Errors: The analyses of these data provide much the same results as were obtained with the tracking error scores, with one exception: during training, task organization did not differentially affect the occurrence of omissive errors as it did the tracking performance data (table 2 vs. table 5). Thus, training performance, in terms of omissive errors, simply shows the influence of amount of training, and from table 5 it can be seen that groups 1-4 generated an average of 0.250 omissive errors during the last training session, while groups 5-8 exhibited an average of only 0.085 omissive errors.

In terms of retention performance, the analysis of the omissive error difference score data provided the same pattern of results as obtained with the tracking difference scores (table 3): both amount of training and the training by task organization interaction are statistically significant. The omissive error data defining this interaction are summarized in figure 8.

Similar findings were obtained with the absolute measures of retention, except that task organization was also significant (see table 4).

Commissive Errors: The analysis of these data for the final block of training trials (see table 6) indicates no statistically significant influence by either experimental variable, nor was there a significant interaction of these

TABLE 6

ANALYSES OF VARIANCE OF PERFORMANCE DURING THE LAST BLOCK OF TRAINING TRIALS

Source	df	Omissive Errors		Commissive Errors		Response Time	
		MS	F	MS	F	MS	F
Amount of Training (T)	1	0.8845	5.81*	0.6874	0.79	10.5226	1.08
Task Organization (O)	1	0.0882	0.58	0.5486	0.63	4.4776	0.46
T × O	1	0.0038	0.02	0.0570	0.07	0.0015	0.00
Residual Error	124	0.1522		0.8685		9.7114	

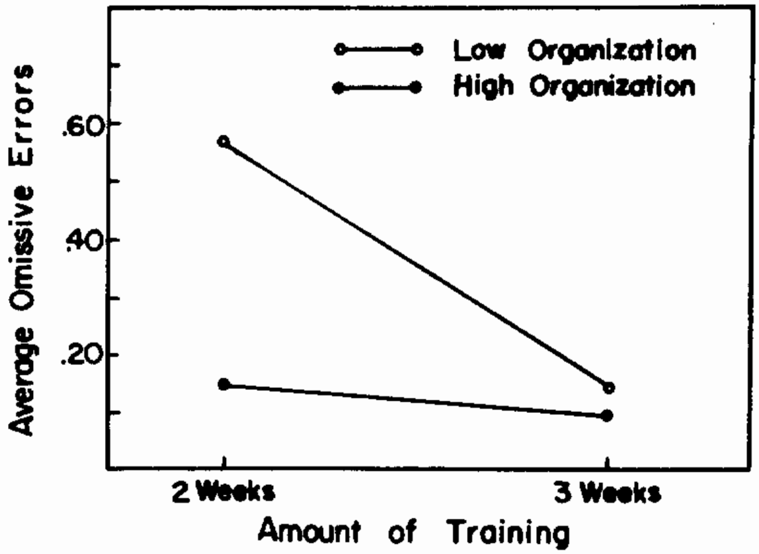

Figure 8. Retention Performance as Measured by Average Omissive Errors on Procedural Task as a Function of Task Organization and Amount of Training.

two variables. However, in the difference score analysis of the retention test performance data (table 3), retention interval emerged as a significant variable, and this effect may be seen in table 5 by noting that the average performance of groups 1, 3, 5, and 7 (1-week retention) involved fewer commissive errors upon retention test than did the performance of groups 2, 4, 6, and 8 (4 weeks of retention). These data are the only ones to show any differential effect on performance of retention interval (1 vs. 4 weeks). The absolute scores showed significant differences due to training and to the length of the retention interval.

Response Time: The analyses of the response times in training and test (tables 3, 4, and 6) resulted in only one statistically significant F ratio. This was rather surprising since response time has long been a "standard" dependent variable. However, an examination of the trends of the response time scores during the training sessions showed that whereas group averages systematically decreased with practice, there was substantial irregularity in the trends. Thus, these data exhibited considerable unreliability, and this may very well account for the lack of statistical significance from the analyses of these measures.

DISCUSSION

It should be noted that two methods of analysis were used on all the data for all measures of performance. These two methods of analysis actually examine two quite different questions. The difference score analyses (training performance—retest performance) provide an indication of amount of relative loss in skill as a function of the experimental conditions. Thus, controlling for the final level of training performance, they ask which groups showed the largest decrements in retention. The two most sensitive dependent variables for this question were tracking performance and omissive errors. It was not surprising to find tracking behavior sensitive to changes in the procedural task, as Bahrick and Shelley (ref. 2) previously had demonstrated the sensitivity of tracking when used as a supplemental performance criterion.

The pattern of significance was identical for these two measures, with amount of training and the training by task organization effects both accounting for a significant portion of the variance in retention performance. By far the largest single variance source in both cases was due to amount of training, indicating that practice is indeed an exceedingly potent, if not overpowering, variable. This is also demonstrated by the significant interaction, since only those groups having only 2 weeks of training showed any differences in retention as a function of task organization.

The existence of an interaction between these two variables illustrates a point made often in the past: the necessity for more multivariate studies in the area of skill retention. Both the amount of training variable and the organization variable previously have been suggested or shown to be important retention parameters (ref. 9). The interaction found in this study indicates, however, that task organization is only important under conditions of lesser degrees of training—a condition which is probably not too likely to exist in most practical learning situations.

The analyses of the absolute retention scores, on the other hand, examined a quite different question. They examine the absolute retention exhibited by the groups as a function of the experimental conditions. This latter question is a somewhat more practical question, and therefore perhaps somewhat more limited in generality to the particular task employed in the study.

The variable of amount of training was found to be a significant variable with all measures. Both the omissive error scores and tracking performance (mentioned earlier as being the most sensitive measures) indicated significant retention differences due to task organization. This variable was not found to be significant with the difference score analysis except under conditions of lesser training. Thus, task organization does influence the amount of absolute retention, but apparently does not play a role in determining the relative loss in retention.

This finding is interesting, since it suggests that the rate of loss of a skill is not a function of task organization, and that absolute retention differences are due to degree of original skill acquisition. Since more organized skills are usually more thoroughly acquired in a given amount of time, this would appear to be reducing task organization to a special case of amount of training.

One further aspect concerning the fact that both the tracking task and the procedural task exhibited the same pattern of behavioral effects should be stated. Although no attempt was made to equate the two subtasks, the similarity of results indicates that there is a certain amount of generality inherent in the findings which is pertinent to skill retention per se, regardless of whether the task is continuous or discrete in nature. Thus, the conclusion that the "type" of task (e.g., procedural or discrete) is not necessarily a pertinent variable in retention (ref. 9) seems supported.

One of the more surprising outcomes was the failure of length of the retention interval to affect retention significantly in the case of the difference scores (with the exception of tracking performance). This was not the case with the absolute scores, where both tracking and commissive errors demonstrated retention interval to be significant. These results are similar to those obtained by Krueger (ref. 7), who utilized retention periods up to 2 weeks in duration, and Jahnke (ref. 5), who studied retention periods of 1 week.

In general, those studies which have investigated amount of training have found results similar to those reported here. Thus, Bell (ref. 3) showed that groups having lesser training showed greater decrements in retention, as did Duncan and Underwood (ref. 4) and Ammons et al. (ref. 1). An interesting point regarding certain previous studies dealing with amount of training is that those studies, such as Krueger (ref. 7) and Rubin-Rabson (ref. 10), which have studied overlearning typically have found little or no facilitation of retention through such methods. The results of the present study, which indicate amount of practice to be extremely powerful as a retention variable, might suggest that 100% acquisition results in retention of such a high degree over fairly extended periods that little or no deterioration is likely.

Certain comments are relevant regarding the remaining two dependent variables, commissive errors and response time. These retention measures showed fewer differences between groups than either of the other measures. The major difference between these two measures and those previously discussed was in terms of the speed of feedback of performance. With omissive errors and tracking error the feedback was immediate, as both types of errors were evident on their respective displays. Response time scores and commissive errors were only obtained by S at the end of each day's session of training. This may have led to a tendency on the part of Ss to pay more attention to those aspects of performance about which immediate knowledge of results was given.

SUMMARY

The study investigated three independent variables (amount of training, task organization, and length of the retention interval) and their effects upon the long-term retention of both a discrete and a procedural task. Two levels of each variable were studied: 2 and 3 weeks of training, high and low organization of the procedural task, and 1- and 4-week retention intervals. Sixteen subjects were assigned to each of the eight experimental groups, assignment to groups being based upon a matching criterion of tracking performance.

Amount of skill loss was found to be a function of amount of original training and task organization (under conditions of 2 weeks of training only). Amount of absolute retention, however, was found to be a function of all three independent variables. General conclusions of the study were that (a) amount of training is an extremely important retention variable; (b) under lesser training conditions, task organization becomes important; (c) task organization as a variable is probably a special case of degree of learning; and (d) there is a need for more multivariate retention studies.

REFERENCES

1. Ammons, R. B., R. G. Farr, E. Block, E. Neumann, M. Dey, R. Marion, and C. H. Ammons, "Long Term Retention of Perceptual-Motor Skills," Journal of Experimental Psychology, Vol 55, pp 318-328, 1958.

2. Bahrick, H. P., and Carolyn Shelley, "Time Sharing as an Index of Automatization," Journal of Experimental Psychology, Vol 56, pp 288-293, 1958.

3. Bell, H. M., "Rest Paulses in Motor Learning as Related to Snoddy's Hypothesis of Mental Growth," Psychological Monographs, Vol 54, Whole No. 1, 1942.

4. Duncan, C. P., and B. J. Underwood, "Retention of Transfer in Motor Learning after Twenty-Four Hours and after Fourteen Months," Journal of Experimental Psychology, Vol 46, pp 445-452, 1953.

5. Jahnke, J. C., "Retention in Motor Learning as a Function of Amount of Practice and Rest," Journal of Experimental Psychology, Vol 55, pp 270-273, 1958.

6. Koffka, K., Principles of Gestalt Psychology, Harcourt, Brace and Company, New York, N. Y., 1935.

7. Krueger, W. C. F., "Further Studies in Overlearning," Journal of Experimental Psychology, Vol 12, pp 502-514, 1929.

8. Mengelkoch, R. F., J. A. Adams, and C. A. Gainer, The Forgetting of Instrument Flying Skills as a Function of the Level of Initial Proficiency, Technical Report NAVTRADEVCEN 71-16-18, U. S. Naval Training Device Center, Port Washington, N. Y., 1960.

9. Naylor, J. C., and G. E. Briggs, Long-Term Retention of Learned Skills: A Review of the Literature, ASD Technical Report 61-390, Aeronautical Systems Division, Wright-Patterson Air Force Base, Ohio, August 1961.

10. Rubin-Rabson, Grace, "Studies in the Psychology of Memorizing Piano Music: VII. A Comparison of Three Degrees of Overlearning," Journal of Educational Psychology, Vol 32, pp 688-696, 1941.

11. Saul, E. V., and C. E. Osgood, "Perceptual Organization of Materials as a Factor Influencing Ease of Learning and Degree of Retention," Journal of Experimental Psychology, Vol 40, pp 372-379, 1950.

12. Siegel, P. S., "Structure Effects within a Memory Series," Journal of Experimental Psychology, Vol 33, pp 311-316, 1943.

APPENDIX

TABLE 7

SCHEDULE OF STIMULUS CONDITIONS FOR THE NINE LIGHT PAIRS
(CODED 1 THROUGH 9 FROM TOP TO BOTTOM OF PANEL)
USED DURING TRAINING

Session	Trial	Red Lights	Off Lights	Session	Trial	Red Lights	Off Lights
1	1-4	1-2-3-9	None	9	1-4 5-8 9-12	2-3-5-7 1-3-4-8 1-3-6-7	2-4-6-8 2-4-6-8 2-4-6-8
2	1-3 4-6	1-2-4-9 3-4-5-6	None None	10	1-4 5-8 9-12	3-4-6-8 2-3-6-9 1-3-6-7	2-4-6-8 2-4-6-8 2-4-6-8
3	1 2-3 4 5-6	6-7-8-9 6-7-8-9 1-2-5-8 1-2-5-8	None 2 None 2-4	11	1-4 5-8 9-12	2-3-4-9 4-6-7-8 2-5-6-8	2-4-6-8 2-4-6-8 2-4-6-8
4	1 2-3 4 5-6	1-3-6-7 1-3-6-7 3-4-6-7 3-4-6-7	None 2-4-6 None 2-4-6-8	12	1-4 5-8 9-12	3-5-6-8 5-7-8-9 3-4-7-8	2-4-6-8 2-4-6-8 2-4-6-8
5	1 2-3 4-6	1-4-5-9 1-4-5-9 4-5-8-9	None 2-4-6-8 2-4-6-8	13	1-4 5-8 9-12	1-3-5-8 2-3-5-6 2-5-8-9	2-4-6-8 2-4-6-8 2-4-6-8
6	1-4 5-8 9-12	2-3-4-9 4-6-7-8 2-5-6-8	2-4-6-8 2-4-6-8 2-4-6-8	14	1-4 5-8 9-12	2-3-5-7 1-3-4-8 1-3-6-7	2-4-6-8 2-4-6-8 2-4-6-8
7	1-4 5-8 9-12	3-5-6-8 5-7-8-9 3-4-7-8	2-4-6-8 2-4-6-8 2-4-6-8	15	1-4 5-8 9-12	3-4-6-8 2-3-6-9 2-3-4-5	2-4-6-8 2-4-6-8 2-4-6-8
8	1-4 5-8 9-12	1-3-5-8 2-3-5-6 2-5-8-9	2-4-6-8 2-4-6-8 2-4-6-8	Retest	1-4	1-2-3-9	2-4-6-8

Note:—Four lights in each series were always red and five were amber. The number of off-lights was gradually raised from zero (day 1) to the operational task condition of four per trial on day 4 (trials 5 and 6).