R E P O R T   R E S U M E S

ED 014 876   24   EM 004 009

REPETITION AND SPACED REVIEW IN THE LEARNING OF CONNECTED DISCOURSE.

BY- REYNOLDS, JAMES H.   GLASER, ROBERT
PITTSBURGH UNIV., PA., LEARNING RES. AND DEV. CTR.

REPORT NUMBER CRP-1343-1   PUB DATE   DEC 63
EDRS PRICE MF-$0.25 HC-$1.92   46P.

DESCRIPTORS- *TEACHING MACHINES, *PROGRAMED INSTRUCTION, *LEARNING PROCESSES, *VERBAL LEARNING, *RETENTION, TIME FACTORS (LEARNING), MIN-MAX I, MIN-MAX II

3 EXPERIMENTS WERE DESIGNED TO FILL THE GAP BETWEEN LAB-ESTABLISHED VERBAL LEARNING PRINCIPLES AND THEIR APPLICATION TO COMPLEX VERBAL LEARNING BY USING THE TEACHING MACHINE TO CONTROL PRESENTATION VARIABLES SUCH AS AMOUNT AND ORDER OF MATERIAL EXPOSED. EXPERIMENT I MEASURED THE EFFECTS OF REPETITION AND SPACED REVIEW ON RETENTION, EXPERIMENT II THE EFFECTS OF SPACED REVIEW UPON RETENTION WITH RECENCY (TIME INTERVAL BETWEEN LAST PRACTICE TRAIL AND RETENTION TESTS) CONTROLLED, AND EXPERIMENT III THE EFFECTS OF PRIOR TESTING AND AN EXTENDED FORGETTING INTERVAL UPON RETENTION. INTACT GRADE 8 CLASSES EQUATED ON INTELLIGENCE WERE CHOSEN FOR ALL EXPERIMENTS. NO CONTROLS WERE INTRODUCED TO ACCOUNT FOR SUBJECTS' LACK OF PREVIOUS EXPOSURE TO PROGRAMMED INSTRUCTION. A SERIES OF EXPERIMENTAL FRAMES WAS INCLUDED IN A LARGER PROGRAM SEQUENCE, YIELDING 5 TREATMENT CONDITIONS IN EXPERIMENT I AND 2 IN II AND III. GROUPS VARIED FROM 10-35. A PRE-TEST OF RECOGNITION OF MATERIAL WAS GIVEN, AND LATER REPEATED WITH DELAYED POST-TESTS OF AIDED AND UNAIDED RECALL. 3 MAIN RESULTS WERE--REPETITION WAS NOT A MAJOR VARIABLE INFLUENCING RETENTION OF COMPLEX MATERIAL, SPACING OF REVIEW SEQUENCES BETWEEN INTERPOLATED LEARNING MATERIALS FACILITATED RETENTION, AND RETENTION WAS MAINTAINED AT FULL STRENGTH FOR AT LEAST 6 WEEKS, DURING WHICH LARGE AMOUNTS OF INTERFERING MATERIAL WERE PRESENTED. (LH)

ED014876

REPETITION AND SPACED REVIEW IN THE LEARNING OF CONNECTED DISCOURSE

James H. Reynolds
Robert Glaser

Learning Research and Development Center
University of Pittsburgh

Project No. 1343
Cooperative Research Program
U. S. Office of Education

December, 1963

EM 004 009

Repetition and Spaced Review in the Learning of Connected Discourse[1]

James H. Reynolds[2] and Robert Glaser
University of Pittsburgh

Background and Problem

In research on verbal learning, investigations using essentially non-meaningful tasks such as paired associates and serial lists have far outnumbered those using more meaningful connected discourse material. A major reason for this imbalance is that presentation of disconnected serial or associative materials permits a degree of precision in the control of experimental conditions which ordinarily is not possible when connected discourse materials are used. Memory drums, card-presentation apparatus, film strips, and other presentation devices which control such learning variables as time, order, and amount of material exposed are difficult to adapt to the experimental manipulation of extensive reading passages or subject matter of an academic nature, making it necessary to use less controlled methods with this type of material. One consequence of this problem has been that discrepancies between the results of research on meaningful and non-meaningful materials are often attributed to control differences in the experimental procedures, with credence usually given to the non-meaningful findings. For this reason researchers in verbal learning have tended to develop and test hypotheses with disconnected material and have hesitated to apply the principles developed in the laboratory to outside-of-laboratory learning situations because of the

---

[1] The research reported is part of a project supported through the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education and Welfare.

[2] Now at the Institute for Human Learning, University of California, Berkeley, California.

control criticisms which almost inevitably arise. The resulting gap between research which confirms learning principles by using serial lists or paired associates and research which attempts to generalize these principles to meaningful connected-discourse learning contexts is a broad one.

The research reported in this paper was an attempt to help fill the gap between laboratory-established verbal learning principles and their application to complex verbal learning by using the teaching machine as a control apparatus. It was considered that the teaching machine, in a manner similar to a memory drum, would permit control over such presentation variables as amount and order of material exposed, and could also, if desired, regulate the time of exposure. Moreover, the increased control would allow the use of a rather large amount of complex academic subject matter as the experimental learning task, rather than limiting connected discourse to brief passages for rote memorization, as has been the case in most of the connected discourse research reported. The main criticism of investigation employing academic material has been that human teachers are required, bringing into the experimental situation variations in instructional stimuli, reinforcing and punishing contingencies, and other subjective influences. These problems are minimized when printed material such as a programmed learning sequence serves the functions of instructing, requiring learner responses and providing feedback in a standard and objective fashion, and under the control of the experimenter.

Two variables, which have received wide research attention in the laboratory because of their relevance to current learning theories, were studied to determine their effects upon the retention of complex material presented in the controlled situation provided by the teaching machine. One variable, repetition, is of fundamental importance in incremental learning theory (Estes, 1960). When applied to verbal learning, incremental theory postulates that each reinforced practice trial of a verbal association being learned increases the strength of that association, making it more resistant to interference and consequently to forgetting. Experimentally, this hypothesis predicts that an increase in reinforced repetitions during learning will facilitate retention.

The research testing this prediction is not entirely consistent, and seems to depend to some extent upon the type of learning material used. Studies of overlearning by Krueger (1929) and Postman (1961), using noun lists, have demonstrated a positive relationship between retention and the number of repetitions permitted following original learning to a criterion of one perfect trial. Recent research on one-trial verbal learning by Rock (1957), Rock and Heimer (1959), Jones (1962), Reynolds (1963), and others have also shown facilitation in retention as a result of increased repetition in the learning phase. Estes (1960), on the other hand, has reported a one-trial learning investigation in which no relationship between reinforced repetitions and retention was found. In still another area of laboratory research, retroactive inhibition studies have demonstrated almost without exception that susceptibility to interference from interpolated tasks is inversely related to number of repetitions during original learning (Slamecka and Ceraso, 1960, p. 452).

These investigations of overlearning, one-trial learning, and retroaction, all of which employed either serial or paired-associate tasks, generally support the incremental hypothesis. However, an early review of verbal learning research by Welborn and English (1937) cites a number of studies indicating that repetition has little or no effect upon retention of meaningful material. Later experiments with connected discourse by Slamecka (1959) contradict these early investigations, and imply that the discrepancy may be due to differences in control over the experimental variables. One of the experiments reported here tested this implication, predicting that retention of complex academic material would vary directly with repetition when the experimental conditions were controlled by the use of teaching machines.

The second objective of the present experiments was to investigate the effect of spaced review upon retention of a complex academic learning task. Spacing of review was defined as the presentation of repeated short practice sequences of an original learning task, inserted between interpolated tasks which follow original learning. While spaced review, as used here, bears some similarity to laboratory investigations of distributed practice, the procedures which define spacing necessarily differ from those defining the distribution of practice variable because of differences in the type of material employed. In general, in distribution of practice studies, serial tasks are presented over a number of learning trials and distribution of practice is determined by the size of the intertrial interval (Underwood, 1961). The size of the intertrial interval normally does not exceed three minutes, and no interpolated learning task is presented during the interval. These operations, designed to test theoretical hypotheses, are quite inappropriate for complex learning. (In Underwood's words, "Even under the most favorable conditions for facilitation of DP, one could not recommend its use in an applied setting where verbal materials are to be mastered" (1961, p. 230).) Consequently, distribution of practice as defined in the laboratory was not tested directly in its application to meaningful material. Rather, the implication from past research that retention is facilitated under distributed time conditions (McGeoch and Irion, 1952, pp. 156-158) was revised to apply to a meaningful learning context. In this context, distribution over time consisted of presenting review trials of previously learned material at varying times subsequent to original learning, with new learning tasks interpolated between reviews. This situation matches many real learning conditions in which a variety of tasks are learned and each one is interpolated with respect to the others. The prediction was that distribution of review sequences over time would facilitate retention to a greater extent than would learning and review which was massed into one continuous sequence.

The first experiment reported was performed to evaluate the effects of both repetition and spaced review upon meaningful retention. Results of this initial experiment raised new questions which required two additional investigations, which are reported as Experiments II and III.

Experiment I: Effects of Repetition and Spaced Review Upon Retention

Method

Materials

The program sequence used in Experiment I was a 1280-frame Biology chapter taken from a larger junior high school General Science program that was in the final stages of development. The Biology chapter covered ten topics in biology, arranged in the standard linear order without spaced review. By rewriting certain topical sequences, and rearranging the order of presentation of some topics, five variations of the Biology program were constructed for experimental use. The resulting experimental programs included three versions containing different levels of repetition, and two versions containing spaced review.

Within the original program, a 115-frame programmed sequence teaching the topic of Mitosis was selected for experimental variation. This sequence required that each S learn 11 new technical terms associated with the mitosis process, and learn also the changes in cell structure which take place during mitosis and the order in which the changes occur. The 11 terms which the S was required to learn, and be able to use in describing cell reproduction, were mitosis, interphase, prophase, metaphase, anaphase, telephase, spindles, cell plate, indentation, equator, chromosomes.

Repetition Sequences. The original 115-frame Mitosis sequence, called the M-1.0 sequence, was used as a basis for constructing three repetition levels. First, the number of stimulus and response repetitions of the 11 new terms were tabulated. Two new sequences were then constructed, one sequence containing one-half as many repetitions of each of the 11 terms, and the other sequence containing one-and-one-half times as many repetitions of the 11 terms, as the original sequence M-1.0. These additional sequences, designated M-.5 and M-1.5 respectively, constituted the low and high levels of repetition of the new material being learned, and the original M-1.0 section was considered the intermediate repetition level. Any one of the three sequences could then be inserted into the larger Biology program of 1280 frames, as is shown in Table 1, Col. 1. The total program thus consisted of 581 frames of proactive learning material (PL), covering the topics Cells, Protozoa, Tissues, Organs and Systems, and Green Plants; following by one of the experimental Mitosis sections, which in turn was followed by 582 frames of retroactive learning material (RL), covering the topics of Plant Reproduction, Animal Reproduction, Biological Classification, and Heredity.

Table 1

Title and Order of Biology Topics Presented to Three Repetition Groups and Two Spaced Review Groups

Order of Topics	(1) Repetition Groups (M-.5, M-1.0, M-1.5)[1]	(2) Review Group (R-1.0, R-1.5)
1	Cells	Cells
2	Protozoa	Protozoa
3	Tissues (581 frames)	Tissues (581 frames)
4	Organs and Systems	Organs and Systems
5	Green Plants	Green Plants
6	Mitosis (.5 or 1.0 or 1.5)	Mitosis (.5 or 1.0)
7	Plant Reproduction (107 frames)	Plant Reproduction (107 frames)
(Review)	- - - - - -	Review Mitosis (28 frames)
8	Animal Reproduction (217 frames)	Animal Reproduction (217 frames)
(Review)	- - - - - -	Review Mitosis (22 frames)
9	Classification (165 frames)	Classification (165 frames)
10	Heredity (95 frames)	Heredity (95 frames)

[1]The numbers 0.5, 1.0, and 1.5 indicate the relative amounts of repetition used in the different experimental versions.

Review Sequences. The Mitosis topic was also reorganized in a manner that permitted two versions containing spaced review. Two review sequences, consisting of 28 and 22 frames containing practice in the 11 critical Mitosis terms, were written. The total number of stimulus and response repetitions in these two sequences combined was equivalent to the number of repetitions in the M-.5 section. By adding these sequences to the Biology program after the two topics that followed mitosis, i.e., after Plant Reproduction and Animal Reproduction respectively, the requirements for spaced review were met, and at the same time the amount of repetition necessitated by the additional spaced review frames was controlled. As a result of this procedure, a spaced review program (R-1.0) containing an amount of repetition equal to M-1.0, and another review program (R-1.5), containing an amount of repetition equivalent to M-1.5, was constructed. The structure and content of these review versions are illustrated in Table 1, Col. 2.

Tests. Measures of unaided recall, aided recall, and recognition of the 11 mitosis terms presented in the five experimental treatments were used to assess retention at various points during the experiment. The Unaided Recall test required the S to reproduce drawings of the mitosis process from memory and to describe in technical terms the changes in cell structure that take place at each stage. This test was scored for 22 possible answers. The Aided Recall test (15 items) consisted of incomplete sentences which required the S to use the 11 experimental mitosis terms as fill-ins. The Recognition measure was a multiple-choice test in which recognition of the 11 experimental terms was required in answering 18 questions.

Two additional measures, an Aided Recall (completion) test and a Recognition (multiple-choice) test covering material from other program topics (Cells, Plant Reproduction, and Animal Reproduction) were also used. These tests (39 and 20 items, respectively) were called control measures since all groups had received the same treatment for each of the three topics tested. The particular control topics chosen represented both PL and RL, so that differential effects of massing and review upon retention of prior and subsequent learning could be assessed.

Subjects

A total of 75 junior high school students participated in the experiment. Scholastic aptitudes, as measured by the Otis Quick-Scoring Mental Ability Test (Beta), ranged from 100 to 134, with a median I.Q. of 117.[3] At the time of the experiment, all Ss belonged to one of three classes taking a general science course. None of the Ss had taken previous courses in biology, and none had had previous experience with programmed instruction.

Design and Procedure

Prior to the experiment, Ss with equivalent intelligence were assigned to one of five groups by a randomized blocks method (Edwards, 1960, Ch. 11). Each of the five groups received one of the experimental versions of the program. The programs were administered with Min-Max I teaching machines to all groups in 20 work sessions, each session lasting 40 minutes. At the beginning of every work session, the teacher, who served as E for all groups, assigned to each group the number of frames that were to be completed in that session. Ss finishing an assignment early were permitted to use the remainder of the session as a study period, provided they did not study biology. Since the programs contained slightly different numbers of frames because of the experimental variations, daily work assignments to the groups varied from session to session, ranging from assignments of 50 frames to a maximum assignment of 80 frames for any single 40-minute session. By regulating the daily assignments, all of the five groups completed the experimental Mitosis section during work sessions 10 and 11, and completed the entire program in the twentieth session.

Before beginning the program, Ss were given the Recognition test as a pretest to determine the equivalence of the five groups on pre-learning knowledge of mitosis. The first retention testing (T_1) was administered in the two days immediately following completion of the program. In T_1, Aided

---

[3]We wish to thank Warren D. Shepler, Assistant Superintendent, and J. Ernest Harrison, Director of Curriculum at the Baldwin-Whitehall Schools, for their aid in providing subjects for Experiment I.

Recall tests were given first for both the experimental and control material. These measures were followed by the experimental and control Recognition tests. After a three-week interval, during which Ss were not exposed to any of the material learned in the programs, a second retention testing (T_2) was administered. T_2 was composed of four separate tests, presented in order of decreasing difficulty. First the Unaided Recall test was administered, which had not been given at T_1 but was used in T_2 both as a warm-up task and as an additional retention measure; then the Aided Recall (completion) and the Recognition (multiple-choice) tests of Mitosis, and finally the Recognition test for the control materials, were given. The Ss were unaware that this T_2 battery was to be administered, and all four tests were given in a single session to prevent the possibility of reviewing.

Results

Table 2 summarizes the means and standard deviations of all groups on the pretest and the various measures obtained during T_1 and T_2. Analyses of variance, employing the randomized blocks technique (Edwards, 1960, Ch. 11), were used to compare the groups on each of these measures. Several Ss were absent at various times during the series of testings. Each absence required that the entire block with which the absent S was matched be eliminated from the analyses, reducing the size of all groups by one. Fortunately, the absences were distributed over the testing periods in a way that necessitated removal of only one or two blocks of Ss from each of the analyses made. However, it was necessary to remove different blocks on different analyses, so that the group N's of 13 or 14 that were used in the various analyses did not always represent the same Ss in each analysis.

As can be seen in Table 2, mean scores among the five groups on the pretest ranged from 2.54 to 3.62. An analysis of variance showed that the pretest differences among groups were not significant ($F = 1.22$; df/4, 64; $P > .05$). A series of correlated t tests, made for each group on the differences between the Recognition pretest scores and the T_1 Recognition scores, yielded significant t values ranging from 2.34 to 3.84, indicating that the higher mean scores for each group at the time of T_1 were due to the effect of the program treatments rather than chance.

Table 2

Summary of Means and Standard Deviations for all Groups on Pretest and Retention Tests Administered at T_1 and T_2

Test	Total Possible Score		Repetition M-.5	Repetition M-1.0	Repetition M-1.5	Spaced Review R-1.0	Spaced Review R-1.5
Pretest							
Recognition (Mitosis)	18	$\bar{X}$	3.62	2.92	2.54	3.92	2.92
		s	1.66	1.98	1.76	2.43	2.22
T_1							
Aided Recall (N = 14)							
Mitosis	15	$\bar{X}$	3.79	5.21	7.57	10.00	9.79
		s	3.49	4.59	5.67	4.76	4.69
Control	39	$\bar{X}$	18.57	17.50	17.79	22.14	19.36
		s	9.15	7.05	9.25	9.85	7.37
Recognition (N = 13)							
Mitosis	18	$\bar{X}$	6.00	7.08	8.54	10.00	8.92
		s	4.26	5.22	4.79	5.93	3.66
Control	13	$\bar{X}$	11.31	12.77	12.69	13.08	12.38
		s	3.55	4.17	3.61	3.25	2.47
T_2							
Unaided Recall (N = 14)							
Mitosis	22	$\bar{X}$	2.93	4.29	3.14	6.57	6.14
		s	3.60	4.92	3.80	4.90	4.87
Aided Recall (N = 14)							
Mitosis	15	$\bar{X}$	4.86	6.14	6.71	9.50	10.07
		s	4.45	5.19	5.40	4.77	4.75
Recognition (N = 14)							
Mitosis	18	$\bar{X}$	6.79	7.71	8.29	10.50	10.93
		s	4.21	5.92	5.72	6.35	6.08
Control	20	$\bar{X}$	12.29	12.86	12.86	15.00	13.93
		s	3.77	3.51	3.57	2.83	2.62

Repetition Effects. The M-.5, M-1.0, and M-1.5 groups received the experimental T_1 measures of Aided Recall and Recognition on the tenth day following original learning (OL) of the Mitosis topic. A simple analysis of variance showed no significant differences among the groups on the T_1 Recognition test ($F = .98$; df/2, 24; $P > .05$). For the T_1 Aided Recall test, however, a significant difference among means was indicated ($F = 4.50$; df/2, 26; $P < .025$). Further analysis of the Aided Recall results showed that the M-.5 group mean was significantly lower than the mean for the M-1.5 group ($t = 3.12$; df/13; $P < .01$) but that all other mean differences were within chance limits. The results of analyses of variance performed for the T_1 control measures were not significant ($F < 1.00$ and $F = 1.17$ for Aided Recall and Recognition control tests, respectively), implying that the reliable differences obtained for the experimental Aided Recall measure were due to the repetition treatments received and not to general learning differences among the groups.

Twenty-one days after T_1, and 30 days following OL on the topic of Mitosis, the T_2 measures were administered. Groups M-.5, M-1.0, and M-1.5 were given the experimental Aided Recall and Recognition tests again, and received the Unaided Recall test as well. The F values obtained in the analyses of variance made for these three retention measures were all below 1.00, indicating no significant mean differences among the repetition groups on any type of retention. Mean differences on the single control measure given at the time of T_2 were also well within chance limits ($F < 1.00$; df/2, 26; $P > .05$). The partial effect of repetition upon retention found at the time of T_1 apparently dissipated during the rest interval between T_1 and T_2 leaving the M-.5, M-1.0, and M-1.5 groups with equivalent levels of retention on the thirtieth day following OL.

Effects of Spaced Review. The R-1.0 and R-1.5 groups received the same T_1 and T_2 measures as were obtained for the M-1.0 and M-1.5 groups. These four groups were compared on all measures, using a series of 2 x 2 analyses of variance in which the effects of two levels of repetition (1.0 and 1.5), two levels of review (R and M), and possible interaction effects could be determined.

Results of the factorial analyses for the experimental T_1 Aided Recall and Recognition tests are presented in Tables 3a and 3b. Neither the repetition effects nor the interaction between repetition and review were significant on either measure. The effect of review was not significant on the Recognition test. On the Aided Recall test, however, the performance of the R groups, receiving review, was superior to that of the non-review M groups beyond the .01 level of significance.

Tables 3c and 3d report similar analyses made for the control T_1 Aided Recall and Recognition tests. No significant differences were obtained for either treatment or the interaction on the control Recognition measure. As can be seen in Table 3c, however, a reliable difference was found between the R and M groups on the control Aided Recall test. Further analysis of the items on the latter test was made to determine the source of this difference between groups on retention of the materials that were not varied experimentally. Three analyses of variance, made separately for the items covering Cells (10 items), Plant Reproduction (18 items), and Animal Reproduction (11 items), showed that the M and R groups were equivalent in their retention of the Cells and the Plant Reproduction material ($F = 3.79$ and 1.04 respectively; $df/1, 39$; $P > .05$), but that performance of the spaced review groups was significantly higher than that of the non-review groups on the Animal Reproduction items ($F = 4.37$; $df/1, 39$; $P < .05$). The Animal Reproduction topic was the one interpolated between the two spaced review sequences received by the R groups but not by the M groups. Possibly spaced review facilitated aided recall at the time of T_1 not only on the material reviewed, but on the material interpolated between the review sections as well, and this facilitation produced the difference in control-item performance shown in Table 3c.

Tables 3e, 3f and 3g show the analyses of the data obtained from the M-1.0, M-1.5, R-1.0 and R-1.5 groups at the time to T_2. At this time, the R groups demonstrated significantly higher performance than the M groups on all three retention measures taken -- Unaided Recall, Aided Recall, and Recognition. The effect of repetition, and the interactions, were not significant for any of the retention measures. The fourth measure given at T_2 was the control Recognition test, on which no T_1 differences had been found. As shown in Table 3h, however, the difference between the M and R groups in recognition of control materials was significant at the time of T_2.

Table 3

Analysis of Variance Results for Comparisons Between Two Levels of Repetition and Massed versus Spaced Review

Table 3a

The T_1 Aided Recall Test

Source	SS	df	F
Repetition	16.08	1	1.24
Review	171.50	1	13.17 ($P < .01$)
Rep. x Rev.	23.13	1	1.78
Within			
Blocks	764.36	13	
Error	507.79	39	
Total	1482.86	55	

Table 3b

The T_1 Recognition Test

Source	SS	df	F
Repetition	.48	1	- -
Review	35.56	1	2.31
Rep. x Rev.	20.94	1	1.36
Within			
Blocks	631.31	12	
Error	553.76	36	
Total	1242.06	51	

Table 3c

The Control T_1 Aided Recall Test

Source	SS	df	F
Repetition	21.88	1	- - - -
Review	135.16	1	4.22 ($P < .05$)
Rep. x Rev.	33.01	1	1.03
Within			
Blocks	2478.09	13	
Error	1246.69	39	
Total	3914.84	55	

Table 3d

The Control T_1 Recognition Test

Source	SS	df	F
Repetition	1.92	1	- - - -
Review	0.00	1	- - - -
Rep. x Rev.	1.23	1	- - - -
Within			
Blocks	261.73	12	
Error	303.35	36	
Total	568.23	51	

Table 3e

The T_2 Unaided Recall Test

Source	SS	df	F
Repetition	7.35	1	- - - -
Review	88.82	1	7.76 ($P < .01$)
Rep. x Rev.	.15	1	- - - -
Within			
Blocks	672.10	14	
Error	480.42	42	
Total	1248.85	58	

Table 3f

The T_2 Aided Recall Test

Source	SS	df	F
Repetition	9.60	1	- - - -
Review	166.66	1	13.15 ($P < .01$)
Rep. x Rev.	2.40	1	- - - -
Within			
Blocks	882.73	14	
Error	532.33	42	
Total	1593.73	59	

Table 3g

The T_2 Recognition Test

Source	SS	df	F
Repetition	4.81	1	- - - -
Review	88.81	1	4.54 ($P < .05$)
Rep. x Rev.	3.76	1	- - - -
Within			
Blocks	1131.23	14	
Error	822.36	42	
Total	2050.98	59	

Table 3h

The Control T_2 Recognition Test

Source	SS	df	F
Repetition	4.01	1	- - - -
Review	36.16	1	5.56 ($P < .05$)
Rep. x Rev.	4.02	1	- - - -
Within			
Blocks	264.80	13	
Error	253.55	39	
Total	562.55	55	

Reminiscence. Mean scores of the combined M-1.0 and M-1.5 groups and the mean scores of the combined R groups on the T_1 and T_2 Aided Recall test, Recognition test, and Control Recognition test, are platted in Figures 1, 2, and 3. Figure 1 shows essentially no change in mean aided recall score for either the M or R conditions over the 21-day period between T_1 and T_2. In Figures 2 and 3, it can be seen that the mean scores of both the M and R groups were higher on the experimental and control Recognition tests given at the time of T_2 than they were on the T_1 administrations of these measures. On each of the recognition tests, the R group increment was greater than that of the M group, accounting for the significant T_2 differences found in the analyses of variance (Tables 3g and 3h). Possibly the effect of the M and R treatments was not to produce different amounts of forgetting over the 21-day period, but rather to produce different degrees of reminiscence.

Discussion

The results from the T_1 testing, which was given to the M-.5, M-1.0 and M-1.5 repetition groups after each had received the same amount of retroactive learning (RL), indicated that the degree of OL repetition had a partial effect upon short-range retention, facilitating recall of the M-1.5 group relative to the M-.5 condition. However, the effect of increased repetition upon T_1 recognition was not significant. These T_1 results only partially support previous studies which have shown that increased OL is directly related to retention following equivalent amounts of RL (cf. Slamecka and Ceraso, 1960). Many of the retroaction studies have employed relearning as the retention measure rather than recognition and recall, which may account in part for the discrepancy in findings. Nevertheless, the present results suggest that OL repetition does not have as consistent an effect upon retention as has previously been assumed, at least with regard to complex academic material.

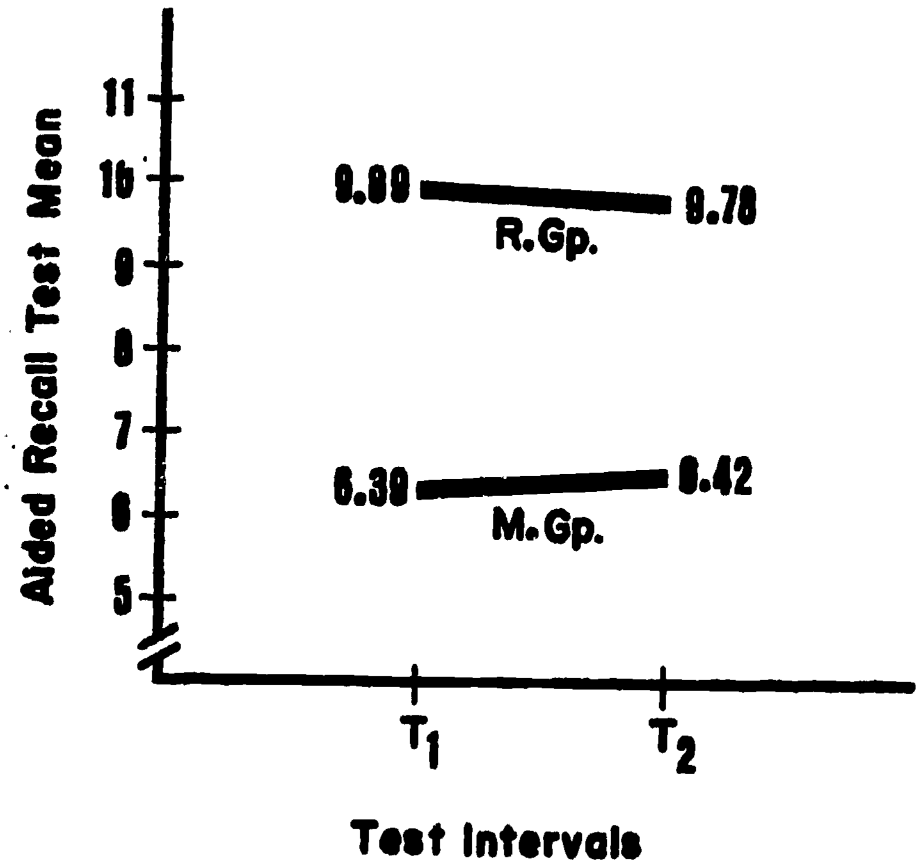

Figure 1. Retention of M and R Groups on the Experimental Aided Recall Test

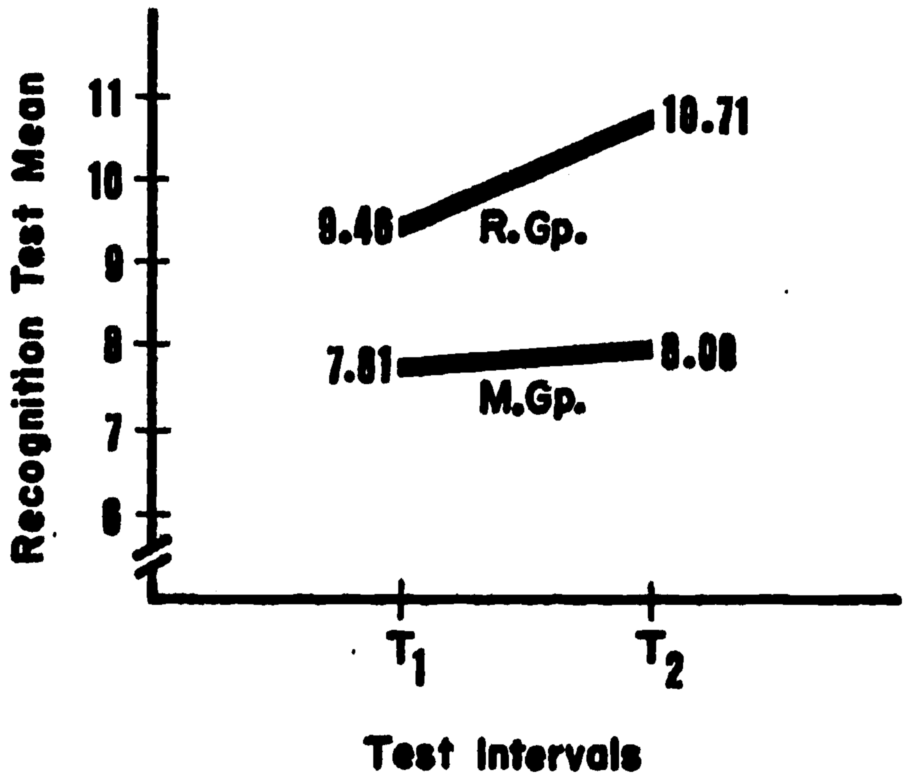

Figure 2. Reminiscence of M and R Groups on the Experimental Recognition Test

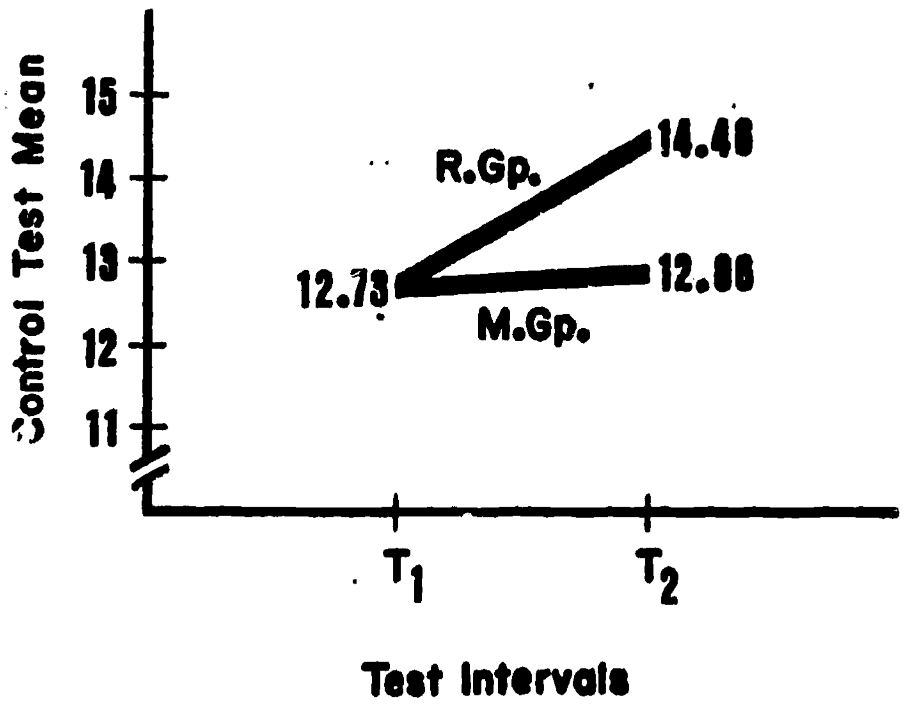

Figure 3. Reminiscence of M and R Groups on the Control Recognition Test

Without exception, no differences in recall or recognition among the repetition groups were found 30 days following the OL task, indicating that repetition had only transitory effects upon retention of complex materials. These T_2 results agree with the early research with meaningful material described by Welborn and English (1937). Alternatively, a contradiction is found when the T_2 findings are compared with the long-term retention reported in Krueger's (1929) investigation of overlearning. In that study, differences in OL repetition of noun lists produced related differences in recall after forgetting periods as long as 28 days. It is possible, however, that this difference in the long-term findings was due to methodological differences in testing for long-term retention, rather than in the type of material retained. Kreuger's data was based upon groups which received recall tests only once; i.e., only at the 28-day interval. In the present study the T_2 results are based upon groups which had received the Aided Recall and Recognition tests once before at T_1. Ammons and Irion (1954) have shown that an initial test may function as a learning trial, raising performance on later retention testings. Exposure to T_1 on the tenth day following original learning may have constituted a relearning trial for all of the repetition groups, giving them equal opportunity to restore material partially forgetten and resulting in equivalent retention among the groups at T_2. The fact that some reminiscence occurred between T_1 and T_2 supports such an interpretation. This methodological problem was investigated further in Experiment III.

In contrast with the short-term and transitory effect which repetition apparently had upon retention, the spaced review treatment produced significant differences favoring the review condition at both the T_1 and T_2 retention intervals. The superiority of the spaced review groups on T_2 is reflected in the warming-up test of Unaided Recall as well as the Aided Recall and Recognition measures. The findings appear to support an assumption that distribution of practice between interpolated learning tasks facilitates retention of the distributed material. However, several problems arise concerning the present data which would make such a conclusion premature. First, it is possible that the significant contribution of spaced review to learning and retention was due to the effects of time between the last practice trial and the tests, rather than to distribution of practice as such.

Groups receiving the massed and review treatments were exposed to the OL materials at the same time, i.e., on the tenth and eleventh days of the experiment, after receiving equivalent amounts of PL. The amount and order of new material (i.e., RL) introduced between OL of Mitosis and the two testings was also equivalent for the two conditions. For the massed groups, which received the final Mitosis trial on day 11, the intervals between the final learning trial and T_1, and between the final learning trial and T_2, were 9 and 30 days, respectively. However, the review groups did not receive the final spaced review trials of the experimental material until day 16, making the time intervals between the final exposure to the experimental material on the T_1 and T_2 tests 4 and 25 days, respectively. The superiority of the review group on the retention tests may have been due in part to the shorter intervals between final practice trials and testing, rather than to the spacing of review sequences alone. The possible influence of these differences in recency of practice prior to retention testing was investigated further in Experiment II.

A second consideration to be made before interpreting the effect of distribution of practice upon retention concerns the results obtained on the control items of the T_1 Aided Recall test. The superior retention of the review group on the control materials was unexpected, since all groups were matched in intelligence and had presumably received the same learning treatments on the control topics. One explanation of the finding is suggested by the fact that on T_1 the differences in control test performances were restricted to one of the topics interpolated between the review sequences, i.e., Animal Reproduction. Possibly insertion of review sequences both before and after this topic gave the review group practice in discriminating between the Mitosis and Animal Reproduction materials. Such discrimination training could have resulted in less interference between the experimental and control material at the time the retention tests were given. A second possibility is that some of the Ss making up the review group had an advantage of prior knowledge in that particular control topic which the others did not. The latter explanation is reasonable in view of the fact that the participating Ss came from varied school backgrounds and might easily have been exposed to different science curricula in preceding science courses. These tentative explanations are discussed further in the next experiment.

A final problem to be considered in attempting to evaluate the spaced review effect is the occurrence of reminiscence on both the experimental and control Recognition tests administered at T_2. The retention performance of the spaced review group was not significantly higher than the massed group on either of these tests at time T_1, and its superiority at T_2 was due to an increase in its own performance rather than to retention decrement for the massed group. This increased retention performance after a period of forgetting is similar to the reminiscence phenomenon first discovered by Ballard (1913), and subsequently shown by Ammons and Irion to be due to the learning effects of an earlier test. The Ammons and Irion finding suggests that the reminiscence obtained in the present study was a function of practice effects from the administration of T_1. However, explanations of why spaced review instruction produced greater amounts of reminiscence over a 21-day period than did non-spaced instruction, or why reminiscence was evident on the recognition measures only, are not readily available from past research evidence. Reminiscence studies have generally indicated that the phenomenon is not evident after rest intervals of more than a few days, and that reminiscence is less likely to occur following distributed practice than massed practice (McGeoch and Irion, 1952, Ch. 5). In contrast with the past findings, the present results suggest that distribution of practice produces a long lasting reminiscence effect, at least when the learning materials are academic subject matter and periodic retention testings are administered. A further investigation of these implied relationships among distributed practice, testing, and reminiscence is reported in Experiment III, which investigates the effect of prior testing upon retention after spaced review.

Experiment II: Spaced Review Effects Upon Retention With Recency Controlled

In the first experiment, the type, amount, and presentation order of the materials were equivalent for the massed and review groups, making it possible to control the type, amount, and order of PL and RL tasks. Control of these variables, however, precluded equating the groups in terms of the time interval elapsing between the last practice trial and the retention tests. The diagram below shows that, when type and amount of RL materials (A, B, C D, E) are made equivalent, the recency interval must necessarily vary because the last spaced review trial will, by definition, occur at a later time than will the last massed review trial:

MASSED: | (PL) | OL + Review | A - B - C - D - E - | T_1

Recency Periods

SPACED: | (PL) | OL | -A- | Review | -B- | Review | - C - D - E - | T_1

|———— Time ————>

An attempt to control for recency necessitates either (a) presenting OL and the spaced review trials earlier, which would relinquish equivalence between groups in the type and amount of PL material received, or (b) presenting additional material to the spaced review group between the end of the program and administration of T_1, which would result in differences between the groups in type and amount of RL material received. In either case, some of the control over interference that was present in the first experiment must be relinquished if control of the recency variable is to be achieved. Two considerations were made in determining that the latter alternative was most appropriate for controlling recency in Experiment II. First, Underwood (1957) has shown that PL has a greater effect upon forgetting than RL, indicating the greater need for control over the amount of PL presented. Second, the topics in the program being used were not entirely independent of each other, so the change in topical order which would result from presenting the experimental material earlier in the sequence to one group might precipitate a learning disadvantage for that group.

For these reasons, in controlling for recency the PL material was held constant for a massed and a spaced review group, and the RL material was allowed to vary as in the diagram below:

MASSED: | (PL) | OL + Review | - A - B - C - D - E - |

Recency Periods

SPACED: | (PL) | OL | -A- | Review | -B- | Review | - C - D - E - | (6 days) T_1

| ---------- Time ----------→

This design replicated the massing and spacing procedures of the first experiment in all respects except that the spaced review group received the first retention testing six days following completion of the program instead of immediately after completion, making the total time interval between the last practice trial of the experimental material and the retention test ten days in duration for both massed and review conditions. In doing so, however, the spaced review group necessarily received six days more RL than did the massed group.

Method

Subjects. Two intact eighth grade science classes were selected to participate in the experiment on the basis of the following criteria:[4] neither had been exposed to biology during the school year, neither had had previous familiarity with programmed instruction, and the mean I.Q.'s of the two groups were equivalent. The classes chosen were from different schools and had different science teachers.

[4] We wish to thank Evan W. Ingram, Associate Superintendent of Instruction and Ralph Scott, Director of Instructional Services, Secondary Schools of the Pittsburgh Public Schools, for their aid in providing experimental classes for the second and third experiments.

Materials. Prior to Experiment II certain modifications were made in the first five sections of the original Biology program in order to make it a more effective instructional tool.[5] Modifications consisted of making minor changes in the wording of certain frames, and adding new frames in some cases to clarify difficult sections. The resulting program differed from that used in Experiment I in that it was 64 frames longer (totaling 1344 frames) and was divided into fourteen sections rather than ten. Table 4 summarizes the newer revision, giving the names, order, and sizes of each topic. Since the changes occurred only in control topics, their possible effects upon the data of Experiment II were equivalent for the experimental treatments compared. All programs were administered with Min-Max II teaching machines.

Table 4

Description of Biology Program Presented to All Groups in Experiment II

Unit	Topic	No. of Frames
1	Introduction to Cell Structure	150
2	The Plastids	74
3	The Nucleus	43
4	The Cytoplasm	74
5	Animal and Plant Cells	31
6	Protozoa	39
7	Tissues, Organs, and Systems	50
8	Green Plants	183
9	Mitosis	115
10	Reproduction of Seed Plants	107
11	Animal Reproduction	217
12	Men in Biology	39
13	Classification of Plants and Animals	123
14	Heredity	99
	Total Frames	1344

[5] The revised edition used in this experiment is part of the General Science Program published in 1962 by TMI-Grolier, New York. It was reprinted for experimental purposes with permission of the publisher.

The experimental Mitosis sections inserted into the revised program were in the same sequences used in Experiment I, and the retention tests were also identical to those previously described.

Design and Procedure. A massed-learning group (M) received the M-1.5 treatment as described in the first experiment followed by a T_1 retention testing at the end of the program. Because of the increased length of the revised program, the time taken to complete it was one day longer than in Experiment I. Consequently, the recency interval between the last trial of the Mitosis topic and T_1 was 10 days for Group M. The spaced review group (R) was administered the same treatment as described previously for the R-1.5 group, except that the T_1 testing was not given immediately following completion of the program. Instead, Group R received five science periods of teacher instruction in Astronomy after finishing the program, and received the T_1 testing in the sixth science period. These interpolated Astronomy periods extended the Group R interval between the last review trial of Mitosis and T_1 to 10 school days, making it equivalent to the recency interval of Group M. As in Experiment I, a T_2 retention testing was administered to both groups three weeks following T_1.

Testing procedures varied slightly from those used in Experiment I so that more information concerning prior knowledge and learning of the control materials could be obtained. The pretests consisted of the Recognition measures for both experimental and control materials, providing an assessment of equivalence of the groups in general knowledge of biology as well as knowledge of Mitosis. At both T_1 and T_2 the Aided Recall and the Recognition tests for the experimental and control topics were administered. As in the first experiment, the Unaided Recall test of Mitosis was given only at T_2, as a warm-up task and an additional retention measure.

Daily administration of the program was accomplished as previously described. On each day, certain frames were assigned by the Es, and Ss were instructed to complete only those frames. Thus the number of learning days and the amount of material presented per day was controlled for the two conditions. All tests at T_1 and T_2 were administered in one session, without prior warning from the E that they would be given. In the interval between T_1 and T_2 both groups received instruction in science topics unrelated to biology, minimizing the probability of systematic practice and review during the three-week forgetting period.

Results

Two Ss in Group R and six Ss in Group M failed to complete the entire program and all retention tests because of absence, and were eliminated from the final data. The results for the remaining subjects are summarized in Table 5, including the measures taken for the R and M groups prior to administration of the program, and also at the retention testing intervals which occurred 10 days (T_1) and 31 days (T_2) following the last review trial of the Mitosis material. Differences between the group means were evaluated by a series of independent two-tailed *t* tests, and the resulting *t* values are also included in Table 5.

There were no significant differences between any of the pretest measures, indicating that the groups were equivalent in intelligence and also in pre-program knowledge of the Mitosis and control materials. The means on the Recognition pretest of Mitosis are no higher than would be expected from guessing on an 18-item 5-choice multiple-choice test, suggesting that neither group had any knowledge of the Mitosis topic prior to taking the program. Control-item means for the Recognition pretest were above chance limits of guessing, however, reflecting some degree of pre-program knowledge of the control materials.

As shown in Table 5, no significant differences between Group M and Group R were found for any of the control tests at either T_1 or T_2. The two groups, having been exposed to the same treatments for the control material, apparently retained their learning to the same degree. On the experimental Mitosis materials, however, significant differences were found between the groups on all tests administered at T_1 and T_2. All differences were in the same direction, with Group R demonstrating higher retention performance than Group M regardless of the type of retention tested or the length of the retention interval. The consistent superiority of Group R indicates that, with the time interval between the last review trial of Mitosis and the retention measures controlled, spaced review of the experimental material produced significantly greater recall and recognition of that material than did the massed-learning treatment.

Table 5

Means and SD's of R and M Groups on All Tests and Results of t Tests for Differences Between Means

Test	R (N = 23)		M (N = 35)		
	$\bar{X}$	s	$\bar{X}$	s	t
Pretests					
I.Q.	118.74	6.81	119.17	9.98	1.18
Recognition					
Mitosis	2.87	1.49	3.37	1.93	1.05
Control	10.30	2.28	11.31	2.52	1.55
T_1					
Aided Recall					
Mitosis	11.00	3.44	8.14	4.41	2.62*
Control	20.30	7.55	20.37	6.39	0.04
Recognition					
Mitosis	12.22	4.00	9.23	4.73	2.50*
Control	13.96	2.79	14.91	2.28	1.42
T_2					
Unaided Recall					
Mitosis	11.74	4.21	7.00	3.49	4.58***
Aided Recall					
Mitosis	11.13	3.40	8.17	4.71	2.60*
Control	21.52	8.08	21.66	6.09	0.75
Recognition					
Mitosis	12.83	4.18	9.34	4.44	3.00**
Control	14.09	2.25	14.17	2.72	0.12

* significant at .02 level,
** significant at .01 level,
*** significant at .001 level,

with 23 + 35 - 2 = 56 degrees of freedom.

The Aided Recall and the Recognition tests were administered to both groups at both T_1 and T_2. Inspection of differences in means for the same group between the two retention intervals reveals that the performance of each group was very slightly higher on the T_2 measure than on the T_1 measure for nearly all of the tests. A single exception is found on the control-item Recognition test for Group M, where the mean dropped from $\bar{X}_{T1} = 14.91$ to $\bar{X}_{T2} = 14.17$. These slight but consistent increases over the 21-day interval between retention testings are similar to those found in the previous experiment.

Discussion

With the recency interval between review and testing controlled, the review groups in the second experiment were consistently superior to the massed groups on all retention tests of the experimental topic, despite the fact that controlling for recency necessitated exposure of the review group to more RL material than the massed group received. These results confirm and extend the findings from Experiment I, indicating that spacing of review sequences facilitates both recall and recognition of complex subject matter over periods at least as long as 30 days following the last learning contact with the material reviewed. The findings from the present experiment leave little doubt that it was the distribution of review sequences per se, rather than attendant artifacts which necessarily arise from the distribution procedures used, which facilitated retention in both the first and second experiments.

In the first experiment, the spaced review group demonstrated superior aided recall of control topics at T_1, and superior recognition of control materials at T_2. Additional control tests were administered in Experiment II at pretesting and at the two retention intervals in order to determine more precisely the effect of spacing upon retention of the control topics. However, results of the second experiment indicated no retention differences between the spaced review and massed groups on any control measure. Two explanations need to be considered in accounting for this discrepancy.

First, it is possible that extension of the recency interval for the experimental material adversely affected the review group's retention of the control topics. The additional six-day interval imposed upon this group in Experiment II may have permitted an initially superior retention of the control material to dissipate to a level not significantly superior to that of the massed group. An alternative explanation, however, is that the retention differences between the massed and review groups on the control tests in Experiment I were due to unmeasured differences between the groups in prior knowledge of that material. In Experiment II the control pretest showed no such differences, and the groups were equivalent on all retention tests of the control topics. The data do not offer conclusive evidence for either explanation. However, the first explanation assumes that retention performance is sensitive to differences in recency, and that forgetting occurred in the review group when the recency interval was extended. Such an assumption is tenuous in view of the fact that neither forgetting nor recency effects were demonstrated in either experiment. Lacking evidence to the contrary, it seems best for the present to attribute the control-item differences obtained in Experiment I to variations between the groups in prior knowledge of biology rather than to effects of the experimental treatments.

The results obtained in Experiment II are inconsistent with the previous results which suggested that spacing of practice facilitated reminiscence. In the present experiment, the increments observed between T_1 and T_2 are too small to suggest an improvement in retention over time, and it is doubtful that they should be considered as evidence for reminiscence. Instead, they indicate simply that retention was maintained at full strength during the 21-day forgetting interval. Whether this maintenance of retention is an artifact of the testing procedures used, and how much time must elapse following learning before a retention decrement is observed, are questions which were explored in Experiment III.

Experiment III: Effects of Prior Testing and an Extended Forgetting Interval Upon Retention

The failure of the review group in Experiment II to demonstrate a greater degree of reminiscence than the massed group contradicted the assumption made in the first experiment that reminiscence is facilitated by the spaced review treatment as such. However, the consistent observations in both experiments of slight degrees of improvement in retention over the 21-day forgetting period, rather than retention decrement, was unexpected. A third experiment was performed to determine if this sustained resistance to forgetting might be an artifact of the prior test practice experienced immediately following the completion of the learning task. An attempt was also made to observe retention decrement after a longer forgetting interval than was used in the previous experiments.

Method

Subjects. Two intact eighth grade classes were used. The classes were selected from two schools, after first determining that they were equivalent in intelligence. Neither class had studied biology during the current school year, and neither had had previous experience with programmed instruction. Both were taking a general science course which met for three school periods each week.

Materials. The spaced review program used was the same one described for Experiment II, and all pretests and retention measures were identical to those employed in the two preceding experiments. The programs were administered in Min-Max II teaching machines.

Design and Procedure. A modification of the design used by Ammons and Irion (1954) was employed. Two groups were first given the spaced review program. One Group (Group $T_1 - T_2 - T_3$) then received an immediate retention testing and a second testing three weeks later, as in previous studies, but also was given a third testing three weeks following the second. The second group (Group T_2) was not tested immediately following the program, but received its first retention testing three weeks later, at a time (T_2) corresponding to the second testing of Group $T_1 - T_2 - T_3$. Retention performances of the two groups at time T_2 were compared to determine if the prior testing given Group $T_1 - T_2 - T_3$ resulted in higher performance than the group which had no prior test. Also, the performance trends of Group $T_1 - T_2 - T_3$ over the two successive 21-day forgetting intervals were examined for evidence of retention decrement.

Each group was given the experimental and control Recognition tests before beginning the program. Then the spaced review program was presented during the three weekly science periods for 20 consecutive sessions. As in the former experiments, the groups were given specific work assignments in each session to control the amount of learning material presented per session. The procedure for program administration was identical to the one used in the previous experiments except that the program sessions took place only three periods per week, rather than five.

Group $T_1 - T_2 - T_3$ received the experimental and control Aided Recall and Recognition tests at T_1, and both groups received these tests plus the Unaided Recall test in all other testings. As in the previous experiments, no notice of the testing was given until the period in which they were administered. All science periods between testings were devoted to instruction in topics unrelated to biology.

Results

Prior Testing Effects. Eight Ss in Group $T_1 - T_2 - T_3$ and two Ss in Group T_2 failed to complete the program and all required retention tests because of absence, and were eliminated from the final data. Table 6 summarizes the results of the pretests and the first two retention testings for Group $T_1 - T_2 - T_3$, which received test practice prior to T_2, and Group T_2 which had no prior test. The results of independent two-tailed t tests for differences between pretest means indicated that the groups did not differ significantly in intelligence, but that prior knowledge of both the control and mitosis materials was significantly higher for Group T_2 than Group $T_1 - T_2 - T_3$.

Inspection of the pretest protocols revealed differences between the groups in the number of test items omitted, suggesting that the significant pretest differences may have been due to guessing rather than prior knowledge differences.[6] Most of the Ss in Group T_2 answered all of the mitosis items,

[6]The difference in guessing was apparently a result of a departure from the test instructions by one of the teachers.

while many in Group $T_1 - T_2 - T_3$ omitted items they could not answer. A chance score for random guessing on this 18-item multiple-choice test was 3.60. Since both group means are below chance level it appears that neither group had prior knowledge of the mitosis topic, and so the mean difference does not indicate a true superiority of one group over the other. Such guessing differences between groups were not as evident for the control pretest, however, and the significance of the difference between means on this test probably reflects a real difference between groups in prior knowledge of biology in favor of Group T_2.

Consistent with the previous studies, Group $T_1 - T_2 - T_3$ showed no indication of forgetting during the 21-day interval between T_1 and T_2. The means of Group $T_1 - T_2 - T_3$ were exactly the same on the Aided Recall tests for Mitosis given at T_1 and T_2, and on the Aided Recall control test the T_2 mean was slightly higher than at T_1. For the Recognition tests also, mean performance increased a small amount from T_1 to T_2 on the mitosis items, and only a small decrease was found on the control test. To evaluate whether this maintenance of retention during the interval between T_1 and T_2 was due to administration of the T_1 tests, Group $T_1 - T_2 - T_3$ was compared with Group T_2 on all tests given at T_2. As can be seen in Table 6, the mean performance of Group T_2 was somewhat higher than that for Group $T_1 - T_2 - T_3$ on the T_2 tests, and in two cases the difference was statistically significant. Apparently Group T_2 maintained retention over the entire forgetting interval to at least as great a degree as did Group $T_1 - T_2 - T_3$, even though it did not have an opportunity for test acquaintance and test practice at time T_1.

Table 6

Means and Standard Deviations of the T_1-T_2-T_3 Groups on T_1 and T_2 Tests and Results of *t* Tests for Differences Between Means

Test	Group T1-T2-T3 (N = 22)		Group T2 (N = 27)		
	$\bar{X}$	*s*	$\bar{X}$	*s*	*t*
Pretests					
I.Q.	107.00	10.78	109.04	9.41	0.71
Recognition					
Mitosis	0.64	1.38	3.44	1.45	6.83 ($P < .001$)
Control	6.59	3.02	10.04	2.19	4.66 ($P < .001$)
T_1					
Aided Recall					
Mitosis	5.68	4.41			
Control	8.86	5.07			
Recognition					
Mitosis	6.36	4.09			
Control	11.23	2.07			
T_2					
Unaided Recall					
Mitosis	4.09	3.78	4.56	3.75	.44
Aided Recall					
Mitosis	5.68	3.93	8.52	3.68	2.61 $P < .02$)
Control	9.09	4.90	13.37	5.45	
Recognition					
Mitosis	6.77	4.95	7.96	4.96	.84
Control	10.77	2.37	12.04	3.06	1.61

Although there was some difference between groups in the amount of guessing which occurred on the T_2 tests, the effect of guessing was too small to account for the significantly higher performance of Group T_2 over Group $T_1 - T_2 - T_3$ on two of the five T_2 measures. However, it was considered possible that the generally higher retention performance of Group T_2 was related to its superiority in pretest knowledge of Biology. To eliminate any effect this pretest difference might have had upon the retention data, ten pairs of Ss matched on both I.Q. and pretest control-item scores were drawn from the two groups for further comparison. Test means for the matched groups selected, and the results of correlated *t* tests comparing mean differences, are presented in Table 7. The matching procedure failed to eliminate a significant difference between the paired groups on the pretest of Mitosis; but this difference was due in all probability to the guessing variations previously described, and neither mean is as large as the chance score of 3.60 that could be obtained simply by guessing on this test. With the groups otherwise equated on pretest measures, all *t* tests for differences between T_2 means were non-significant, still failing to indicate any forgetting in Group T_2 relative to the performance of Group $T_1 - T_2 - T_3$ after the same time interval. If testing at the T_1 interval had any practice effect which contributed to maintaining or increasing retention performance 21 days later, the effect was too slight to be noticeable in the present data.

Long-Range Retention. The final testing of the $T_1 - T_2 - T_3$ group permitted observations of the effects of a longer forgetting interval than used in the previous studies. The progression of mean retention performances for Group $T_1 - T_2 - T_3$ over the three retention intervals are reconstructed in Table 8. For three of the five measures that were administered, performance improved as the length of the retention interval progressed. Changes in either direction were small, however, and differences in mean performance over the six-week period from T_1 to T_3 were less than 1.00 for all tests. Although these slight changes do not represent reliable reminiscence effects, they do indicate that retention was maintained at a consistent level throughout the six-week forgetting period.

Table 7

Means of T_1-T_2-T_3 and T_2 Matched-Pair Groups on T_1 and T_2 Tests and Results of *t* Tests for Differences Between Correlated Means

Test	Group T1-T2-T3 (N = 10)		Group T2 (N = 10)		
	$\bar{X}$	s	$\bar{X}$	s	t
Pretests					
I.Q.	111.40	6.67	111.50	6.67	0.26
Recognition					
Mitosis	0.80	1.62	3.40	1.5	2.99 ($P < .02$)
Control	8.80	1.55	9.10	1.45	0.88
T_1					
Aided Recall					
Mitosis	6.30	5.08			
Control	9.70	5.56			
Recognition					
Mitosis	7.90	4.31			
Control	12.40	2.12			
T_2					
Unaided Recall					
Mitosis	6.00	4.19	4.90	4.89	0.48
Aided Recall					
Mitosis	6.60	4.30	8.20	3.29	0.82
Control	10.20	4.94	12.60	3.66	1.08
Recognition					
Mitosis	8.30	5.46	6.50	4.30	0.71
Control	11.40	1.78	11.40	3.27	0.00

Table 8

Changes in Means of Group T_1-T_2-T_3 Over Three Retention Intervals

Test	T_1		T_2		T_3
Unaided Recall	--	--	4.09	>	4.05
Aided Recall					
Mitosis	5.68	=	5.68	<	6.14
Control	8.68	<	9.09	<	9.55
Recognition					
Mitosis	6.36	<	6.77	<	7.09
Control	11.23	>	10.77	>	9.77

Discussion

The group which had no test practice prior to retention testing consistently demonstrated recall and recognition performance at least as high as the prior-test group, indicating that retention maintenance was not benefitted by the opportunity for previous practice. The Ammons and Irion finding, obtained with one simple learning task, apparently is not relevant for the more complex and varied learning materials that were used in the present studies. The results of Experiment III permit a conclusion that the evidence of strong resistance to forgetting obtained in the previous experiments was not an artifact of the testing procedure. Instead, the data from the three studies indicate sustained retention over forgetting periods as long as six weeks following the programming task, during which large amounts of interference from other learning tasks occurred.

For three of the five measures that were administered, performance improved as the length of the retention interval progressed. Changes in either direction were small, however, and differences in mean performance over the six-week period from T_1 to T_3 were less than 1.00 for all tests. Although the slight changes do not represent reliable reminiscence effects, they do indicate that retention was maintained at a consistent level throughout the six-week forgetting period.

General Discussion and Summary

The three main findings from this series of experiments are (a) that repetition was not demonstrated to be a major variable influencing retention of complex subject matter, (b) that spacing of review sequences between interpolated learning materials facilitated retention, and (c) that retention of complex learning material was maintained at full strength for a period of at least six weeks, during which time large amounts of interfering material were presented for learning. These results have certain implications for current learning theory and research with meaningful and non-meaningful verbal materials which deserve further consideration. Since the experiments included the use of programmed instruction, the findings are relevant also to research and development of programming techniques.

The procedure used in most investigations of the effect of repetition upon retention of meaningful material has been to vary the number of repeated readings of short verbal passages, and assess the effect of the variations upon rote memory of these passages following a period of rest or interpolated learning. Under these conditions, the several early studies cited by Welborn and English (1937) found no evidence that repetition contributed to retention of meaningful tasks. At least one recent investigation with connected discourse material, however, has demonstrated a positive repetition effect (Slamecka, 1959). The results of the present research, using more complex material and an apparatus which presumably controlled variations in both stimulus and response repetition more precisely than had previous methods, add support to the earlier studies. The findings raise problems for theories of learning which are based upon an incremental hypothesis. Although some interpretations of the present results in the context of incremental theory may be attempted, they are not very convincing. One such explanation is that meaningful material is more susceptible to rehearsal or overt practice during the forgetting interval than is disconnected material simply because rehearsal would have meaning for the learner, and therefore repetitions which were beyond experimental control may have occurred and influenced the results. It seems doubtful, however, that such extra-experimental influences, even if they were present, could alone account for the equivalent retention of groups differing in OL repetition by as much as 200%, since this would imply that one group rehearsed 200% more than another group to gain equivalence on the retention measure. Other incremental explanations are equally unconvincing. Nor can the present results be attributed to a testing artifact, since Experiment III showed that the test procedures used had no significant influence upon retention.

Concern over the apparent failure of research using meaningful connected material to conform to the predictions of an incremental theory is found in a recent article by Gagné (1962). In reviewing a number of military training studies, Gagné has concluded that certain of the learning variables which receive wide theoretical attention because of their predictive power in laboratory situations have little application in meaningful learning contexts. Although repetition may well produce significant learning effects in certain laboratory circumstances, much work needs to be done before its relevance and function in the complex verbal learning situation can be specified.

The consistent and significant effects of spaced review are difficult to relate to previous laboratory research because the manner in which review was distributed in the present studies does not replicate in several respects either the distribution of practice or the retroaction paradigms used in the laboratory. The procedures which ordinarily define distribution of practice consist of presentation of a series of serial-list or paired-associate trials, each trial separated by a period of rest ranging from fifteen seconds to three minutes (Underwood, 1961). In contrast, the present experiments spaced review sequences over several days following original learning, with large amounts of new learning interpolated between the reviews. The extent of the similarity between distribution of practice in the laboratory and spacing of review seems to be limited to the condition that both demonstrate facilitation of retention under certain circumstances that include interpolation of non-practice periods between periods of practice with a given set of learning material. Further generalization of results between distribution of practice and spacing of review does not seem warranted when the procedures have so little in common.

The comparability of spaced review, as used here, and the retroaction paradigm is little better. In their review of research on retroaction inhibition, Slamecka and Ceraso (1960) cite several studies demonstrating that retroactive interference operates upon connected discourse material in the same way as upon serial word lists and paired associates. Since the present investigations presented RL tasks between OL of connected discourse material and a task requiring retention of OL, it might be expected that some generalization from retroaction research to the present results concerning spaced review would be possible. In the present studies, however, the basic retroaction paradigm was altered to include a series of several interpolation-then-review occurrences, while in the experimentation on retroaction only a single interpolated learning period is ordinarily employed between OL and a retention measure. This fundamental difference in procedure stems from the fact that the major objective of retroaction investigations is the study of interference effects rather than spacing effects, and consequently the retroaction data which has been reported is not interpretable in terms of the relative effect of spaced and massed review sequences upon retention following interpolated interference tasks. In brief, then, the spaced review conditions used with complex academic material in the present investigations employed procedures which were similar in some respects to distribution of practice and in others to retroactive inhibition, combining them in a way which produced significant effects upon retention but which cannot be directly compared with past laboratory investigations of verbal learning.

In contrast with the spacing results, the finding that retention of complex material persists over relatively long periods of time is quite consistent with past research using both meaningful and non-meaningful tasks. Davis and Moore (1935) have summarized the data from 61 retention studies which demonstrate Ebbinghaus's forgetting curve, i.e., a rapid decrement in retention during a short period after OL, followed by very little further decrement over longer periods of time. When the Davis and Moore data were plotted, with separate curves for the investigations using meaningful and non-meaningful tasks, the only major difference in the curves was that the long-term retention of meaningful material was generally higher than that for non-meaningful material (Hovland, 1951). In shape, both curves showed rapid forgetting during the first two to three weeks, followed by maintenance at a consistent level for periods as long as 400 days. In the present studies, the first testing occurred ten school days, or two weeks, following OL, a period of time during which the Davis and Moore data indicate that most of the forgetting should have occurred. Thereafter, retention remained stable for the three or six week periods which elapsed before further testing, again coinciding with the earlier findings. As previously discussed in Experiments II and III, the increments observed as time progressed were too small to be considered indicators of real retention improvement, or reminiscence. Rather, the present results simply confirm the previous research showing that consistent retention levels of meaningful material may be expected over a long-range interval (six weeks). It might be assumed, in keeping with previous studies, that initial forgetting did occur immediately after learning, but no measurements were made in the three studies to assess this.

The past and present findings regarding retention decrement imply that any learning procedures which facilitate retention have their facilitating function during what may be called a "critical period" (approximately two to three weeks following OL) and once this period has passed the extent of further forgetting is relatively slight. Within this context, it would seem appropriate to conceptualize spaced review as an operation which was performed during such a critical period, and had a relatively permanent facilitating effect upon retention. The spacing effect was superior to increased repetitions which, in this context, occurred at the beginning of the critical period. Further research, directed at exploring the utility of a critical-period conception of forgetting, and investigating its parameters, might prove useful.

Since programming procedures were used, mention should be made concerning implications of the current results for programmed learning. First, these experiments have demonstrated that the teaching machine is a useful experimental apparatus for the investigation of the learning of connected discourse. A programmed learning sequence permits more control over experimental presentation of complex material in an extra-laboratory learning situation than do methods requiring human instructors and their attendant subjective influences. A second point to be made, however, is that programs, despite any similarities to the memory drum, should probably not be considered as direct extensions of the verbal learning laboratory and therefore subject to the same principles established in theoretical verbal learning research. It has been shown that two of the three main findings from the present experiments bear little resemblance to laboratory results, i.e., retention is not a function of repetition per se, and the spaced review paradigm useful for the investigation of connected discourse academic material differs from the usual laboratory experimental arrangement. The source of the inconsistency encountered in moving from the memory drum to the teaching machine appears to be in the type of material being learned, rather than in the experimental procedures themselves. Increasingly, investigators are examining this problem of the interaction between kind of learning taking place and the properties of the tasks involved, (cf. Glaser, 1964). The implication is that many currently established learning principles may be unreliable bases upon which to determine procedures for either constructing programs or predicting learning outcomes in applied programming situations unless the additional variable of task properties is considered. The gap between laboratory findings and the complex learning situation is not insuperable, but it does exist, and a considerable amount of research will be needed in order to close it. It appears that something like the program and the teaching machine, used as experimental tools, may be of significant help in reaching this objective.

References

Ammons, A. & Irion, A. L. A note on the Ballard reminiscence phenomenon. J. exp. Psychol., 1954, 48, 184-186.

Ballard, P. B. Oblivescence and reminiscence. Brit. J. Psychol. Monog., 1913, #2.

Davis, R. A., & Moore, C. C. Methods of measuring retention. J. gen. Psychol., 1935, 12, 144-155.

Edwards, A. L. Experimental design in psychological research (Revised Ed.). New York: Holt, Rinehart and Winston, 1960.

Estes, W. K. Learning theory and the new "mental chemistry." Psychol. Rev., 1960, 67, 207-223.

Gagne', R. M. Military training and principles of learning. Amer. Psychol., 1962, 17, 83-91.

Glaser, R. Implications of training research for education: In E. R. Hilgard (Ed.), 1964 Yearbook of the National Society for the Study of Education. (In press)

Hovland, C. I. Human learning and retention. In S. S. Stevens (Ed.), Handbook of experimental psychology. New York: Wiley, 1951, Pp. 645-646.

Jones, J. E. All-or-none versus incremental learning. Psychol. Rev., 1962, 69, 156-160.

Krueger, W. C. S. The effect of overlearning on retention. J. exp. Psychol., 1929, 12, 71.

McGeoch, J. A., & Irion, A. L. The psychology of human learning. (2nd ed.) New York: Longman, 1952, Pp. 156-158.

Postman, Leo J. Retention as a function of degree of overlearning. Science, 1961, 135, 666-667.

Reynolds, J. H. Repetition factors in one-trial learning of paired-associates. J. exp. Psychol., 1963, 65, 362-367.

Rock, I. The role of repetition in associative learning. Amer. J. Psychol., 1957, 70, 186-193.

Rock, I., & Heimer, W. Further evidence for one-trial associative learning. Amer. J. Psychol., 1959, 72, 1-16.

Slamecka, N. J. Studies of retention of connected discourse. Amer. J. Psychol., 1959, 72, 409-416.

Slamecka, N. J. & Ceraso, J. Retroactive and proactive inhibition of verbal learning. Psychol. Bull., 1960, 57, 449-475.

Underwood, B. J. Interference and forgetting. Psychol. Rev., 1957, 64, 49-60.

Underwood, B. J. Ten years of massed practice on distributed practice. Psychol. Rev., 1961, 68, 229-247.

Welborn, E. L., & English, H. Logical learning and retention: A general review of experiments with meaningful verbal materials. Psychol. Bull., 1937, 34, 1-20.