COGNITION AND INSTRUCTION, 1984, *1* (3) 341-358

Contextual Enrichment and Distribution of Practice in the Classroom

Steven M. Smith
Texas A & M University

Ernst Z. Rothkopf
Bell Laboratories

Small scale experiments have led us to believe that teaching a lot in a short time is inefficient, perhaps because it overtaxes student resources. This principle, however, has not been adequately tested in realistic instructional settings. Another untested practical notion derived from psychological theory is that diversifying environmental contexts of teaching and providing contextual aids for the organization of course content can make instruction more effective, particularly when it is massed. These predictions were tested with an 8-hour statistics course, consisting of four videotaped lessons. The lessons were presented either within a single day or over 4 days. Environmental context was either diversified, by teaching the lessons in four different rooms, or was held constant by using a single room setting. Retention was tested 5 days after training in a totally new environment. Field dependence of subjects was measured by the Group Embedded Figures test.

We found: (1) Distribution of lessons over 4 days was more effective than a single day presentation; (2) diversification of context by using a different room for each of the four lessons resulted in better productive performance for field-dependent students. Diversification was equally effective for time-massed and distributed teaching.

We concluded that both massing and context effects were primarily retrieval, rather than learning, phenomena because performance at the end of each lesson was not affected by experimental manipulations while recall after 5 days was. Our results indicate that efficiency of instruction can be increased by diversifying the teaching environment. Such context diversification would be a particularly useful instructional strategy in settings where transportation costs necessitate massed teaching.

Requests for reprints should be sent to Dr. Steven M. Smith, Department of Psychology, Texas A&M University, College Station, TX 77843.

This experiment had two major purposes. The first was to test whether the diversification of the physical settings of instruction leads to better retention of content, a hypothesis derived from earlier laboratory research. A second purpose was to test, with realistic teaching material, the widely held belief that cramming a lot of instructional information into a short time period produces poor results. To our knowledge this hypothesis had not been tested with practical subject matter of substantial character. We also wanted to explore the possibility that the diversification of the physical context in which lessons were delivered would overcome any deleterious effects produced by temporally massed teaching. We considered the information sought by our study to be pertinent particularly to short intensive courses, such as might be offered during presessions of professional meetings or in technical training, but we believe our findings would also be relevant to ordinary classroom practices.

There are many reasons to believe that presenting much new information in a short time can overtax students' resources and thus lead to inefficient teaching. Nevertheless, practical considerations, such as geographically dispersed students and high transportation costs, often dictate that substantial amounts of new material be taught within a relatively short time period. This experiment aimed to explore one technique for overcoming inefficiencies in temporally massed teaching.

Evidence for the advantage of distributing learning widely in time comes from small scale laboratory experiments (dating as far back as Ebbinghaus, 1885). These have shown that exposures to a given amount of material distributed over a long period of time leads to better retention than massing the same material into a short time span. Favorable effects of distribution of practice in time have been observed with a wide range of material and experimental conditions. There are indications that this favorable effect on retention, usually referred to as distribution of practice effect, is actually not a unitary phenomenon, but depends upon different underlying psychological mechanisms for various tasks and experimental circumstances.

The psychological mechanisms that have been identified in this connection include (1) fatigue-like states, such as reactive inhibition (Hull, 1952), that dissipate between instructive events during widely spaced acquisition conditions but accumulate during massed conditions and interfere with learning; (2) attention and other relevant processes necessary for learning, that extinguish during massing but recover during the intervals between instructive events during distributed conditions (Rothkopf & Coke, 1966); (3) longer intervals between the presentation of two related instructive events,[1] which re-

[1]It is a reasonable assumption that a series of lectures or other teaching materials include presentations of phrases, words, or other representations that must be considered repeated presentations of the same instructional information. For example, the idea of *variability* will occur many times in a statistics course. The exact nature of instructive events about *variability* will not necessarily be in verbatim-identical forms although they will be relevant to the same underlying competence (see Rothkopf, 1981).

sults in more widely dispersed records in memory. If only a limited region of memory can be accessed in recall, the chances of finding at least one record of an event will be greater under distributed conditions of presentation than during massed teaching (Landauer, 1975; Ross & Landauer, 1978); and (4) distributed acquisition places events into juxtaposition with a variety of contextual elements, whereas massed conditions cause learned material to share common contextual information. A diversity of contextual information results in distinctive memorial codes, which are more favorable for retrieval than codes that share contextual elements among many items (Bower, 1972; Melton, 1970).

The last of these hypothesized mechanisms is of special interest because it is not completely dependent on the duration of the interval between instructive events. Environmental context is shared by temporally contiguous events almost coincidentally. Context may be diversified by systematic manipulation of the environment even under temporally massed conditions of learning.

Regardless of interpretation, consistent reports of distribution phenomena have led to the indirectly supported belief that teaching would be more effective if it were conducted under distributed rather than massed conditions. Much of the fundamental experimentation, however, on which this conjecture is based involved very small quantities of learning material compared to the amounts of instructional material usually transmitted in practical teaching situations. Nonetheless, much instruction is distributed in time not because of learning-theoretic ideas but because of sensible assumptions about the physical endurance of both teachers and students, and for logistic reasons of various kinds. Consequently, widely spaced teaching is commonly practiced, not only because it is a reasonable extension of known experimental facts, but also because it is frequently convenient. The tri-weekly meeting of many college classes is a good illustration of this.

Because massed teaching appears to be often dictated by practical constraints, it seems reasonable to look for ways in which the negative effects of cramming can be alleviated. One interesting possibility is suggested by the conjecture (Number 4 previously mentioned) that one of the causes for the inefficiency of massed teaching is the uniform enviromental context shared by much of the teaching material. If environmental context could be greatly diversified and enriched, the inefficiencies of massed instruction might be reduced. Distinctive environmental stimuli associated with each instructional segment could provide additional addresses in memory. These stimuli, as Smith, Glenberg, and Bjork (1978) and Smith (1979) have suggested can be expected to aid subsequent recall because additional addresses provide extra access paths to the stored information (see Anderson & Bower, 1973).

An important factor to be considered in this connection is that the added context cues should not distract the learners or tax their resources. One environmental variable is a particularly attractive source for such nondistracting context enrichment, namely place. Hasher and Zacks (1979), and Rothkopf, Fisher, and Billington (1982) have reported evidence that location provides automatically processed cues, in the sense suggested by Shiffrin and Schneider (1977), that is, that they are acquired without intention by the learner, and do not make resource demands. Smith et al., (1978) and Smith (1979), have furthermore demonstrated that location provides excellent cues for retrieval of information. They have shown that word lists are recalled better in rooms where they were learned rather than in other familiar rooms, and that recall was aided even when subjects only imagined the room in which learning took place.

Since most of our ideas about the detrimental effects of massing instructional material in time come from experiments that involved a few minutes of learners' time and verbatim repetition of the same material, we wanted to find out whether the distribution of practice effect would hold for realistic situations as well. A short industrial or academic training course, for example, might involve 10 to 100 times more material than that used in laboratory studies, and would, typically, include little repetitive drill.

We also wanted to find out whether inefficiencies caused by massed instruction, if found, could be reduced by conducting lessons in a variety of experimental contexts. The additional contextual information provided by environmental changes was expected to function like the contextual variability caused by long intervals between lessons.

To test the effects of contextual variations, temporal spacing of lessons, type of test, and certain subject factors on the learning and memory of educationally realistic materials, we developed a special 8-hour cram course in introductory statistics. The four lectures of this course, videotaped to provide the same instructional content for all conditions, were presented either within a single day, or spread over a period of 4 days. In order to investigate contextual variation, the lectures occurred either in a single classroom or they were shown in four different room settings. Five days following the last lesson, subjects were given a battery of memory tests covering the material, including General Recall (free recall), Cued Recall, Matching (recognition), and Problems. The final test session occurred in a totally new room so that the test environment would not serve as a differentially useful retrieval cue.

Two additional factors that guided the design and analysis of this experiment should be mentioned. The first was the nature of the tests used to measure achievement. The cue specificity rule proposed by Eich (1980) states that the more specific the cues provided by a memory test, the less will be the effects of experimental contextual manipulations. For this reason, a free recall test was used to provide the main dependent variable. Free recall tests provide only very general, nonspecific cues (e.g., recall all the words from the list you have memorized). Cued recall provides more retrieval information than free recall, and recognition still more. Studies of environmental context (Smith et al., 1978) and of pharmacological context, or drug state (Eich, 1980) have shown that memory's dependence on contextual information is most likely to occur for free recall, and least likely for recognition tests. Performance on the free recall (called General Recall here) was therefore the dependent variable we expected to show the effects of context. We administered this test first in the sequence of memory tests, because we did not wish general recall results to be contaminated by other prior memory tests. A cued recall and a recognition test (matching) followed because we were interested in exploring whether context effects would be diminished by amount and specificity of retrieval information. Interpretation of the cued recall and recognition test would be clouded, however, because of confounding with prior administration of the general recall test.

A second factor explored in the present experiment is the possibility that certain cognitive styles may be more sensitive than others to manipulations of environmental context. Although no previous work on this topic has been reported, it seemed reasonable to suspect that some subjects would be more susceptible to context dependence than others, and we made some guesses about what cognitive style might be relevant to the hypothesized phenomenon. The style we selected was field dependence.

Briefly, field dependence refers to a tendency to perceive a stimulus in relation to the background or field in which the stimulus is embedded. A field-independent subject tends to perceive a stimulus independently of its background. We conjectured that field-dependent subjects might have more context-dependent memories because they attended to background environmental context as well as to the focal stimuli. Field-independent subjects, on the other hand, might learn material independently of the background environment, and might therefore be insensitive to contextual manipulations.

METHOD

Design and subjects. The two main independent variables were context enrichment (1 Room vs. 4 Rooms) and distribution of instruction in time (1 Day vs. 4 Days). These were used to form four experimental groups: (1) 1 Room:1 Day (1R1D), (2) 4 Rooms:1 Day (4R1D), (3) 1 Room:4 Days (1R4D) and (4) 4 Rooms:4 Days (4R4D). Each group consisted of 25 subjects, all paid volunteer Texas A & M undergraduates without previous statistics training. In assigning subjects, an effort was made to match the groups on their scores on a mathematics pretest. Exact matches proved impossible because of constraints of the subjects' academic schedules.

Subjects were ex post facto assigned to either the field-dependent or field-independent group according to their score on the Group Embedded Figures Test (GEFT). Because it was logistically impossible to match groups in advance on this variable, each treatment group was split at its median score on the GEFT. Fortunately the median scores on all four groups were the same as the group median.

Procedure and Material

Pretesting. Pretesting was done 2 to 3 weeks prior to the instructional experiment. This session was used to provide information about the experiment to subjects, and to gather information for matching and scheduling groups of subjects for the study. Everyone was given the Math Pretest, the GEFT, and was asked to supply their grade point average (GPA), major in school, age, and sex. They were also asked to list the times when they could attend the experiments.

The Math Pretest was constructed to test the basic quantitative skills of subjects prior to the lectures. The 50 questions on the Math Pretest covered operations with negative numbers, fractions, square roots, algebra, graph reading, summation rules, and measures of central tendency.

The Group Embedded Figures Test (GEFT) is a standard test for field dependence (Witkin, Oltman, Raskin, & Karp, 1971), with lower scores indicating field dependence and higher scores indicating field independence. The GEFT requires subjects to find and trace simple geometric figures which are embedded in more complex figures. The shading and the distracting patterns in the complex figures are designed to disguise the simple figures within, even though the simple figures are present in the exact same size, shape, and orientation as shown in the samples. After seven warmup problems, subjects were given 18 critical test problems to work in 10 minutes. A maximum of one correct solution was counted towards the subject's total for each problem, even though several problems required multiple solutions. Thus, a perfect score was 18.

The pretests were administered to groups of students in a room which was never again used in the experiment. The room was not in the building in which the experimental lectures took place.

Experimental lessons and problem exercises. All groups of subjects received four 2-hour periods of experimental instructions in statistics. Each 2-hour period consisted of a televised lecture, a set of problem exercises, and a televised segment in which the answers to the exercises were described. The videotaped lectures were shown on a single color TV monitor. For exploratory purposes and to further increase context variety, two television monitors were used for Lessons 2 and 4 in the four-room condition. For the two-monitor arrangement, the lecture would switch back and forth between a color and a black-and-white TV set. Switches occurred at topical breaks in the lecture at approximately 5-minute intervals. Each student was given a notebook for taking notes during the four lessons. The notebooks were kept by the experimenter between sessions and turned over to the student after completion of recall testing. The reason for retaining the notebooks between lessons was to minimize subjects' opportunities to review course material.

The four rooms used for the lessons in the four-room treatments were all in the building which houses the Texas A & M University Psychology Department. The sizes of the rooms varied from small conference rooms to medium-sized classrooms holding 75 to 80 students. The rooms varied along a number of other dimensions, including color of walls, presence of windows, type of desks and seats, illumination, presence of chalkboards, and miscellaneous objects scattered about. In general, the rooms were quite distinct and different from each other. Instruction for the one-room groups took place in one of these four rooms.

To enhance contextual differences between the four environments used for the lessons, the image on the TV screen was also systematically manipulated in the multiple-room condition. In Room 1 the picture was presented with its normal color; in Room 2 the color was set to show unnaturally reddened pictures; in Room 3 the picture was unnaturally green and in Room 4 it was shown in black-and-white. Also, in the 4 Room condition, multiple TVs were used to present material for Lessons 2 and 4. In this situation the video image and sound switched about every 5 minutes (during natural breaks in the lectures) between two television monitors.

The videotaped lectures and problem sessions were prepared specifically for the present experiment, and were not taken from regular course lectures. The lecturer (Dr. Arthur Glenberg of the University of Wisconsin) prepared a set of four condensed lectures, which restricted the material to a finite but coherent set of testable concepts. Although these concepts are typically presented within a broader context of a full-blown statistics course, the minicourse lectures represented a complete and integrated set of information rather than an excerpt cut from a large series of lectures.

The topics covered by the lectures included descriptive statistics (Lessons 1 and 2) and inferential statistics (Lessons 3 and 4). Lesson 1 covered summation rules, scale types, measures of central tendency, and definitions of basic concepts such as *statistic, parameter, sample,* and *population.* Lesson 2 covered measures of variability, bias in parameter estimation, and computational vs. definitional formulas. Lesson 3 included statistical generalization, random sampling, degrees of freedom, and confidence intervals. Lesson 4 dealt with hypothesis testing, independent sample *t*-tests, Type I and II errors, and decision rules.

The videotapes showed Dr. Glenberg lecturing in front of a chalkboard. The camera stayed primarily on the lecturer but occasionally zoomed in for close-ups when there was writing on the board. Each lecture lasted about 90 minutes. At the end of each lecture, a set of problem exercises was administered, derived from the material presented on that lesson. This was done in the following way. At the end of each lecture the videocassette recorder was turned off, and copies of problem sets were handed out. Subjects worked 15 minutes on each problem set, and then they were given red pens to correct their own work. Correcting was done by a final videotaped segment where the lecturer worked through the correct solutions to the test problems. The red pens were used so that subjects would not be able to amend their original penciled answers, because we were interested in their initial performance on those problem sets.

None of the problems used in the end of lecture exercises was repeated on a final test, but the types of questions were the same as those used on the final tests. Question types included matching, computations, word problems, and recall of terms and definitions. The maximum scores possible for the four problem sets were 23 for Problem Set 1, 26 for Set 2, 18 for Set 3, and 24 for Set 4. These scores were not differentiated according to type of question.

Subjects were led to believe (erroneously) that a bonus of $5 was to be given to those subjects who scored in the top 50% on the four problem sets. We believed that this procedure would increase motivation and attention during the lessons. In fact, all subjects were paid the bonus.

Between lessons, subjects were free to go where they wished. For the 1-Day groups, 20 minutes were given between Lessons 1 and 2 and between 3 and 4. Between Lessons 2 and 3, a 45 minute lunch break was given. For the 4-Day groups, lessons were given at 1-day intervals. Subjects were told they would have to attend a posttest session to receive payment, but they were not told that they would be tested on the minicourse lectures.

Posttests. Five days after the fourth lesson, we tested subjects' memory for information taught during the minicourse. Because we did not want to encourage extra-experimental studying, subjects had not been told about the posttest. They came to the posttest session to receive their honorarium. The posttest was conducted in a room different from any other room used during the experiment, located in a building different from that in which the experimental lectures were given.

Four types of tests were given during the final testing session. These were General Recall, Cued Recall, Matching, and Problems, in that order. The order of presentation of the tests corresponded roughly with the amount of information provided by the test; that is, tests given the least amount of retrieval information were presented first.

The General Recall test was analogous to an uncued or free recall test. Subjects were given 10 minutes to outline the lectures thoroughly, listing all of the topics and symbols covered by the lectures, such as "measures of central tendency," "null hypothesis," or "$\hat{S}$." They were told not to define or describe any of the terms, but simply to list as many as they could recall.

A list of symbols (e.g., $\bar{X}$, Ho, S^2), statistical terms (e.g., "variance," "median," "confidence limits"), and topics (e.g., "inferential statistics," "measurement scales," "hypothesis testing") was compiled for scoring purposes. Each item on a subject's General Recall test was counted as one point. The rare cases of questionable responses were scored according to a consensus reached between the experimenter and the first author. The highest score received was 32 correct.

The Cued Recall test (15 minutes) provided retrieval cues for specific information, as compared with the general listing of terms in the General Recall test. It consisted of 23 questions asking for specific terms, symbols, and definitions. Sample questions were: (1) Name four types of measurement scales, (2) Draw the correct symbol for the sample variance, (3) Define "parameter," and (4) What is the term for the middle-ranked score? A total of 15 minutes was given for the 23 questions. A perfect score was 47 points.

On the Cued Recall test one point was given for each correct answer. No partial credit was given. In cases where multiple answers were requested (e.g., "Name three measures of central tendency") one point was given for each answer. A formula was counted as incorrect if any part of the formula was incorrect.

The Matching test (4 minutes) required subjects to match a list of 16 definitions (e.g., "A means of summarizing data") with an equal number of terms and symbols (e.g., "descriptive statistic," "range"). There was only one correct term for each definition.

The Problems test (20 minutes) included computations (e.g., "Calculate the 90% confidence interval for a sample of scores") and short answers (e.g., "Why is random sampling important?") There were 19 questions on this test. Each completely correct answer or computation was given one point. No partial credit was given. As with Cued Recall, the rare cases of ambiguous answers were scored according to a consensus reached between the experimenter and the first author. Answers were quite short, requiring fewer than 10 words apiece.

Following the Problem test, all subjects were paid $30, debriefed, and dismissed.[2]

RESULTS

Lesson Problem Exercises. Scores on the Problem Exercises are shown in Table 1 for the four main treatment groups. Those scores were submitted to an analysis of covariance, using Rooms (2) × Days (2) × GEFT (2) × Problem Set Number (4) as independent variables, and Math Pretest and GPA as covariates. Problem Set Number was a repeated measure (each subject did all four sets of problems) and the others were grouping factors.

---

[2]In evaluating this significance level, the cost of this experimental operation should be considered. It included the development of a special televised statistics course as well as 1½ days out of the lives of 100 student participants.

TABLE 1
Mean Number and Proportion of Correct Responses for the Four Treatments on the Problem Sets Presented at the End of Each Lesson

Problem Source	Treatments							
	1 Room – 1 Day		4 Rooms – 1 Day		1 Room – 4 Days		4 Rooms – 4 Days	
	$\overline{X}$	prop.	$\overline{X}$	prop.	$\overline{X}$	prop.	$\overline{X}$	prop.
Lesson 1	17.6	.76	18.4	.80	17.4	.75	18.4	.80
Lesson 2	19.9	.76	19.3	.74	20.0	.77	18.9	.73
Lesson 3	18.8	.82	14.0	.78	14.4	.80	15.2	.84
Lesson 4	9.8	.70	10.0	.71	10.7	.76	11.0	.78
All Lessons	62.0	.77	61.7	.76	62.4	.77	63.4	.78

The only significant effect in the analysis was found for Problem Set Number, $F(3,276) = 6.77, p < .001$. Problem Set 3 was done best (81% correct), and Set 4 was worst (74%). The problems in the four sets were not matched for difficulty and therefore the problem effect cannot be meaningfully interpreted.

More important for the present experiment were the lack of main effects for Rooms, $F(1,90) = .79$, or for Days, $F(1,90) = 1.46$, as well as the absence of interactions between Problem Set Number and Days, $F(3,276) = 1.84$, or between Problem Set Number and Rooms, $F(3,276) = 1.52$. Had such effects been found, it might have indicated that the treatments could have caused differences in learning (as measured by Problem Set scores) rather than memory (measured by General Recall). This point will be discussed again later.

General Recall. An analysis of covariance was computed for the General Recall scores, using Rooms (2) × Days (2) × GEFT (2) as grouping factors, and Math Pretest and GPA as covariates. Math Pretest and GPA scores were used as covariates because those measures correlated with performance on the experimental tests (Table 2). As can be seen in this table, GEFT scores did not correlate highly with GPA, Math Pretest scores, or other measures of performance.

As shown in Figure 1, the 4-Day groups recalled about 13% more information than the 1-Day treatments. The day effect on General Recall was of borderline significance, $p = .07$[3], $F(1,90) = 3.38$. Room effects were not reliable, $F(1,90) = 1.86$, $p = .18$, although average recall for the 4-Room treatments exceeded the 1-Room groups by about 10%. Field-Dependent students recalled less than field-independent subjects, $F(1,90) = 3.86$, $p = .05$.

The most noteworthy finding was not in the main effects, but in the interaction between context enrichment and field dependence, $F(1,90) = 3.94$,

TABLE 2
Intercorrelations Among Pretest and Selected Performance Test Measures

	1	2	3	4	
1. Mathematics	–				
2. Grade Point Average	.33	–			
3. GEFT	.25	.07	–		
4. Problem Set Total	.69	.46	.27	–	
5. General Recall	.47	.34	.22	.44	–

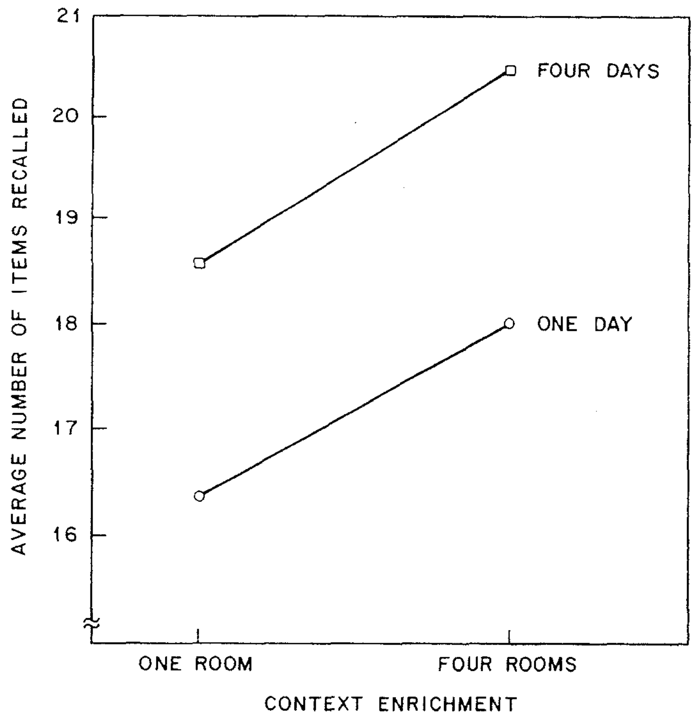

Figure 1 Effects of Rooms and Days upon number recalled on the General Recall test.

$p < .05$. As shown in Figure 2, Field-dependent subjects performed substantially better on the General Recall test when instructed in the diverse context, that is, the multiple room setting. This was found for both the massed and the temporally distributed teaching conditions. Field-independent students, on the other hand, performed about equally well under the single and the multiple room treatments.

Cued recall. An analysis of covariance was computed for the cued recall scores, again using Rooms (2) $\times$ Days (2) $\times$ GEFT (2) as grouping factors, and Math Pretest and GPA as covariates. There was a main effect for Days, $F(1,90) = 7.52, p < .01$, indicating that the mean cued recall of 4-Day groups was higher than for 1-Day groups. The adjusted mean for the 1-Day group was 26.49 items correct, whereas for the 4-Day groups it was 30.22 items, an improvement of 14% over the 1-Day group's performance. The only other significant effect in the analysis was for GEFT, $F(1,90) = 7.70$, $p < .01$, showing that field-independent subjects scored higher (30.37) than field-dependent ones (26.34). Unlike the results of the General Recall test, there were no effects for Rooms, $F(1,90) = .82$, nor for the Rooms $\times$ GEFT interaction, $F(1,90) = .34$.

Matching. The analysis of covariance for Matching test scores again used Rooms (2) × Days (2) × GEFT (2) as grouping variables and GPA and Math Pretest as covariates. The main effect for GEFT was significant, $F(1,90) = 4.31$, $p < .05$, again indicating higher scores for field-independent subjects. The Days × GEFT interaction was also significant, $F(1,90) = 4.04$, $p < .05$, indicating that for field-independent subjects, the 4-Day condition (13.84) yielded better scores than the 1-Day condition (12.81), whereas for field-dependent subjects, 1-Day (12.58) was better than 4-Days (11.58). The reason for this interaction was not clear. No other reliable effects were found. The low performance of field-dependent subjects under the 4-Day condition was primarily due to the unusually weak performance of the 4-Day-1 room treatment on the matching test.

Problems. Scores on the Problems test were submitted to an analysis of covariance, again using Rooms (2) × Days (2) × GEFT (2) as grouping factors, and Math Pretest and GPA as covariates. No reliable effects were found.

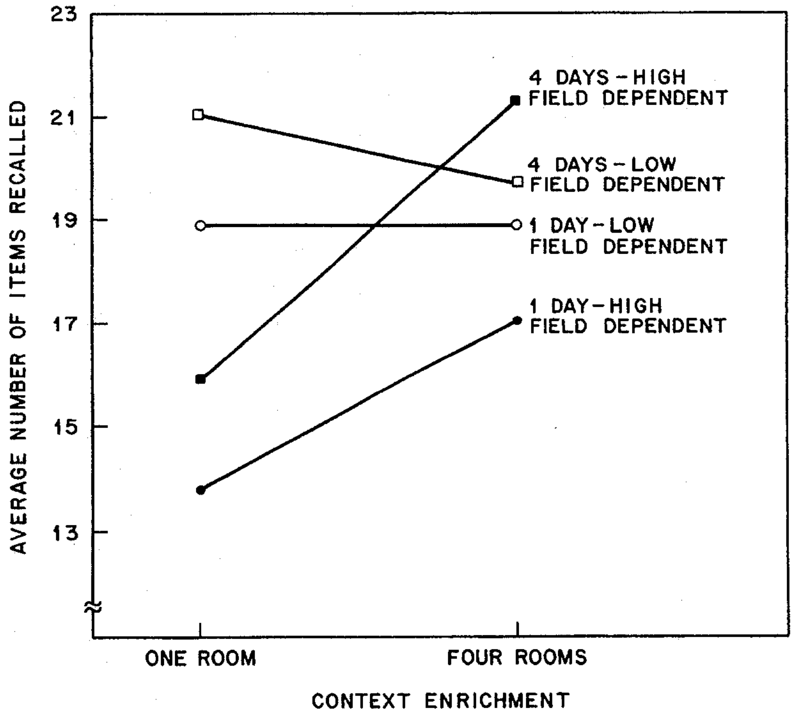

Figure 2 Effects of Days, Rooms, and Field Dependence on the General Recall test: Open circles and squares are field-independent subjects, solid circles and squares are field-dependent subjects.

DISCUSSION

Our experiment confirmed that temporally massed teaching can lead to inefficient learning. We tested this hypothesis with substantial, practical teaching matter delivered under carefully controlled experimental conditions. The study also partially confirmed the theoretical expectation that the detrimental effects of massing instruction in time could be reduced by enriching the teaching environment. However, our results provided important qualifications for the original hypothesis.

For field-dependent subjects contextual enrichment (i.e., use of multiple learning environmental contexts) caused an improvement in General Recall, as compared with recall performance for the group given lessons in only one context. The benefits of contextual enrichment, however, occurred for both the massed and distributed instruction conditions; the lack of an interaction between Days and Rooms (Fig. 1) shows that the Days effect was not influenced by contextual manipulations. Viewed this way, it can be concluded that contextual enrichment did not remove the effects of massed practice since the 1-Day group recalled less than the 4-Day group for both the "enriched" and "plain" context conditions. The theoretical conclusion from this result is that the mechanism that caused the superiority of the distributed over the massed treatments does not depend entirely on the greater contextual diversity of time-distributed teaching. The practical implication of these results is that contextual enrichment should improve recall for any type of course arrangement, whether the classes are massed or distributed in time.

Massed instruction of the minicourse was found to have detrimental effects upon both cued and uncued recall, as compared with presenting course lectures at spaced intervals. This result is in keeping with results of small-scale laboratory studies that have demonstrated that memory for information given massed repetitions is worse than memory for material given spaced presentations (e.g., see Hintzman, 1974 for a review). The finding of massed vs. distributed presentation effects in the present experiment is important for two reasons: (1) Whereas most prior research on spacing of presentations has considered memory for verbatim-repeated events, the present experiment found spacing effects for materials that were not repeated in identical form (although certain concepts, such as mean, were necessarily mentioned in multiple lectures), and (2) The materials used were ecologically realistic, and could be used for educational purposes.

The most challenging qualification was that the effects of context enrichment interacted with perceptual style, that is the field-dependence variable. The recall of field-dependent subjects was reliably improved when they were taught in the multiple-room context. For field-independent subjects, on the other hand, context enrichment produced only small and unrealiable increases in recall.

Field dependence (see Goodenough, 1976 for a review) is considered to be one aspect of a general personality dimension defined by a global (field dependent) vs. an analytical (field independent) orientation to the world. In a perceptual task, the ability to overcome embedding contexts in order to focus on a figure is the definition of field independence, whereas field dependence is the inability to ignore such contextual information. Although field dependence is typically measured with perceptual tasks like Rod-and-Frame tests or Embedded Figures tests, it is clear that the concept of field dependence has important implications for memory processes, as well. No reported studies of the relation between field dependence and memory, however, have examined the interaction of field dependence and contextual memory dependence, as the present study has.

We propose that contextual manipulation affected only field-dependent subjects because of their perceptual styles. Field-independent subjects may have focused on the instructional stimuli alone, ignoring the environmental context in which the stimuli were presented. Field-dependent subjects might not have ignored the context, and therefore contextual information was encoded, causing it to affect memory performance. Alternatively, it could be that differential use of contextual information between field-dependent and -independent subjects occurs at the time of retrieval rather than during encoding. It could be that field-independent subjects used meaningful associations to guide their retrieval of the minicourse material, whereas field-dependent subjects were more likely to use contextual information to guide memory retrieval.[3] These explanations, if correct, suggest: (1) Phenomena of contextual memory dependence might be more efficiently studied if experiments focus upon field-dependent subjects; and (2) Previous findings of contextual memory dependence (e.g., Smith et al., 1978) may have been "watered down" by results from field-independent subjects. Finally, it is also interesting, and perhaps surprising, to note that differences in perceptual styles could have such a dramatic effect on the higher level cognitive processing necessary for learning material in a statistics course.

The present results can be further explained by considering the issue of availability and accessibility (Tulving & Pearlstone, 1966). Information is available in memory if there exists an encoded representation, whereas it is accessible only if that encoded information can be retrieved. Uncued (General) recall of an item requires that the item must be not only available, but

[3]If different individuals encode differing amounts of contextual information, then this casts doubt on the idea that environmental context is automatically encoded, since the Hasher & Zacks (1979) notion of automaticity is not sensitive to individual differences. The retrieval hypothesis is more compatible with the notion that environmental information is privileged or automatically encoded. According to this hypothesis, contextual enrichment could affect retrieval strategies, which are applied to automatically encoded context information.

accessible as well. Tests which provide retrieval cues, such as Matching, Problems, and to some extent Cued Recall, are better indicators of availability than of accessibility because the provided cues help alleviate the need for retrieval processes. In the present study contextual enrichment improved General Recall, but did not affect scores on the initial Problem Exercises, Cued Recall, Matching, or Problems tests. The implication is that contextual enrichment improves accessibility of information rather than availability, since improvements in availability should improve scores on all tests. This improved accessibility is presumably accomplished by the increased organization of material into segments corresponding to multiple environments, even though that organization may not have affected the other performance measures.

Aside from these theoretical issues, an educator may worry about the generality of General Recall results to other educational situations. We believe, however, that uncued retrieval is commonly required under circumstances where learning must be used and applied, ranging from situations where one must synthesize an essay or a speech, to situations when one must martial one's statistical knowledge to analyze a set of data. In the latter case, for example, it may occasionally happen (outside of the classroom examination situation) that one will be asked the meaning of a statistical concept or the method for a statistical procedure (as in the Cued Recall test in the present study), but it will more commonly be the case that implementation of statistical skills begins with a much more general question, such as "How are these data to be analyzed?" This less specifically directed question seems to us to require the same sort of general retrieval processes tapped by the General Recall test.

Although it has already been noted that memory tests that followed the General Recall test were contaminated by the previous tests, the results of those tests are suggestive. The finding that distribution over days improved General Recall and Cued Recall, but not Matching, indicates that either the results of the Matching test were essentially uninterpretable because they are confounded with test sequence, or that the massed vs. distributed presentations affected retrieval, but not recognition processes. Whereas the Days factor influenced both General and Cued Recall performance, the Rooms variable affected only the General Recall. Although the General Recall effects did not quite reach statistical significance, these results suggest two conclusions: (1) The effects caused by temporal manipulations are different from those produced by contextual factors, and (2) The Cue Specificity rule (i.e., memory tests that provide less retrieval information lead to greater memory dependence on environmental context) holds not only for contextual reinstatement phenomena (e.g., Smith et al., 1978), but for contextual enrichment effects as well.

To briefly review the results, we found (1) that the effect of contextual enrichment was a memory (as opposed to learning) phenomenon, (2) that it affected uncued recall of educationally realistic material, (3) that only field-dependent subjects were affected, and (4) that contextual enrichment acted independently of temporal spacing of lessons. An explanation of the results states that during learning (for field-dependent subjects) contextual information is encoded and associated with learned material, forming a potential retrieval cue for the learned information. If recall is uncued, subjects use environmental context information as a self-generated memory cue. In such a case, one single context as a cue would be more likely to fail on retrieval than the ensemble of four contextual cues used in the 4-Room condition. Subjects should recall more with four different contextual retrieval starters than with a single contextual cue. This is similar to the notion that, up to a point, the more categories or divisions into which a list is divided, the better free recall performance will be (Mandler & Pearlstone, 1966).

Mention should be made of some recent experiments by Smith (1982) that have also looked at the effects of multiple learning environments on memory performance. Although the materials used in that study were unrelated word lists, and the experimental sessions lasted less than an hour, it is interesting to note that Smith also found beneficial effects of multiple learning contexts. This finding was limited, however, to situations where subjects were tested in a totally new room. When tested in an old familiar learning environment, number of learning environments had no effect upon recall. This last result has an important implication for educational applications of contextual enrichment procedures. Because much learning is tested in the classroom but used in settings away from the school environment, the results of classroom exams could be misleading. If contextual enrichment is used during lessons, then the resulting memory improvements might not be noted unless testing occurs outside of the original learning environment.

ACKNOWLEDGMENTS

The authors are greatly indebted to a number of people for their comments and assistance during all stages of the project. In particular we would like to thank Art Glenberg for giving the lectures and making comments on the project; Margaret Bradley for the videotaping, collection and analysis of pilot data, and help with establishment of experimental procedures; Dennis Egan, Anthony Bourgeois, and John Riskind for advice concerning our subject variables; M. J. Billington for data analyses; J. B. Francks for conducting experimental sessions and scoring the data, and E. U. Coke for helpful comments on the manuscript.

The experimental work was performed while Steven M. Smith served as a consultant for Bell Laboratories.

REFERENCES

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory.* Washington, DC: Winston.

Bower, G. H. (1972). Stimulus-sampling theory of encoding variability. In A. W. Melton & E. Martin (Eds.), *Coping process in human memory* (pp. 85-124). Washington, DC: Winston.

Ebbinghaus, H. (1885). *Über das Gedächtniss: Untersuchungen zur experimentellen Psychologie.* Leipzig: Duncker & Humblot.

Eich, J. E. (1980). The cue-dependent nature of state-dependent retrieval. *Memory and Cognition, 8,* 157-173.

Goodenough, D. R. (1976). The role of individual differences in field dependence as a factor in learning and memory. *Psychological Bulletin, 83,* 675-694.

Hasher, L. & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108,* 356-388.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp. 77-99). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hull, C. L. (1952). *A behavior system.* New Haven, CT: Yale University.

Landauer, T. K. (1975). Memory without organization: Properties of a model with random storage and unidirectional retrieval. *Cognitive Psychology, 7,* 495-531.

Mandler, G., & Pearlstone, Z. (1966). Free and constrained concept learning and subsequent recall. *Journal of Verbal Learning and Verbal Behavior, 5,* 126-131.

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 596-606.

Ross, B. H., & Landauer, T. K. (1978). Memory for at least one of two items: Test and failure of several theories of spacing effects. *Journal of Verbal Learning and Verbal Behavior, 17,* 669-680.

Rothkopf, E. Z. (1981). A macroscopic model of instruction and purposive learning: An overview. *Instructional Science, 10,* 105-122.

Rothkopf, E. Z., & Coke, E. U. (1966). Variations in phrasing, repetition intervals, and the recall of sentence material. *Journal of Verbal Learning and Verbal Behavior, 5,* 86-91.

Rothkopf, E. Z., Fisher, D. G., & Billington, M. J. (1982). Effects of spatial context during acquisition on the recall of attributive information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 126-138.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127-190.

Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 460-471.

Smith, S. M. (1982). Enhancement of recall using multiple environmental contexts during learning. *Memory and Cognition, 10,* 405-412.

Smith, S. M., Glenberg, A. M., & Bjork, R. A. (1978). Environmental context and human memory. *Memory and Cognition, 6,* 342-353.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 5,* 381-391.

Witkin, H. A., Oltman, P. K., Raskin, E., & Karp, S. A. (1971). *A manual for the Embedded Figures Test.* Palo Alto, CA: Consulting Psychologists Press.