ATTENTIONAL AND NON-ATTENTIONAL EFFECTS OF PREQUESTIONING

By

JAMES C. TAYLOR, B.A.Sc.

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree
Doctor of Philosophy

McMaster University

© Copyright by James C. Taylor, September 1996

ATTENTIONAL AND NON-ATTENTIONAL EFFECTS OF PREQUESTIONING

DOCTOR OF PHILOSOPHY (1996) McMaster University
(Psychology) Hamilton, Ontario

TITLE: Attentional and Non-attentional Effects of Prequestioning

AUTHOR: James C. Taylor, B.A.Sc. (University of Toronto)

SUPERVISOR: Dr. I.M. Begg

NUMBER OF PAGES: viii, 155

ABSTRACT

Educators occasionally present students with prequestions before the students read textual material. Prequestions increase retention of the prequestioned portion of the text. The consensus in the educational psychology literature is that the increased retention is caused by selective displacement of attention from unprequestioned to prequestioned material.

The experiments in this dissertation employ techniques and theories from generation effect research to show that the memorial benefits of prequestioning derive from more complex process changes than mere shifting of attention. Experiment 1 establishes that the materials used in the dissertation (general knowledge statements) support a prequestion effect in a traditional within-subject design. Experiments 2 and 3 employ a between-subjects manipulation of prequestioning. The presence of a prequestion effect within these experiments demonstrates that selective displacement of attention is insufficient to explain the full prequestion effect.

In Experiments 4-6, alternative explanations for the effect of prequestions are explored. Question

recognition and answer recognition are tested in Experiment 4, in order to assess the impact of prequestions on item-specific information. Item-specific effects of prequestioning were strong for the questions, but not the answers. In Experiment 5, the effect of time lag on prequestion effectiveness was examined. No appreciable lag effect was found. Experiment 6 examined the importance of the relationship between a prequestion and a provided answer. The absence of a prequestion effect for meaningless question-answer pairs in this experiment rules out explanations based solely on cue-specific information. Experiment 7 examines the effectiveness of prequestioning in comparison with dual presentation of the to-be-learned material. For the factual prequestions employed, dual presentation provided a much greater benefit in cued recall.

Finally, a proposal is presented to explain the effects of prequestions. The proposal is a supplement to explanations based on selective displacement of attention.

ACKNOWLEDGEMENTS

The completion of this thesis would not have been possible without the input and assistance of several people.

Thanks should be most gratefully extended to Dr. Ian Begg, my thesis supervisor. He has offered tremendous support as a technical critic and advisor throughout this endeavour. The polished form of this document is in part the result of his patient commentary on many drafts.

I would also like to thank the other members of my supervisory committee, Dr. Lee Brooks and Dr. Larry Jacoby, for the time that they spent following the progress of this project. Their willingness to offer commentary and suggestions on this work is very much appreciated.

Ann Anas deserves special mention for the hours she spent running subjects (a task she finds stressful) at McMaster University while I was teaching at another institution. The thesis could not have been finished without you Ann - thank you.

Negotiating the bureaucratic jungle to complete my degree was greatly facilitated by Erie Long in the

Psychology department office. Thanks for taking care of stuff at the far end of the e-mail line.

Finally, there are all the important people in the background. Thanks must go to all my friends and family, who shaped this thesis by shaping me. Thanks especially to my parents, John and Sylvia Taylor; to my sister, Sheryl; and to the woman who was my girlfriend, then my fiancée, and is now my wife (did the thesis really take that long?), Laureen.

TABLE OF CONTENTS

Descriptive Note	ii
Abstract	iii
Acknowledgements	v
Table of Contents	vii
List of Tables	viii
Chapter 1 - Introduction	1
Chapter 2 - Prequestion Literature Review	5
Chapter 3 - Generation Effect Literature Review	25
Chapter 4 - Experiments: Part 1	44
Chapter 5 - Experiments: Part 2	69
Chapter 6 - General Discussion	115
References	132
Appendix A: Further Analysis of Experiment 4 Target Recognition	145
Appendix B: Stimuli for Experiments 6 & 7	149
Appendix C: Experiment 6b	154

LIST OF TABLES

Table 1 - Cued Recall, Experiment 1 53

Table 2 - Cued Recall, Experiment 2 59

Table 3 - Cued Recall, Experiment 3 64

Table 4 - Cue Recognition Ratings, Experiment 4 80

Table 5 - Cued Recall, Experiment 4 83

Table 6 - Target Recognition Ratings, Experiment 4 85

Table 7 - Cued Recall, Experiment 5 92

Table 8 - Cued Recall, Experiment 6 101

Table 9 - Cued Recall, Experiment 7 109

Table 10 - Needham & Begg (1991), Experiment 5 130

Table A1 - Target Recognition Ratings, Experiment 4 146

CHAPTER 1 - INTRODUCTION

Students rarely remember as much from textual material as their instructors would like. Some loss of memory for the detail of a text is probably unavoidable, but it often seems that a lot of material is forgotten that might have been, or should have been, remembered. Because mastery of textual material is a salient demand in modern educational systems, the discovery of ways to maximize the transfer of material from text to student is of tremendous importance.

The retention of textual material by a student is determined in part by the quality of the mental processes that are aroused in the student when he or she is presented with a text. Mental processes that occur while a student reads are not determined by the text alone, but are additionally dependent upon the cognitive state of the student as each passage in the text is encountered. For example, a student reading in a state of deep inebriation will likely retain far less of the read material than the same student would have retained from reading while sober (Parker, Birnbaum, & Noble, 1976). At a more subtle level, a student with appropriate mental concepts in his or her background knowledge will usually gain more from reading a text than a student who lacks appropriate mental concepts (Tierney & Cunningham, 1984).

One method of altering the cognitive state with which a student approaches a text is to present him or her with prequestions about material in the text. Research on prequestions has been conducted since at least the 1960's, as part of an interest in adjunct questions in general (e.g. Rothkopf, 1966; Frase, 1967). The consensus in the educational psychology literature appears to be that prequestions serve to focus a student's attention on prequestion-related material, and that the selective attention results in superior memory for that material (Hamilton, 1985; Hamaker, 1986). At best, such an explanation is incomplete. Although it is obvious that failure to attend to a stimulus will result in poor memory for that stimulus, attending to a stimulus is not sufficient to guarantee memorability. For example, extensive phonemic processing of a stimulus will result in poor memory for that stimulus on tests that rely upon semantic memory (Craik & Lockhart, 1972). Clearly, the nature of the attention that is devoted to a stimulus is as critical as is the fact that attention is devoted at all.

The simplest plausible interpretation of the selective attention hypothesis is that students do "more of the same" with prequestioned material. That is, whatever it is that control subjects do when they are confronted with a text, prequestioned subjects do more of it to the prequestion-related material and less of it to the incidental material. To the extent that the control subjects engage in mental activity that promotes memory, prequestioned subjects will benefit from having done more of it with the prequestion-related material.

The thesis to be advanced here is that the simple attentional explanation of the effect of prequestions is insufficient. That is, any explanation that suggests that the entire memorial benefit derived from prequestioning can be explained in terms of attention displacement from unprequestioned material to prequestioned material is incomplete. Furthermore, evidence will be presented that is inconsistent with several other plausible explanations for the memorial benefits of prequestioning.

The material in the chapters that follow will be presented in order to substantiate the thesis advanced above. Chapter 2 is devoted to a review of the experimental literature on prequestioning. Chapter 3 contains a selective review of the cognitive literature on the generation effect, with an emphasis on material that will be informative for later discussion of the effect of prequestions. Chapter 4 presents a series of experiments that attempts to rule out an explanation of the prequestion effect based on selective attention. Chapter 5 presents further experimental data, in order to explore the limits of the prequestion effect, and also to explore some plausible non-attentional explanations of the effect. A general discussion of the experimental findings will be found in Chapter 6.

CHAPTER 2 - PREQUESTION LITERATURE REVIEW

Educational psychologists generally accept that questioning students about material the students have just read, or about material they are about to read, affects the amount of material that is remembered later (Tierney & Cunningham, 1984). Questions employed for such purposes are called adjunct questions, and research on their use as supplements to text has a long history (e.g. Pressey, 1926). Despite this long history, attempts to provide a theoretical explanation for the benefits of questioning do not appear to have had much impact until the 1960's.

Beginnings: Rothkopf, Frase, and Information Rejection

Research on adjunct questions gained momentum in the 1960's because of the increased use of programmed instruction. Programmed instruction textbooks were designed to require frequent written responses from students. Questions were interspersed with material that was to be learned in order to achieve this goal. For example, the question "The small whale which is popularly called a porpoise in America is really the ________ dolphin." might be inserted shortly before or after the relevant material was presented in the text (example from Rothkopf & Bisbicos, 1967). Theoretical justification for programmed instruction appears to have been derived from Skinner's operant conditioning theory (Skinner, 1958). Within this theoretical framework, the purpose of questioning is to develop and sustain appropriate behaviours during study, although some researchers apparently assumed that the responses to the questions were themselves a requirement for effective learning (Rothkopf, 1963).

As an explanation for the effects of adjunct questions, Rothkopf (1963) proposed that adjunct questions affect subjects' inspection behaviours. As discussed by Rothkopf, inspection behaviours include articulation responses to the text, intonation patterns, and mathemagenic processing. This last term was defined by Rothkopf as follows: "....mathemagenic processing (from Greek *mathemema*, that which is learned; and *genic*, to form, to give birth), includes both rehearsal-like activities and the emission of mediating response chains or problem-solving." (Rothkopf, 1963, pg. 36). Because the conversion of the stimulus object (the text) to the effective stimulus (the psychological consequence of the subject's exposure to the text) depends heavily on the quality of the inspection behaviours applied to the text, improvements in inspection behaviours can have a substantial effect on learning. Inspection behaviours were considered by Rothkopf to be modifiable as free operants, and thus could be encouraged or discouraged through reinforcement or extinction. Reinforcement and extinction of the inspection behaviours were achievable through the strategic use of adjunct questions.

Rothkopf's interest in adjunct questions was focused on demonstrating general facilitative effects of adjunct questions on the quality of student inspection behaviours. Rothkopf (1966) examined the effects of factual prequestions (questions appearing before relevant text) and factual postquestions (questions appearing after relevant text). He found that both prequestions and postquestions substantially increased the recall of directly questioned material. In addition, the presence of postquestions increased recall of facts that were incidental to the material that was directly questioned, though no more than did an exhortation to read slowly and carefully. The presence of prequestions resulted in performance on incidental facts that was equivalent to an unexhorted control group. Rothkopf concluded that the presence of postquestions had both specific and general facilitative effects on inspection behaviours, but that the presence of prequestions had effects that were specific only to the prequestioned text.

Like Rothkopf (1966), Frase (1967) examined the effect of adjunct questions on the retention of material that was directly questioned and on the retention of material that was not questioned. Frase compared predictions from the Mathemagenic Hypothesis (which by then referred to all inspection behaviours, Rothkopf & Bisbicos, 1967) with predictions from the Cybernetic Hypothesis. Frase's interpretation of the Cybernetic Hypothesis was that adjunct questions provide an internalized criterion for selectively attending to text content. When question-relevant text is encountered, appropriate attentional behaviours are elicited. Thus, the Cybernetic hypothesis predicts positive effects of prequestioning, but not of postquestioning.[1]

Contrary to the findings of Rothkopf (1966), Frase (1967) found no effect of prequestions alone on either directly prequestioned material or on incidental material. The failure to detect a direct effect of

---

[1] Later work by Frase (1968a) and others (Sagaria & Di Vesta, 1978) appears to have included higher level control processes as aspects of the Cybernetic Hypothesis. "If he reads the passage and has not found the answer an error signal is generated. The error is negative feedback which is used to control reading behaviours until the student matches criterion." (Frase, 1968a, pg. 320).

prequestioning is unusual, but consistent with the general finding that the effects of multiple choice prequestions, employed by Frase (1967), are substantially weaker than the effects of short-answer prequestions (Anderson & Biddle, 1975; Hamaker, 1986). With respect to the learning of incidental material, Frase suggested that there might be an inhibitory effect of the presence of prequestions. Although neither his own results, nor those of Rothkopf (1966), were statistically significant, both experiments found recall of incidental material to be lower in the prequestion group than in the control group. Frase proposed that prequestions "...tend to limit the general facilitating effects of test-like events since they relate to specific content within the passages." (pg. 270) This conclusion is consistent with Frase's interpretation of the Cybernetic Hypothesis; however, Frase rejected the Cybernetic Hypothesis on the basis of his failure to find a facilitative effect of prequestions on directly questioned material.

Later work by Frase (1968b, 1968c) also suggested that prequestions lead subjects to reject information incidental to the prequestions. Frase (1968b) examined the recall of incidental information that was presented either before or after question-relevant information. His results were that, regardless of whether prequestions or postquestions were used, the retention of incidental information was depressed when it preceded question-relevant information in a paragraph. Frase proposed that subjects who are prequestioned about material at the end of a paragraph pay little attention to incidental information at the beginning of the paragraph. In order to explain why postquestions produced the same result, Frase pointed out that for each subject in his experiment the position of question-relevant material was held constant. Therefore, postquestioned subjects who received question-relevant content at the end of each paragraph could be expected to learn that they would not be asked about material at the start of the paragraph.

Because prequestions appeared to enhance the learning of directly questioned material at the expense of incidental material, Frase (1968c, 1971) sought to increase the amount of material relevant to each prequestion. Frase (1968c) examined the effects of specific, comparative, and general prequestions on the retention of prose material. Subjects were given learning material containing eight propositions, e.g. "Jim is a pilot. He was born in 1921. John is a policeman. He was born in 1930. Jack is a butcher. He was born in 1926. Jeff is an engineer. He was born in 1934." (Frase, 1968c, pg. 198). Specific prequestions addressed only one proposition within the presented material, e.g. "When was Jack born?". Comparative prequestions required attention to two propositions, e.g. "Is Jim older than Jack?". General questions required attention to four propositions, e.g. "When were the men in the paragraph born?" Contrary to his initial predictions, Frase (1968c) found that memory for the propositions in the text was worst in the general question condition. Frase argued that this result occurred because the "general" questions encouraged subjects to focus on only one half of each proposition, i.e. the age but not the name. Subjects consequently failed to encode the full content of each proposition. Unintentionally, Frase (1968c) provided further evidence that prequestions may serve to limit the information that is attended in a passage. A later study by Frase (1971) that employed more carefully constructed prequestions found that inferential prequestions, i.e. those that require combining more than one proposition in a text, may enhance the amount of material retained.

<u>Attention, Retention, and Forgetting-Rates</u>

Boyd (1973) proposed a simple additive model to explain the effects of prequestions and postquestions.

His model employed two operations that he labelled *attention* and *retention*. The attention operation was concerned with the placing of information into some kind of storage, and the retention operation referred to a composite of maintaining storage and maintaining retrievability from storage over time. Boyd proposed that prequestions act by cuing the attention process to focus on the prequestioned material rather than incidental material. Prequestions, according to Boyd's model, have no effect on the retention process. Control subjects were proposed to attend to all information at a level intermediate between the level of attention devoted to prequestion-relevant material and the level of attention devoted to incidental material by the prequestioned subjects. Postquestions, on the other hand, were proposed to act only on the retention process, because selective attentional processing could not occur at the time that the text was read. Thus, postquestions do not increase the amount initially learned from a text, but decrease the rate of forgetting for postquestioned material[2]. In

---

[2] Boyd (1973) appears to have entirely neglected any "forward" benefits of postquestioning, e.g. better mathemagenic processing as per Rothkopf (1963, 1966). This is surprising, given that a fair amount of interest was focused on precisely that aspect of postquestions. Boyd did however propose that post-questioning might slightly increase the retention of incidental material in a paragraph. The paper's weakness in this regard was probably affected by the

the event that both prequestions and postquestions were employed, Boyd predicted benefits to both the attention and retention processes, that would combine in additive fashion.

Boyd's (1973) experiment involved eleven different groups given various combinations of prequestions and/or postquestions. In addition, each group received a final recall test at the end of the experimental session. In order to measure initial levels of attention, Boyd examined performance on the postquestions. Boyd found that postquestions were answered better when the same material had been prequestioned prior to the reading of the relevant paragraph. He argued that this finding supported his selective attention proposal regarding the action of prequestions. When examining the retention hypothesis with respect to postquestioning, Boyd found that the rate of forgetting, as measured by differences between postquestion performance and final test performance, was greater for prequestioned material than for postquestioned material. The reported forgetting rate for prequestioned material appears to be mathematically greater than the forgetting rate for control material

---

fact that Boyd's results did not show any facilitative effect of postquestioning on incidental information. This lack of a general facilitative effect of postquestions was later confirmed in a literature review by Hamaker (1986).

in Boyd' experiment, though no statistical comparisons were presented to address this point. Boyd also substantiated his claim of additivity by comparing predicted post-test scores on the basis of an additive model with obtained post-test scores for six of his experimental conditions. No difference between the predicted and actual scores was significant beyond p=.10.

Despite Boyd's (1973) results, other researchers have failed to show different forgetting rates for prequestioned and postquestioned material. Sanders (1973) tested education undergraduates given prequestions or postquestions. Testing occurred either immediately or after a delay of one week. Although overall performance was lower after a one-week delay, Sanders found no two-way interaction between adjunct question position and time of testing, and no three-way interaction between those two factors and relevance (prequestion-related vs. incidental material). He did, however, find a three-way interaction between question position, time of testing, and student ability. This interaction occurred because poor students tended to show the pattern of results predicted by Boyd's model (i.e. greater loss in prequestion groups), whereas good students tended to show the opposite pattern of results (i.e. greater loss in postquestion groups). No statistical analysis of the separate groups was presented, nor were data on the retention of relevant versus incidental information as a function of the three interacting variables (though the lack of a four-way interaction may be informative). Studies by Boker (1974) and Swenson and Kulhavy (1974) also failed to show an interaction between time of test (immediate vs. delayed 1 week) and question position either alone or with retention type (relevant vs. incidental) as a third factor. The Boker and Swenson and Kulhavy results may be subject to question however, because the time of test was manipulated as a within-subject variable in both studies, and the tests employed were the same for immediate testing and delayed testing.

Further evidence in support of Boyd's (1973) contention that prequestioned and postquestioned material have different decay rates comes from a study by Sagaria and Di Vesta (1978). In the Sagaria and Di Vesta study, subjects who received only prequestions showed greater forgetting between the time of reading and the time of the post-test (at the end of the experimental session) than either subjects who received only postquestions or subjects who received a mix of prequestions and postquestions. As a tentative conclusion, it seems safe to say that if one compares immediate retention (as measured by postquestion performance) with slightly delayed retention (as measured by an immediate post-test) there are signs of different forgetting rates for prequestioned and postquestioned material. However, if one compares slightly delayed retention with substantially delayed retention (as measure by a 1-week delayed post-test) there is no difference in forgetting rates. This pattern of results can probably be attributed to the use of postquestions to measure the immediate retention of prequestioned material. Soliciting an answer immediately following the reading of a paragraph that has been prequestioned likely encourages subjects to adopt a reading criterion of immediate recallability for prequestioned material, encouraging superficial encoding (e.g. rote processing).

The issue of the decay rate for prequestioned material is important because it could provide evidence that prequestioning is beneficial beyond selective attention. In Boyd's additive model, selective attention does not change the decay rate for attended material. Therefore, if prequestioning benefits decay rates, more than just selective attention processes are at work. However, the idea that prequestioning might act in ways other than focusing attention has received little attention.

A Non-Attentional Theory: Curiosity Drive Reduction

Berlyne (1954, 1966) is one of the rare researchers to propose a non-attentional explanation for prequestion effects. Berlyne (1954) proposed that prequestions might act by arousing epistemic curiosity.

> Epistemic curiosity, according to the theory, is aroused by 'thematic probes'(Skinner, 1947, 1953). The clearest examples of these are questions, which are assumed to evoke, among others, drive-producing meaning-responses (r_{mD}). The curiosity-drive-strength is assume to increase with (1) the intensity of the drive-stimulus (s_{mD}) produced by the r_{mD} and (2) the degree of conflict (F) between the symbolic meaning-responses. (pg. 256)

Berlyne (1954) proposed that the learning of new material could be reinforced by reduction of the curiosity drive. The reading of material related to a prequestion could be expected to recall the prequestion to mind, with a concomitant re-arousal of the curiosity drive. Therefore, material that had been prequestioned could benefit from reinforcement through drive-reduction in a way that unprequestioned material could not.

In order to test his theory, Berlyne (1954) conducted an experiment that examined prequestion effects and curiosity. In order to demonstrate the role of the degree of conflict in curiosity, Berlyne created a study list of statements about twelve different animals. Some of the animals were well-known, and others were "exotic". Berlyne proposed that the more familiar animals would have more associations than the less familiar animals, and thus prequestions about familiar animals would be capable of giving rise to more numerous and stronger competing symbolic response-sequences. Thus, the familiar animals would have greater potential for creating curiosity drive strength, and would show greater benefit from prequestioning than the less familiar animals.

Berlyne's (1954) experiment supported his predictions. Prequestioning increased retention of newly learned material, and increased curiosity about the animals in the experiment. Furthermore, prequestioning benefited familiar animals more than it benefited less familiar animals. Berlyne also noted that statements that were identified at study as answering a previous prequestion were more likely to be correctly answered in the final recall test than were statements not recognized as prequestioned. Berlyne explained this result by pointing out that statements that were not recognized at study as answering a prequestion could not reinstate curiosity, therefore the statements could not benefit from curiosity drive reduction.

Later work by Berlyne (1966) examined the impact of forcing subjects to respond to a prequestion and the impact of delayed provision of an answer to a prequestion. Subjects in the experiment were presented with a list of 28 quotations, as well as two or three names per quotation. The subjects were asked to learn which name was correctly associated with each quotation. Half the subjects in the experiment were required to make a guess prior to being told which name was correct; the other subjects were merely informed of the correct name. The manipulation of guessing was factorially combined with a manipulation of when subjects were informed of the correct name for each quotation. Some subjects were told the correct name to associate with the quotation immediately after presentation of the quotation. Other subjects were presented with all 28 quotations first, and then told the correct author for each quotation. Berlyne's results were that, on a final recall test, subjects who were forced to guess outperformed subjects who were not forced to guess. His explanation for this fact rested on an assumption that forcing people to guess intensifies the strengths of the competing responses, creating greater curiosity drive, and hence greater reinforcement through drive reduction when the answer is presented.

Delayed presentation of the author's name resulted in superior memory on the final recall test as well. Berlyne (1966) argued that there might be an "incubation" effect for curiosity, such that material that is delayed (within some upper time limit) will invoke greater curiosity as the delay increases. A potential interpretation not explored by Berlyne is that the subjects who received immediate presentation of the author's name without guessing do not in fact generate any curiosity, because the answer is provided before they have a chance to (or are required to) have any opinion or curiosity about the matter. Thus, the scores for immediate subjects might be depressed selectively in the subjects that did not guess. However, the two-way interaction between delay and guessing was confounded with student ability, so no data were presented on this interpretation. As support for his epistemic curiosity theory, Berlyne pointed out that a measure of extended curiosity correlated ($r=.22$) with recall, and that the extended curiosity scores showed the same pattern of results as the recall scores with respect to guessing and delay, though not significantly.

Since Berlyne's investigations, little interest has been shown in curiosity drive reduction as an explanation for prequestion effects, though some researchers have included debriefing questions about curiosity in their designs (Peeck, 1970). Bull and Dizney (1973), however, combined Berlyne's (1954, 1966) idea of conceptual conflict with the concept of generalized arousal. They argued that questions that emphasized conceptual conflict would result in an aroused state that would enhance learning. If the state of arousal was sufficiently strong to prevent quick habituation to incoming stimuli, benefits would be seen not only on prequestion-relevant material, but also on incidental material. The emphasis in this theory is on arousal as a driving force in learning (probably mediated through attention), rather than on any benefit of curiosity drive reduction. Bull and Dizney compared the effects of prequestioning with high-arousal questions and low-arousal questions on the delayed retention of prequestion-related and incidental material. The results of the experiment were weak, the only statistically significant result under an ANOVA being a difference between the incidental and related conditions. However, the results were in the direction of the predictions of the theory. By a Dunnett's test, overall retention was higher in the high curiosity question group than the control group, and the low curiosity question group did not differ statistically from the control group. A separate analysis of the relevant and incidental scores does not appear to have been conducted, perhaps due to lack of power. However, the high curiosity question group scored mathematically higher on both the related and incidental retention measures than either the low curiosity question group or the control group.

Current Conceptions of Prequestion Effects

As can be seen, the vast majority of theorization about the effects of prequestions has been devoted to the idea that they cause subjects to attend selectively to prequestion-related material. Two major reviews of the adjunct question literature (Hamaker, 1986; Hamilton, 1985) have argued that verbatim prequestions result in attention being focused on prequestion-related material and withdrawn from incidental material[3]. Hamilton (1985) also argues that because the verbatim prequestions can be answered given the normal, superficial processing that subjects apply to text, no change in the nature of the mental processing of the text is to be expected.

Recently, the unwavering focus on attention has been criticized. Pressley, Tanenbaum, McDaniel, and

---

[3] This contrasts somewhat with Hamilton's (1985) views on conceptual level prequestions. Hamilton argues "...for semantic prequestions, the prequestions would not only focus the subject on the relevant prose material, but also induce more than a superficial semantic level of processing of the target material." (pg. 77)

Wood (1990) commented that "Little is known about prequestion effects on processes other than selective attention." (pg. 28) They conducted an experiment in which they contrasted the learning of textual material by subjects who were required to attempt to answer prequestions with subjects who merely read prequestions to check for comprehensibility of the question, and also with subjects who received no prequestions. In an attempt to have subjects attend to both prequestion-related and incidental material while reading a text, subjects in all conditions were asked to judge each paragraph of the text for ease of comprehension. The results of the experiment were that subjects who correctly answered a prequestion outperformed the other subjects on a test of material related to that prequestion. The performance of the subjects who merely read the prequestions for comprehensibility was not significantly different from the unquestioned control subjects. Furthermore, there were no differences among the three groups of subjects with respect to the recall of incidental information. The lack of an inhibitory effect of prequestions on the incidental material suggests that attention was not differentially allocated by the prequestioned subjects, but the lack of a measure of reading time makes this conclusion less certain. Pressley et al. also noted that incorrectly answering a prequestion did not appear to prevent prequestion benefits for related text material. The recall of material that was related to incorrectly answered prequestions was intermediate between the recall of material related to correctly answered prequestions and the recall of incidental material.

Pressley et al. (1990) argued that there must be more to prequestion effects than mere selective attentional processes. They reasoned that a lack of selective attention was demonstrated in their experiment, because there was no inhibition of incidental information. However, robust prequestion effects still emerged. Furthermore, the activation of information in semantic memory appears to be important, because merely reading the prequestions for comprehensibility resulted in little memorial benefit. Pressley et al. suggest that knowledge mobilization aids in the acquisition of new information, but do not suggest a mechanism by which this might function.

In summary, with the exception of the Pressley et al. (1990) paper, and the work by Berlyne (1954, 1966), there has been little investigation of effects of prequestions on anything other than attentional processes.

CHAPTER 3 - GENERATION EFFECT LITERATURE REVIEW

The similarities between the effects of prequestions, documented in the educational psychology literature, and the effects of generation, documented in the cognitive psychology literature, are both striking and informative. This section contains a brief review of the generation effect literature, with an emphasis on those papers relevant to discussion of the effects of prequestions.

First Description

The generation effect was first identified in the cognitive psychology literature in 1978. Three separate papers presented data that demonstrated better memory for material that subjects had generated for themselves than for material that they had only read. Slamecka and Graf (1978) showed that a variety of generation rules, including rhyme (save-c___), category relative (ruby-d___), and synonymy (rapid-f____) resulted in superior memory for generated words, in comparison with read words (e.g. save-cave). Enhanced memory was demonstrated for recognition that the words had been presented, for recall of the words cued by their study list associate (e.g. save-____), and for free recall. Slamecka and Graf found that memorial benefits accrued only to the generated member of a stimulus pair, and not to the member used as a cue for generation. Although they offered some potential explanations for the effect, Slamecka and Graf were unable to present firm conclusions regarding the source of the generation effect.

Another paper that documented the generation effect was presented by Kane and Anderson (1978). Kane and Anderson showed that when subjects were forced to complete a sentence, e.g. "The <u>dove</u> is a symbol of ______.", they demonstrated better memory for that sentence than if they merely read the sentence "The <u>dove</u> is a symbol of peace." The advantage was apparent both when the subject noun (dove) was used as a cue for recall of the object noun (peace), and vice versa. Interestingly, the memorial benefit of generation was of comparable size whether the sentence was highly determinate, as in the above example, or relatively indeterminate, e.g. "The dove appeared when the magician said ______." (peace), provided feedback regarding the correct answer was given subsequent to the generation attempt. The provision of the correct answer was necessary for the indeterminate sentences because subjects almost always failed to generate the chosen completion. Kane and Anderson advanced an explanation of their results based on the levels-of-processing framework (Craik & Lockhart, 1972), arguing that generating the completion of a sentence requires deeper processing of the sentence than does merely reading the completion. Deeper processing, according to levels-of-processing theory, leads to better memory.

Jacoby (1978) presented two experiments in which generating a word using a semantic relatedness rule (e.g. foot-s__e) resulted in higher cued recall than did merely reading the word (foot-sole). In one experiment, Jacoby demonstrated that the generated target (sole) had to be retrieved from long term memory in order to gain increased memorability. If the target was readily retrievable from short term memory, no generation effect occurred. Thus, if subjects were exposed to the word "sole" immediately before receiving the cue "foot-s__e", no memory benefit was observed when they were later tested with the cue "foot-____". However, if "sole" was presented to subjects 20 items in advance of the generation cue, a memory benefit for "sole" was observed. A second experiment focussed on the importance of the ease or difficulty of generation. Jacoby contrasted "easy" generation pairs such as "check-m_ney" with "difficult" generation pairs such as "lance-sp__r". Although there were effects of generation difficulty, they were not strong, and did not appear in all conditions. Jacoby's proposals regarding theoretical explanation of his results covered the levels-of-processing framework, as well as the concepts of arousal and habituation.

Generation of Meaningful and Non-meaningful Products

Following the initial description of the generation effect, a lot of the theoretical interest focussed on the memorial effects of generating non-meaningful units. Graf (1980) compared the generation effect for meaningful sentences such as "The blond girl baked the cake." with the generation effect for anomalous sentences such as "The blond leaflet baked the piano." Subjects were tested for recall of the nouns of a sentence, given the verb of the sentence as a cue. The generation task required the subjects to assemble each sentence from a list of words, using a fixed rule. Graf found a generation effect with the meaningful sentences, but not with the anomalous sentences. However, he did find "generation" effects for both meaningful and anomalous sentences when recognition of individual nouns was used as a criterion test instead of cued recall[4]. Graf proposed that

[4] The word generation appears in quotation marks because, although the sentences were generated by a transposition rule, the individual words were not

generating a meaningful sentence results in greater interword organization of the sentence than does reading the sentence. The requirement that the generated sentence be meaningful in order to achieve this result argues for a semantic basis for the increased organization. However, generation also affects recognizability of the individual words. According to Graf, this results from increased attention to the individual words in the generate condition, which is required whether the final construction is meaningful or not.

McElroy and Slamecka (1982) also investigated the generation of non-meaningful units. They examined the effects of generating nonwords using either a letter transposition rule, or a rhyme rule. In three experiments, no generation effects emerged either under recognition or free recall testing. McElroy and Slamecka argued that the lack of an effect for nonwords implicated semantic memory in the generation effect, though they were unable to explain the exact role of semantic memory. The finding that non-meaningful units do not seem to result in a generation effect was later extended by several authors (Gardiner & Rowley, 1984; Gardiner & Hampton, 1985; Nairne, Pusen, & Widner,

---

generated. Thus, it is difficult to call the result at the word level a true "generation" effect.

1985) to include dissociations between meaningful and meaningless number bigrams (28 vs. 2,8), meaningful and meaningless letter bigrams (ET vs. EP), familiar and unfamiliar noun compounds (cheese cake vs. cheese ketchup), and high frequency and low frequency nouns.

Nairne and Widner (1987) argued that part of the reason that generation effects had not been found with nonword stimuli was that researchers were testing the wrong stimulus units. They demonstrated that when words and nonwords were generated by reversing underlined letters in a stimulus (e.g. VEAHEN -> HEAVEN, ZERPIK -> PERZIK) generation effects could be found by testing subjects' memory for the letters that had been underlined, rather than the entire stimulus. Unlike the generation effect for recognition of the entire stimulus, which was present only for word stimuli, the letter-level effect holds for both word and nonword stimuli. This result is similar to the result published by Graf (1980), discussed above.

The absence of generation effects for intact nonwords was challenged by Johns and Swanson (1988). These authors argued that the absence of a generation effect for nonwords was largely due to the lack of visual exposure that these words received in the generate condition of experiments prior to 1988. That is, whereas nonwords which had been read had received visual exposure at the time of study, nonwords that had been generated were never presented visually to the subjects. This lack of visual exposure of the generated nonwords might mask whatever benefits they enjoyed as a result of generation. Johns and Swanson documented the validity of their argument by demonstrating that generation effects could be obtained for nonword stimuli in recognition memory if the nonword was visually presented after it had been generated. It appears, then, that nonwords may in fact benefit from generation, at least in a recognition memory task.

Around the same time that Johns and Swanson (1988) were arguing that a generation effect existed for nonwords, Gardiner, Gregg, and Hampton (1988) argued that a generation effect existed for low frequency words. Earlier, Nairne, Pusen, and Widner (1985) had stated that no generation effect existed for low frequency words. In four experiments Gardiner et al. refuted this claim, demonstrating generation effects for low frequency words in both recognition memory and in free recall. In two of three experiments that allowed such comparisons, the generation effect in low frequency words was the same size as the generation effect in high frequency words. This result is important because it argues against the associative linkage hypothesis proposed by Nairne et al. (1985). This hypothesis proposed that generation effects could be found only for words that possessed extended associative networks. Because low frequency words were felt to have few associations, no generation effect for those words was predicted under the hypothesis. However, the Gardiner et al. (1988) results showing comparable generation effects for low frequency and high frequency words seem to refute the hypothesis.

Nairne and Widner (1988) countered the argument of Gardiner et al. (1988) by noting that the association between word frequency and familiarity is not a perfect one. Nairne and Widner used a letter-switching generation task to have subjects generate nonwords, low-frequency unfamiliar words (savant, bivouac), and low-frequency familiar words (dinosaur, bladder). Their results using recognition testing showed that neither nonwords nor low-frequency unfamiliar words benefited from generation, whereas low-frequency familiar words did benefit from generation. In order to thwart the argument that subjects did not recognize the low-frequency unfamiliar words as being real words, Nairne and Widner conditionalized their data on lexical decisions made by the subjects during study.[5] However, rather than defend the associative linkage thesis, Nairne and Widner (1988) argued that generation of nonwords or unfamiliar words likely encourages different mental processing on the part of the subject than does generation of familiar words. One possibility, consistent with arguments in Nairne and Widner (1987), is that the generation of nonwords or unfamiliar words may encourage the subject to pay attention to only part of the to-be-tested item, i.e. the switched letters. Alternatively, and consistent with Slamecka and Katsaiti (1987), they mentioned the possibility that familiar words would tend to be selectively rehearsed, to the detriment of unfamiliar words or nonwords in the same list.

A Crisis of Faith: Results from Between-Groups Designs

In 1987, two papers were published that seemed to severely limit the importance of the generation effect. Begg and Snider (1987) and Slamecka and Katsaiti (1987) both documented several experiments that demonstrated no generation effect when target

---

[5] In the experiments by Nairne and Widner (1988) there was no presentation of the generated target, other than what the subject wrote down. Once again, there were visual exposure differences between the generated stimuli and the read stimuli, as had been pointed out by Johns and Swanson (1988).

generation versus target reading was manipulated between study lists rather than within a single study list. Begg and Snider used pairs of unrelated nouns, in which one or both members of a pair had to be generated from a fragment form. Generation effects were found in recognition and recall when generate-versus-read was manipulated within a single list, but not when it was manipulated between lists. Begg and Snider argued that generating some items within a list does not enhance memory for the generated items; it inhibits memory for the read items. They reasoned that subjects adopt a criterion of identifiability while studying a list containing generated items. Therefore read items, which are easily identified, receive only cursory processing. Slamecka and Katsaiti conducted a similar comparison of within-list and between-lists designs using synonym and antonym generation rules and testing with free recall. They too found no generation effect in a between-lists design. Slamecka and Katsaiti argued, on the basis of the failure to find between-lists generation effects, that the generation effect stems from selective displacement of rehearsal from the read items in a list to the generated items. In their words, "Peter is robbed to pay Paul." (pg. 600) When subjects were prevented from selectively rehearsing generated items, no generation effect in free recall was found for a within-list design.

Of the two explanations for the generation effect proposed above, the selective displaced rehearsal hypothesis has fared less well. Watkins and Sechler (1988) demonstrated generation effects in free recall with an incidental memory procedure. In two experiments, subjects in incidental conditions were led to believe that the word-generation task was merely a distractor task in a pictorial stimuli experiment. They were not warned of an upcoming test of memory for the word stimuli, nor did they claim to have anticipated a test when interviewed at the end of the experiment. Robust generation effects were found in both experiments. Watkins and Sechler argued that the finding of a generation effect under incidental memory conditions was inconsistent with the selective displaced rehearsal concept proposed by Slamecka and Katsaiti (1987), as no rehearsal was likely to have taken place, either of generated or read stimuli.

Revival: Item-specific and Relational Information

Despite the fact that two preceding papers have concluded that there is no generation effect when a between-lists design is employed, at least with intentional memorization, later research has shown this to be too extreme a claim. Hirshman and Bjork (1988) found generation effects in both cued and free recall using a between-subjects, intentional memory design. However, they also demonstrated that the effect was larger in a within-list design. Thus, accurate estimation of the generation effect seems to require that generate-versus-read be manipulated as a between-subjects (or at least between-lists) variable. Hirshman and Bjork, in the same paper, proposed a two-factor theory to explain generation effects. They argued that the size of the generation effect in cued recall and free recall, and the way in which it responds to a 48 hour delay, mathematically precludes a single-factor item-specific explanation of the generation effect[6,7]. Therefore, in addition to item-specific benefits, Hirshman and Bjork proposed that

---

[6] For the details of the argument, the reader is referred to the original paper.

[7] The term item-specific is used in the sense described by Hunt and Einstein (1981). Hunt and Einstein defined two general classes of information: relational information and item-specific information. Relational information refers to the features that are shared by all elements of a to-be-remembered event, and item-specific information refers to features unique to each element the event. Thus, in the context of the generation effect, if one has only two study items, ANIMAL-D_G and ANIMAL-C_W, one can talk about item-specific information for DOG, which would be the encoded features unique to the stimulus DOG. Also, one can refer to cue-target relational information for ANIMAL-DOG, which would be the encoded features that are shared by both ANIMAL and DOG. Finally, one can also discuss whole-list relational information, which would be the encoded features that are shared by ANIMAL, DOG, and COW.

relational benefits accrue to generated pairs. Similar ideas had been raised by a few earlier authors as well (e.g. Donaldson & Bass, 1980; Graf, 1980).

McDaniel, Waddill, and Einstein (1988) extended Hirshman and Bjork's explanation to a three-factor account by including whole-list relational information within their explanatory framework. McDaniel et al. used free recall performance (which is sensitive to whole-list relational information) and a measure of free recall clustering to show changes to the level of whole-list relational processing as a result of target generation. Their experiments showed that memory for whole-list relational information benefited from generation, at least under conditions in which that information was both salient and informative for the target generation process. It should be pointed out that subjects in McDaniel et al.'s experiments were placed in incidental memory conditions, which probably minimized the active processing of the stimuli by the students in the read conditions. Since 1989, the focus of interest in the generation effect has been on the roles played by item-specific information, cue-target relational information, and whole-list relational information. Begg, Snider, Foley, and Goddard (1989) took the position that generating a word makes the word distinctive in a way that reading does not. By their account, generation within the context of an associate and fragment (e.g. Animal - H_R_E) involves two tasks: production and discrimination. The associate 'Animal' recruits several candidates including DOG, MOOSE, PLATYPUS and HORSE, whereas the fragment H_R_E results in discriminations that end with the selection of HORSE as the generated word. As a consequence of this generative activity, HORSE will be memorially distinct whenever it must be discriminated from the same set of alternatives from which it was discriminated at first. Thus, HORSE will show a memorial benefit under recognition memory (which is sensitive to item-specific information, Hunt & Einstein, 1981) and also under cued recall with the cue Animal, because the discriminations that occurred at study will be useful at test. However, recall of HORSE will not benefit from generation if it is cued with the extra-list cue "mode of transportation", because distinguishing horse from car, truck, etc. requires a different set of discriminations than are provided by the study episode. Begg et al. also argued that generation is not sufficient to increase cue-target relational information, though they conceded that "generating sometimes calls attention to and encodes existing relations among words." (pg. 988). In support of the above arguments, Begg et al. conducted four experiments that showed repeated advantages for generation targets in cued recall and recognition. However, there was no advantage in recall of the generative context given the generation target as cue (i.e. recall of Animal given HORSE as a cue), nor was there an advantage of generating the cue word (e.g. AN_M_L - HORSE) on the recall of HORSE given ANIMAL as a cue. Because of these non-results, Begg et al. concluded that there was no evidence of cue-target relational benefits in their experiments. Thus, generation is not sufficient to improve cue-target relational information.

Support for Begg's claim that cue-target relational information need not benefit from a demand to generate comes from work by Schmidt and Cherry (1989). In four between-subjects experiments, Schmidt and Cherry found negative consequences of generating on free recall of pairs, and also on cued recall of generated targets.[8] By analyzing the number of

---

[8] The finding of a negative effect of generation in cued recall of the target is unusual. However, it only occurred in one experiment under atypical conditions. The instructions used by Schmidt and Cherry (1989) for both read and generate groups were to "concentrate on how the words in each pair are conceptually related or associated with each other." (pg 361) Donaldson and Bass (1980) had earlier shown that the generation effect can disappear when the reading group is instructed to attend to the relation between the cue and target. Furthermore, the study materials were presented on a single sheet of paper, with a block allotment of time for either reading and

recalled pairs, the number of recalled generation cues, and the number of recalled generation targets when subjects attempted free recall of pairs, Schmidt and Cherry showed that generation of the target during study actually decreased the cue-target association with their materials (moderate associates, e.g. CITY-COUNTRY, DOCTOR-LAWYER). Furthermore, whole-list relational information also decreased when subjects generated words. Evidence for that trend came from an experiment in which clustering of recall into categories was examined. The only kind of information that benefited from generation within Schmidt and Cherry's experiments was item-specific information about the generation target, as measured by superior recognition of generated versus read targets.

McDaniel and Waddill (1990), and McDaniel, Riegler, and Waddill (1990) defended the three-factor theory of the generation effect against Begg et al.'s (1989) single factor proposal. McDaniel and Waddill (1990) demonstrated that generation of targets in the presence of related context words can improve cued recall of the context words given the target words as cues. This finding may be taken to support the

---

studying or generating and studying. Subject pacing differences may account for the negative finding in cued recall. These shortcomings were later remedied, but only for the pair free recall test.

existence of cue-target relational benefits of generation, and is in contrast to the results of Begg et al. who showed no effect on related context words. The major difference between the two experiments lies in the fact that the subjects of McDaniel and Waddill (1990) were not informed that their memory would be tested. It seems fair to conclude that generating targets in the context of related words forces more attention to the relationship between context and target than might be provided by cursory reading, but not more than is provided by reading with intention to memorize.

McDaniel, Riegler, and Waddill (1990) published results that support the idea that memory for whole-list information that is salient and that facilitates the completion of generation fragments can be enhanced by generation. The evidence for this hypothesis is best demonstrated in their second and third experiments. In these experiments, subjects were presented with stimuli in semantically related blocks, with a critical member of the block either sharing the semantic relationship with the other block members, or not sharing in the semantic relationship enjoyed by the other block members. The only stimuli that showed a generation effect in free recall were critical target words that occurred last (rather than first) in their block, and that shared in the semantic relationship of the other block members. McDaniel et al. (1990) argued that this indicated that whole-list relational information (in the form of block-relationship information) was not fully developed at the start of each new block, nor could it be of use to the last member of the block if that member did not share semantic associations with the other block members. Furthermore, the pattern of benefits in recognition memory was different from the free recall pattern, indicating that it was unlikely that item-specific benefits to the generated targets were mediating the effect in free recall. Once again, this effect was achieved under incidental conditions, so it is not possible to compare the results with subjects whose intention was to memorize the material presented.

Recent papers have focussed on the impact of generation on item-specific or relational information. A number papers have been published that demonstrate increases in pair-relational information as a result of generation (e.g. Burns, 1990), or costs to whole-list-relational information as a result of generation (Burns, 1992; Nairne, Riegler & Serra, 1991; Serra & Nairne, 1993).

Summary

Perhaps the most reasonable capture of what is now known about the effects of generation was written by McDaniel, Waddill, and Einstein (1988). They stated

> "...generation will enhance processing of whatever information subjects are able to utilize to accomplish the generation task, and generation effects will occur to the extent that the memory test requires that information." (pg. 534-5)

With the perspective provided by this statement it can be seen that the explanation provided by Begg et al. (1989) is essentially correct. The only information that results, *of necessity*, from generation is the record of the discriminations used to arrive at the target. However, it is also true that under many circumstances subjects will strategically attend to additional information, particulary in circumstances in which that information will facilitate the task of generating the required target. This information may then become part of the memorial record of the study event. Generation is not sufficient to insure that processing of cue-target relational information or whole-list relational information will occur, but the demand to generate may make such processing more likely. The remaining task in the area would appear to be to determine what conditions reliably lead to processing of information beyond the item-specific information about the generation target.

CHAPTER 4 - EXPERIMENTS: PART 1

Looked at fairly simply, prequestioning and generating are remarkably similar procedures. In both experimental paradigms subjects are confronted with a stimulus that demands some kind of completion, and their memory is later tested for the provided completion. Despite this similarity, there has been little mention of either literature within the other.

There are three major differences between the generation effect and the prequestion effect that are easily discerned. Two of these are procedural. First, prequestioning, in most experiments, is arranged in such a way that there is a time lag between the presentation of the prequestion and the presentation of the material that answers the prequestion. This time lag can be quite lengthy in some cases, spanning several pages of text. Generation, on the other hand, results in the subject encountering the to-be-remembered material immediately subsequent to the presentation of the generation cue. If the subject fails in the immediate generation of the answer, the item is either discarded from further analysis, or feedback regarding the correct answer is immediately provided.

The second procedural difference between research on the generation effect and research on prequestion effects is in the nature of the materials used. Prequestions, virtually without exception, are semantic in nature, requiring subjects to attend to and remember semantically connected questions and answers. Generation effect research, on the other hand, often employs non-semantic generation tasks such as letter-switching or rhyming. The use of non-semantic materials and generative rules in the generation effect literature is certainly justifiable, but it requires that one be cautious about the comparisons made between the two literatures. Happily, there are many studies of the generation effect that have used semantically meaningful materials and generative rules.

The third, and perhaps most important, difference between the generation effect and the prequestion effect lies in the kind of theoretical explanation offered for the effect. As has been reviewed earlier (see Chapter 2), virtually all explanations of the prequestion effect rest on differences between attention to prequestioned material and attention to unprequestioned material. Similar explanations arose for the generation effect in the late 1980's (Begg & Snider, 1987; Slamecka & Katsaiti, 1987), but it has since been recognized that attentional explanations of the generation effect are insufficient. The recent emphasis in the generation effect literature has been in understanding how the demand to generate affects the quality of processing devoted to item-specific and relational types of information (e.g. Begg, Snider, Foley, & Goddard, 1989; McDaniel, Waddill, & Einstein, 1988)

The finding that selective attention is, by itself, an inadequate explanation for the generation effect raises the possibility that selective attention may be inadequate to explain the prequestion effect as well. In fact, recent evidence tends to support this view (Pressley, Tanenbaum, McDaniel, & Wood, 1990), though not conclusively because of methodological difficulties. In order to establish that prequestion effects result from more complex process changes than a mere shifting of attention from unprequestioned material to prequestioned material, I decided to adopt the methods used by generation effect researchers to establish the same point.

As has been reviewed in Chapter 3, generation effect researchers demonstrated the influence of selective attention by switching from within-list to between-lists manipulation of generate versus read (Begg & Snider, 1987; Hirshman & Bjork, 1988; Slamecka

& Katsaiti, 1987). The consequent reduction in the size of the generation effect was substantial; so substantial that it was believed for a time that the effect had disappeared altogether. However, subsequent work proved this belief to be unfounded, and theorists began to examine other explanations for the generation effect. Nonetheless, following 1987 there has been fairly widespread recognition that within-list designs can tremendously inflate the size of the generation effect, through the selective allocation of processing within a mixed list of generated and read items.

In order to demonstrate the insufficiency of selective attentional explanations of the prequestion effect, it is necessary to conduct an experiment in which prequestioning is manipulated as a between-lists variable. However, for this to be achieved, subjects in a prequestion condition must be prequestioned on all the material that they will study. Such a procedure will prevent subjects from shifting their attention from unprequestioned material to prequestioned material at study. Therefore, if a prequestion effect is observed under such conditions, a selective attention explanation will be insufficient to account for it.

However, before seeking to establish whether or not a prequestion effect can be achieved under the conditions of a between-lists design, it should first be shown that the materials and procedures to be used can provide a prequestion effect under the usual within-list design.

EXPERIMENT 1

Subjects in Experiment 1 were first given a list of 32 general information questions for which they provided learnability ratings. For each question, subjects were required to estimate how likely they would be to be able to produce the answer to the question on a final recall test, if they were given a chance to study the answer between the rating and the test. These ratings served as prequestions on the material to be studied[9]. Following the rating phase, the subjects studied 32 general information statements. Sixteen of the studied statements were answers to questions that the subjects had previously rated, and the other 16 statements were new. In the final recall test, subjects were given 64 questions to answer: 16 tested new material, 16 were identical to prequestions to which the answers had not been provided during the

---

[9] Learnability ratings were used instead of explicit prequestions. The use of learnability ratings was considered adequate because in order to make a learnability rating you first have to determine whether or not you already know the answer. Recent work by Koriat (1995; Koriat & Goldsmith, 1994) indicates that feeling-of-knowing judgments are based on the results of memorial search processes that are responsive to the memory cue.

study phase, 16 were identical to prequestions to which answers had been provided, and 16 were questions that had not been rated, but to which the answer had been provided during the study phase. To the extent that the studied answers to the rated questions are better recalled than the studied answers to the unrated questions, there is a prequestion effect.

Method

Subjects. The subjects were 104 students from an introductory psychology course at McMaster University. They received course credit for their participation in the experiment. The experiment was conducted in group sessions of 11-14 subjects.

Materials. The materials used in Experiment 1 were derived from 64 general knowledge statements selected from the Nelson and Narens (1980) norms. These were divided into a "common knowledge" set and a "rare knowledge" set. The items in the "common knowledge" set were answered by .25 to .61 of the norming sample. The items in the "rare knowledge" set were answered by .004 to .19 of the norming sample.

Each of the 64 statements was rewritten so that it could be turned into a fill-in-the-blank question through the deletion of the final word of the statement, e.g. "The name of Batman's butler is A________". The first letter of the final word was always presented when the rest of the word was deleted. In this paper, "statement" refers to sentences shown with the final word present; "question" refers to sentences shown with the final word absent; "item" is used as a generic reference.

The pool of 64 items was divided into four sets, with eight common and eight rare items in each set. The average recall probabilities of the common items in the four sets were .442, .441, .443, and .442. The average recall probabilities of the rare items in the four sets were .065, .064, .063, and .063. The overall average recall probabilities for the four sets were .254, .253, .253, and .253, based on the Nelson and Narens (1980) norms.

For each group of subjects, the four item sets were assigned to four experimental histories. The histories represented the factorial combination of rating (rated or not rated for learnability) and study (studied or unstudied in statement form). All items were tested on the recall test. Thus, on the recall test, one set of items was new, one set had been rated for learnability but not studied, one set had been studied but not rated for learnability, and one set had been both studied and rated for learnability. The item sets were assigned to each history approximately equally often across subjects.

Learnability ratings were collected in booklets. The items were presented in question form, in order to prevent response learning during the rating phase. For each group of subjects, the two sets of items that were to be rated were arranged in a booklet containing four blocks of eight questions. Each block consisted of four randomly ordered questions (two common, two rare) from each of the two item sets. Below each question was a scale from 0 to 100 marked with numerals at intervals of 20.

The study lists were 32 items long, and were displayed one item at a time on a television monitor. Items appeared in statement form during the study phase. Sixteen of the studied items for each group had been previously rated for learnability, and the other 16 were new. The order of items in the study list was randomized with the same constraints as the order of items in the learnability rating booklets.

The recall test was conducted with booklets. All 64 items were used in the recall test, appearing in question form. Each booklet was composed of eight blocks of eight questions, with each block containing two questions (one common, one rare) from each of the four item sets. The order of items within each block was random.

Procedure. Experiment 1 was conducted in four phases:

an initial rating phase, a study phase, a second rating phase, and a test phase.

In the initial rating phase, subjects were given a learnability rating booklet and were asked to make learnability ratings for each of the 32 questions in the booklet. More precisely, they were asked to estimate, for each question, how likely they were to be able to answer that question on a final recall test, if they were given a chance to study the answer before being tested. The initial rating phase was subject-paced, and required approximately 5 minutes to complete.

The study phase immediately followed the initial rating phase, and consisted of a video-taped presentation of each of the 32 statements in the study list. Statements were presented one at a time on a television screen. Each statement appeared on the screen for 5 seconds, followed by a 1-second blank interval. Thus, the entire study display required slightly over 3 minutes to present.

Prior to the recall phase, subjects were asked to make learnability ratings once again, this time for all 64 items. This was done in order to investigate hindsight effects on learnability ratings. The results are not important for the current purpose, so the second rating phase of the experiment will not be mentioned again. The second rating phase took approximately 7 minutes to complete.

Finally, the subjects were given the recall test. The subjects were given recall booklets and instructed to try to answer as many of the questions as they could, whether they had seen the answer in the study phase of the experiment or not. The cued recall task was self-paced, and required about 8 minutes to complete.

Results

Responses in all experiments of this thesis were scored as correct if they could be interpreted as phonetically equivalent to the correct response, or if they were sloppily written but interpretable as the correct spelling. Analysis of the initial ratings will not be presented, as the accuracy of the ratings is peripheral to the current interest. The results of Experiment 1 are shown in Table 1.

Table 1 - Cued Recall, Experiment 1

	Studied	Unstudied
Rated	.64	.22
Unrated	.47	.23
PreQ Effect	.17	-.01

Recall scores were submitted to a two-factor within-subject analysis of variance, with rating (rated or unrated) and study (studied or unstudied) as the two factors.

The analysis revealed main effects of both rating and study. As expected, studied items were recalled better than unstudied items [.56 vs. .22, $F(1,102)=599.00$, $MSe=0.02$, $p<.001$]. Furthermore, rated items were recalled better than unrated items [.43 vs. .35, $F(1,102)=54.96$, $MSe=0.01$, $p<.001$]. The main effect of rating was centred completely in the studied items: recall of the rated study items (.64) exceeded recall of the unrated study items (.47), whereas the recall of unstudied items did not change with rating (.22 rated vs. .23 unrated). Consequently, the two-way interaction of rating and study was significant [$F(1,102)=80.72$, $MSe=0.01$, $p<.001$].

Discussion

The results of Experiment 1 establish that a prequestion effect in cued recall can be achieved with the materials and procedures employed above. There was a 17% cued recall advantage of prequestioned material over unprequestioned material when prequestioning was manipulated as a within-list variable.

The results above, in addition to demonstrating a strong prequestion effect, argue against explanations of the prequestion effect that are linked to the concept of hypermnesia. Hypermnesia refers to increases in recall that are the result of multiple attempts to retrieve material from a once-studied list (Erdelyi & Becker, 1974). Because the materials used were general knowledge statements, it is conceivable that the multiple retrieval attempts made on the material, first at the time of rating and again at the time of testing, created conditions that favoured the occurrence of hypermnesia. The fact that the recallability of unstudied items did not change as a function of prequestioning argues against this hypothesis.

Having established that the materials and prequestioning technique used in the within-list experiment are capable of producing a prequestion effect, it is important to establish that the effect can be produced in a between-lists design as well. By doing so, I will have eliminated explanations of the prequestion effect that are strictly based on selective displacement of attention from unprequestioned material to prequestioned material.

EXPERIMENT 2

In Experiment 2, prequestioning was manipulated between subjects. Half the subjects completed learnability ratings for 32 questions, were given a chance to study statements that answered those 32 questions, and were finally given a recall test covering the studied items and an additional 32 new items. The other half of the subjects received only the 32-statement study phase and the 64-item recall test.

As noted earlier, the procedure described above prevents the prequestioned subjects from selectively displacing attention from unprequestioned items to prequestioned items. For the prequestioned subjects, there are no unprequestioned items presented at study. Therefore any observed prequestion effect in Experiment 2 cannot be explained in terms of selective displacement of attention.

In addition to examining the effects of a between-subjects design, the common and rare items in the experiment were analyzed separately. The reason for the separate analysis was a concern that the effect of the ratings might be limited to items that are common knowledge. Such a result could imply that the benefits of prequestioning are limited to already known, but momentarily inaccessible, facts. This would greatly limit the educational value of prequestioning; one is presumably more often concerned with teaching students new material than reminding them of material they already know.

Although the educational literature on prequestions seems to indicate that the memorability of new material is affected by prequestions (if only through selective attention) there are a few results elsewhere that urge caution. The research on the generation effect has raised concerns that the effect may not extend to meaningless stimuli (see Chapter 3 for a review) or unrelated word pairs (Begg & Snider, 1987). To the extent that rarely known facts represent novel associations of previously unrelated words, they may not benefit from prequestioning. In addition, a study by Begg, Martin, and Needham (1992) showed that with unrelated cue-target word pairs, interposing a cue-review between an initial cue-target study phase and a second cue-target study phase did not enhance learning during the second study phase, as measured in a final cued recall test. It therefore remains an open question whether previously unknown material will benefit from prequestioning when selective attention factors are not allowed to operate.

Method

Subjects. The subjects were 53 students from an introductory psychology course at McMaster University. They received course credit for their participation in the experiment. The experiment was conducted in group sessions of 13-14 subjects. Twenty-six subjects participated in the rating condition; twenty-seven participated in the no rating condition.

Materials. The four sets of 16 general knowledge items used in Experiment 1 were also used in Experiment 2. For each group of subjects, two item sets were used as studied and tested items. The other two item sets were used as new items on the recall test. The four items sets were assigned to each condition approximately equally often across subjects.

In Experiment 2, half the subjects were asked to provide learnability ratings. These subjects were required to rate each of the items that they would later encounter in the study phase. Thus, a booklet containing the appropriate set of 32 items, in question form, was provided to each of these subjects. The booklets were constructed using the same blocked randomization procedure as in Experiment 1.

The materials for the study phase and the recall phase of Experiment 2 were the same as those used in Experiment 1.

Procedure. Experiment 2 was combined with a word-pair learning experiment to fill an experimental hour. Following completion of the word-pair experiment, the subjects in the rating condition were given the booklets to be used for learnability ratings. Subjects in the no rating condition proceeded directly to the study phase of the experiment. The procedure of Experiment 2 was identical to that of Experiment 1 except in the following three ways: (1) half the subjects gave learnability ratings for every item they studied; (2) the other half of the subjects did not give learnability ratings for any item; and (3) hindsight learnability ratings were not collected from any of the subjects.

Results

The results of Experiment 2 are shown in Table 2.

Table 2 - Cued Recall, Experiment 2

	COMMON ITEMS		RARE ITEMS	
GROUP	Studied	Unstudied	Studied	Unstudied
Rating	.74	.37	.62	.07
No Rating	.72	.39	.47	.06
PreQ Effect	.02	-.02	.15	.01

Recall scores were submitted to a three-factor analysis of variance with study (studied or unstudied items) and difficulty (common or rare items) as within-subject factors, and rating (rating or no rating) as a between-subjects factor.

The analysis revealed main effects of study and difficulty, but not of rating. As expected, studied items were recalled better than unstudied items [.64 vs. .22, $F(1,51)=380.67$, $MSe=.02$, $p<.001$] and common items were recalled better than rare items [.55 vs. .30, $F(1,51)=851.17$, $MSe=.02$, $p<.001$]. Overall, there was no difference in recall between the subjects in the rating condition and the subjects in the no rating condition [.45 vs. .41, $F(1,51)<1$, $MSe=.09$, ns].

The main effects described above were qualified by multi-factor interactions. As in Experiment 1, rating interacted with study [$F(1,51)=4.55$, $MSe=.02$, $p<.05$]. In addition, there was a significant interaction between study and difficulty [$F(1,51)=11.27$, $MSe=.02$, $p<.002$] and a marginal interaction between rating and difficulty [$F(1,51)=3.26$, $MSe=.02$, $p<.08$]. The three-way interaction, however, failed to reach significance [$F(1,51)=1.33$, $MSe=.02$, ns].

In order to investigate the data in more detail, the studied items were analyzed separately from the unstudied items. An analysis of variance conducted on the unstudied items revealed a significant effect of difficulty but no effect of rating. Common items were remembered better than rare items [.38 vs. .06, $F(1,51)=193.46$, $MSe=.01$, $p<.001$], but subjects who had initially produced learnability ratings recalled no more new items than did subjects who did not produce learnability ratings [.22 vs. .22, $F(1,51)<1$, $MSe=.01$, ns]. Furthermore, the two factors did not interact [$F(1,51)<1$, $MSe=.01$, ns].

An analysis of variance conducted on the studied items revealed a significant effect of difficulty, but only a weak effect of rating. Common items were remembered better than rare items [.73 vs. .54, $F(1,51)=34.02$, $MSe=.03$, $p<.001$], but rated items were only slightly better remembered than unrated items [.68 vs. .59, $F(1,51)=2.40$, $MSe=.08$, $p<.125$]. Further, there was a marginal interaction between difficulty and rating [$F(1,51)=3.32$, $MSe=.03$, $p<.075$]. Separate analysis of the studied common and studied rare items revealed a significant effect of rating in the rare items [rated:.62 vs. unrated:.47, $F(1,51)=4.71$, $MSe=.06$, $p<.05$] but not in the common items [rated:.74 vs. unrated:.72, $F(1,51)<1$, $MSe=.05$, ns].

Discussion

The results of Experiment 2 appear to suggest that at least part of the effect seen in Experiment 1 was due to selective attention at study to items that had been previously rated. The influence of the learnability ratings on the effectiveness of study is weaker in Experiment 2 (9%) than in Experiment 1 (17%), and appears to be present only in the rare knowledge items. However in those items, it was a large effect (15%), allaying concern that the effect of prequestions might be confined to previously known material. If anything, the reverse appears to be true.

The results of Experiment 2 allow the conclusion that prequestion effects in between-lists designs may not be as robust as they are in within-list designs, but they are nonetheless present. Therefore, prequestion effects in cued recall cannot be explained merely by arguing that attention is displaced from unprequestioned material to prequestioned material.

In Experiments 1 and 2, a stimulus onset asynchrony of 6 seconds was used at study. For some items, this meant that very little study time was left following the reading of the sentence (e.g. "The name of the author who received a Pulitzer Prize for his writings about Abraham Lincoln is Sandberg.") It has been known for some time that text is read more quickly the second time it is encountered than it is the first time (Kolers, 1975). It is possible that, because of re-reading effects, items rated for learnability were read more quickly at the time of study than items not rated for learnability. If so, subjects in the prequestion condition would have had more time to memorize the items they encountered than the control subjects would have had. This could easily lead to superior memory in the prequestion subjects.

The re-reading hypothesis is important to rule out, because it would reduce the effect of the learnability ratings to an artifact of presentation rate. This would have little practical value in educational terms, as students normally study at a self-directed pace.

EXPERIMENT 3

In order to address the re-reading hypothesis, Experiment 3 used substantially longer presentation rates than Experiments 1 and 2. The stimulus onset asynchrony in Experiment 3 was 12 seconds, double that of Experiments 1 and 2. If the effect of the learnability ratings is caused by a lack of study time, then the effect should disappear or be greatly reduced at this presentation rate. This manipulation is the same as that used by Burns (1992) to rule out a processing time explanation for the effects of generation on response clustering in free recall.

With the exception of the presentation rate during study, Experiment 3 was the same as Experiment 2. That is, the manipulation of rating was conducted between subjects, and common and rare items were analyzed separately.

Method

Subjects. The subjects were 46 students from an introductory psychology course at McMaster University. They received course credit for their participation in the experiment. The experiment was conducted in group sessions of 11-13 subjects. Twenty-four subjects participated in the rating condition; twenty-two participated in the no rating condition.

Materials. The materials used in Experiment 3 were the same as those used in Experiment 2.

Procedure. The procedure used in Experiment 3 was identical to that used in Experiment 2, save in two details. The first change was that statements were presented on the television screen for study at a 12 second rate (11 seconds on the screen, 1 second blank). The second change was that the task preceding Experiment 3 was a short 20 word free recall task, rather than the word-pair task in Experiment 2.

Results

The results of Experiment 3 are shown in Table 3.

Table 3 - Cued Recall, Experiment 3

	COMMON ITEMS		RARE ITEMS	
GROUP	Studied	Unstudied	Studied	Unstudied
Rating	.82	.29	.66	.04
No Rating	.68	.28	.51	.05
PreQ Effect	.14	.01	.15	-.01

Recall scores were submitted to a three-factor analysis of variance with study (studied or unstudied items) and difficulty (common or rare items) as within-subject factors, and rating (rating or no rating) as a between-subjects factor.

The analysis revealed main effects of study and difficulty, and a marginal main effect of rating. As expected, studied items were recalled better than unstudied items [.67 vs. .17, F(1,44)=386.50, MSe=.03, p<.001] and common items were recalled better than rare items [.52 vs. .32, F(1,44)=204.92, MSe=.01, p<.001]. Overall, there was only a marginal difference in recall between subjects in the rating condition and subjects in the no rating condition [.45 vs. .38, F(1,44)=2.95, MSe=.08, p<.10].

The main effects described above were qualified by multi-factor interactions. Study interacted with rating [F(1,44)=8.22, MSe=.03, p<.01]. There was also a significant interaction between study and difficulty [F(1,44)=5.29, MSe=.01, p<.05]. The marginal interaction between rating and difficulty that was present in Experiment 2 was not present in Experiment 3 [F(1,44)<1, MSe=.01, ns]. The three-way interaction also failed to reach significance [F(1,44)=1.15, MSe=.01, ns].

As in Experiment 2, the studied items were analyzed separately from the unstudied items. An analysis of variance conducted on the unstudied items revealed a significant effect of difficulty but no effect of rating. Common items were remembered better than rare items [.29 vs. .05, F(1,44)=338.00, MSe=.01, p<.001] but subjects who had initially produced learnability ratings recalled no more new items than did subjects who did not produce learnability ratings [.17 vs. .17, F(1,44)<1, MSe=.02, ns]. Furthermore, the two factors did not interact [F(1,44)<1, MSe=.01, ns].

An analysis of variance conducted on the studied items revealed significant effects of difficulty and rating. In the studied items, common items were remembered better than rare items [.75 vs. .59, F(1,44)=62.43, MSe=.01, p<.001], and rated items were better remembered than unrated items [.74 vs. .59, F(1,44)=5.47, MSe=.09, p<.025]. There was no interaction between difficulty and rating [F(1,44)<1, MSe=.01, ns].

Discussion

The results of Experiment 3 serve to clarify and extend those of the preceding experiments. The major result is that doubling the amount of study time available did not reduce or eliminate the effect of prequestions. The nine percent advantage seen in Experiment 2 is here a fourteen percent advantage, and is statistically significant. The fact that study time was increased to 12 seconds without eliminating or reducing the size of the advantage argues against the re-reading hypothesis outlined earlier.

In addition to ruling out speed advantages due to re-reading as the sole cause of the prequestion effect, Experiment 3 further confirms that the use of a within-list design in Experiment 1 was not completely responsible for the increased learnability of rated items. The effect of rating was clearly significant in Experiment 3, and was nearly as large as the effect seen in Experiment 1.

The other major result of the experiment is the reconfirmation that the effect of prequestions is not limited to the common knowledge items. There were substantial effects of making learnability ratings on the recall of both common and rare items in Experiment 3. This is somewhat different from the results of Experiment 2, in which no advantage was found for common items. The reason for the difference between the two experiments in this regard is unclear. However, the fact that prequestions have an effect on rare knowledge facts as well as on common knowledge facts means that they may be a useful tool for the teaching of novel material, rather than being limited to review purposes.

Because attentional differences have been ruled out as the sole explanation of the prequestion effect in cued recall, it becomes of interest to discover other potential psychological determinants of the prequestion effect. Several different theoretical lines provide possible explanations. These will be examined in the next chapter.

CHAPTER 5 - EXPERIMENTS: PART 2

The prequestion effect and the generation effect are strikingly similar. Because of this similarity, one may expect that theories used to explain the generation effect may enjoy some success in explaining the prequestion effect as well. The experiments presented in this chapter will attempt to explore how successful those theories, as well as others, can be.

Current proposals regarding the source of the generation effect focus on the distinction between item-specific information and relational information (Begg, 1978; Einstein & Hunt, 1980; Hunt & Einstein, 1981). Hunt and Einstein (1981) propose that, when given a recall cue, subjects use relational information to gain access to a set of potential response candidates, and use item-specific information about the target to select the target from that set. Which part or parts of the recall process benefit from generation is not yet agreed upon. Begg and associates (Begg, Snider, Foley & Goddard, 1989; Begg, Vinski, Frankovich, & Holgate, 1991) base their explanation of the generation effect solely on item-specific benefits for the generation target, denying that generation is sufficient to increase relational information between the generation cue and target. Others, however, propose that cue-target relational information is also increased by generation (Burns, 1990, 1992; Hirshman & Bjork, 1988; McDaniel & Waddill, 1990; McDaniel, Waddill, & Einstein, 1988). To the extent that prequestioning results in mental processes similar to those that occur in a generation task, one may expect that the consequences of prequestioning might include changes in item-specific information about answers, and/or changes in relational information linking questions and answers.

A different explanation of the prequestion effect may be provided by Begg's (1982) Organization-Redintegration hypothesis. This hypothesis proposes that in cued recall for episodic information "...item-specific information allows a cue to identify a memory unit, at which point relational information makes the target available as a response." (Begg & Nicholson, 1994, pg. 401). Therefore, processes at study that enhance cue identifiability have the potential to increase performance on a cued recall test by increasing the likelihood that the recall cue will make contact with the memory of the study episode.

The Organization-Redintegration hypothesis is not alone in predicting that enhanced cue identifiability will result in increased performance on a cued recall test. Ausubel's (1968) Subsumption theory proposes that the learning of new meaningful material requires anchoring the new material to existing ideas in the learner's knowledge base. According to Ausubel,

> "...the learning and longevity in memory of new meaningful material are functions of the *stability and clarity* of its anchoring ideas. If they are ambiguous and unstable, they...provide inadequate relatability and weak anchorage for potentially meaningful new materials..." (pg. 132).

Therefore, increasing cue identifiability can improve cued recall performance in the event that cues that are not used as prequestions are less than adequately identifiable.

Another area of research that may provide insight into prequestion effects is devoted to studying the memorial effect of knowledge mobilization. It has been found that having subjects activate knowledge they already possess can facilitate the learning of new, related material (Mannies, Gridley, Krug, & Glover, 1989; Peeck, 1982; Peeck, van den Bosch, & Kreupeling, 1982). Attempts to explain the effect of knowledge mobilization on memory for prose have principally used concepts from schema theory[10]. The general argument is that knowledge mobilization activates relevant schemata in memory. "New information (whether in lists or passages) is then assimilated to the active schemata, which allows for the retention of greater amounts of information with better organization." (Mannies et al., pg. 122) This proposal is essentially the same as the framework hypothesis described by Brewer and Nakamura (1984). Brewer and Nakamura argue that if one thinks of a schema as being a frame with slots that hold information (e.g. Minsky, 1975), then placing information into a slot of an activated schema will result in better memory for that information than if the information was not placed into an organizing framework.

The difficulty with applying the framework concept to prequestion effects is that one has to assume that a relevant schema would not normally be activated. For many of the demonstrations of framework effects in schema research this is a plausible

---

[10] Schema theory is a class of memory theory that attempts to explain encoding, storage, and retrieval of memorial information on the basis of organizing structures (schemata) in memory that have been abstracted out of many past experiences. For example, recall of your latest visit to a restaurant might be organized and assisted by a "restaurant" schema built up from many previous restaurant visits. This schema might actually lead you to misremember that you were given a menu, when in fact you were not.

assumption (e.g. the washing clothes passage, Bransford & Johnson, 1972). However, most theories of prose comprehension incorporate schema activation as a basic process (e.g. Anderson & Pearson, 1984; Graesser, 1981). With the single sentence materials that were used in the experiments of Chapter 4, it is particularly difficult to see how this argument might apply. It is not obvious that schema activation would be different for "Nairobi is the capital of K______" than it is for "Nairobi is the capital of Kenya."

In short, there are a variety of potential explanations for the prequestion effect. In order to attempt to reduce the number of plausible explanations, it is helpful to examine the consequences of prequestioning on memory tasks other than cued recall. In particular, investigation of the item-specific effects of prequestioning may prove fruitful. Although the presence of item-specific effects of prequestioning does not necessarily imply that those effects are responsible for cued recall benefits, the absence of item-specific effects will rule out certain explanations of the prequestion effect in cued recall. For example, the absence of target-specific benefits of prequestioning would rule out a pure target-based explanation similar to Begg's (Begg, Snider, Foley, & Goddard, 1989; Begg, Vinski, Frankovich, & Holgate,

1991) explanation of the generation effect. Similarly, the absence of cue-specific benefits would rule out explanations based on the Organization-Redintegration hypothesis (Begg, 1982) or Ausubel's (1968) Subsumption theory.

EXPERIMENT 4

Experiment 4 was conducted in order to determine the item-specific effects of prequestioning on the question (cue) and the answer (target). Hunt and Seta (1984) have argued that recognition testing provides a relatively pure measure of item-specific information; therefore, subjects in the experiment were tested with target recognition, cue recognition, and cued recall.

Subjects in Experiment 4 received a study list of 40 general knowledge statements. Prior to studying the list, they were asked to fill out learnability ratings for 40 prequestions. For half the subjects, the 40 prequestions covered the same content as the 40 statements they would later study. The other half of the subjects received 40 prequestions that were unrelated to the 40 statements they would study.[11]

---

[11] This is the first experiment presented in this paper in which irrelevant prequestions have been employed in the control group. Few prequestion researchers have employed irrelevant prequestions in the control group, though there are exceptions (e.g.

Following exposure to the study list, subjects were given tests of cue (question) recognition, target (answer) recognition, and cued recall of the answer given the question as a cue.

Method

Subjects. The subjects were 105 students from an introductory psychology course at McMaster University. They received course credit for their participation in the experiment. The experiment was conducted in group sessions of 11-15 subjects. Fifty subjects participated in the relevant prequestion condition; fifty-five subjects participated in the irrelevant prequestion condition.

Materials. The materials used in Experiment 4 were derived from 120 general knowledge statements selected from the Nelson and Narens (1980) norms. Each of the 120 statements was rewritten so that it could be turned into a fill-in-the-blank question through the deletion of the final word of the statement, e.g. "The name of King Arthur's sword was E________". The initial letter of the final word was always presented when the rest of the word was deleted. The pool of 120 items was divided into three sets of 40. The average recall

---

Rickards & DiVesta, 1974, Rickards, 1976). The presence of the irrelevant prequestions controls for general arousal effects of prequestions, or whatever other effects may occur as a result of mere exposure to questions before studying text.

probabilities for the three sets were .245, .244, and .243, based on the Nelson and Narens norms.

Two study lists were created for presentation on a video monitor. Each study list consisted of the 40 statements from one of the item sets. Only item sets A and B were used for study purposes. Half the groups studied item set A; the other half studied item set B. The order of statements in the study lists was random.

Three booklets were created to collect learnability ratings from the subjects. Each booklet contained the question versions of items from one of the three item sets. Below each question was a scale from 0 to 100 marked with numbers at intervals of 20. Relevant prequestion booklets (item sets A and B) contained the question versions of the 40 items that would later be studied by subjects receiving those booklets. The irrelevant prequestion booklet (item set C) contained the question versions of 40 items that would not be studied. The order of the questions in the ratings booklets was random, with the exception that questions related to statements in the first half of the study list appeared in the first half of the prequestion booklet, and questions related to statements in the second half of the study list appeared in the second half of the prequestion booklet.

Each subject received two test booklets. One booklet was used for cue recognition and cued recall, and consisted of the question forms of the 80 items from sets A and B. Because each subject received either set A or set B as studied items, half the items in this booklet had been presented intact at study, and half were new. Each question was presented beside a rating scale ranging from 1 to 5, to be used for recognition testing. A rating scale at the top of the test booklet indicated that 1 corresponded to "certainly no" and 5 corresponded to "certainly yes". The order of the questions in the test booklet was random, with the same restriction as the learnability ratings forms.

The other test booklet was used for target recognition, and consisted of the last words of the 80 items from sets A and B. Once again, because each subject received either set A or set B as studied items, half the items in this booklet had been studied, and half were new. Beside each word was a 1 to 5 rating scale, with a scale at the top of the booklet again indicating that 1 corresponded to "certainly no" and 5 to "certainly yes". The order of the words in the booklet was random, with the same restriction as the learnability ratings forms.

Procedure. Experiment 4 was conducted in three phases:

a rating phase, a study phase, and a test phase.

In the rating phase, subjects were given a learnability rating booklet and were asked to make learnability ratings for each of the 40 questions in the booklet. More precisely, they were asked to estimate, for each question, how likely they were to be able to answer that question on a final recall test, if they were given a chance to study the answer before being tested. Half the subjects in each group were given a relevant prequestion booklet, and the other half were given an irrelevant prequestion booklet. The rating phase was self-paced, and required approximately 5-7 minutes to complete.

The study phase immediately followed the rating phase, and consisted of a video-taped presentation of each of the 40 statements in the study list. Statements appeared one at a time on a television screen. Each statement appeared on the screen for 7 seconds, followed by a 1-second blank interval. Thus, the entire study phase lasted about 5.5 minutes to present.

Following the study phase, subjects were tested for recognition and cued recall. Half the subjects received the target recognition booklet first. They were instructed to make a judgment for each word, to indicate whether they remembered that word as being the last word of a statement in the study list. The target recognition task required 5-7 minutes to complete. Next, the subjects received the cue recognition/cued recall booklet, and were instructed to try to answer as many of the questions as they could, whether they had seen the answer in the study phase of the experiment or not. The cued recall task required 10-12 minutes to complete. Finally, the subjects were asked to examine each question in the cue recognition/cued recall booklet and make a recognition judgment to indicate whether they remembered those items appearing in the study phase of the experiment. The cue recognition task required 5-7 minutes to complete. The other half of the subjects in the experiment received the three tests in the opposite order, i.e. cue recognition, followed by cued recall, followed by target recognition.

Results

Cue Recognition. Because of the potential for contamination of the cue recognition results by prior cued recall attempts, the cue recognition ratings were analyzed only for the 52 subjects who began the test phase with cue recognition.

For each subject, mean cue recognition ratings were calculated, contingent on cued recall success or failure and on the presence or absence of the item during the study phase. That is, for each subject, mean recognition ratings were determined for recalled/studied items, recalled/unstudied items, unrecalled/studied items, and unrecalled/unstudied items. Four subjects were dropped from subsequent analysis because they failed to answer even one unstudied item in the cued recall phase, rendering the calculation of means impossible. The mean cue recognition ratings for the remaining subjects are shown in Table 4.

Table 4 - Cue Recognition Ratings, Experiment 4

PREQUESTION	RECALLED		UNRECALLED	
TYPE	Studied	Unstudied	Studied	Unstudied
Relevant	4.975	1.005	4.816	1.041
Irrelevant	4.903	1.067	4.006	1.243
PreQ Effect	0.072	-0.062	0.810	-0.202

The mean ratings were submitted to a three-factor analysis of variance with recall (recalled, unrecalled) and study (studied, unstudied) as within-subject factors, and prequestion type (relevant, irrelevant) as a between-subjects factor.

Every main effect and interaction in the analysis of variance was significant at $p<.002$ or better. There was a main effect of prequestion type, with relevant prequestions leading to higher recognition ratings than irrelevant prequestions [2.959 vs. 2.805, $F(1,46)=11.45$, $MSe=0.10$, $p<.002$]. Also, recall success was associated with recognition ratings, recalled items receiving higher ratings than unrecalled items [2.987 vs. 2.777, $F(1,46)=57.98$, $MSe=0.04$, $p<.001$]. Not surprisingly, studied items received higher recognition ratings than did unstudied items [4.675 vs. 1.089, $F(1,46)=7671.31$, $MSe=0.08$, $p<.001$]. However, these main effects were qualified by interactions between prequestion type and recall [$F(1,46)=27.54$, $MSe=0.04$, $p<.001$], between prequestion type and study [$F(1,46)=49.14$, $MSe=0.08$, $p<.001$], and between recall and study [$F(1,46)=126.10$, $MSe=0.04$, $p<.001$]. Finally, the three-way interaction between prequestion type, recall, and study was also significant [$F(1,46)=56.98$, $MSe=0.04$, $p<.001$].

In order to provide a more tractable view of the cue recognition results, the recalled items were analyzed separately from the unrecalled items. A two-factor analysis of variance of the recalled items with prequestion type as a between-subjects factor and study as a within-subject factor revealed no main effect of prequestion type [$F(1,46)<1$]. As expected there was a main effect of study [studied vs. unstudied: 4.939 vs. 1.036, $F(1,46)=18747.53$, $MSe=0.02$, $p<.001$]. In addition, there was a significant interaction between prequestion type and study [$F(1,46)=5.49$, $MSe=0.02$, $p=.02$]. This interaction occurred because prequestioning improved recognition of old items and rejection of new items, which reflects the standard mirror-effect in recognition memory (Glanzer & Adams, 1985). Despite the significant interaction, there were only weak effects of prequestion type within the studied items [t(46)=1.51, p=.13] and within the unstudied items [t(46)=1.65, p=.10].

In the unrecalled items, a two-factor analysis of variance with prequestion type as a between-subjects factor and study as a within-subject factor revealed a main effect of prequestion type [relevant vs. irrelevant: 2.929 vs. 2.624, F(1,46)=19.36, MSe=0.11, p<.001]. There was also a main effect of study [studied vs. unstudied: 4.411 vs. 1.142, F(1,46)=2507.35, MSe=0.10, p<.001]. These main effects were qualified by a significant interaction between prequestion type and study, again reflecting the standard mirror-effect [F(1,46)=60.70, MSe=0.10, p=<.001]. There were significant effects of prequestion type within both the studied items [t(46)=6.92, p<.001] and the unstudied items [t(46)=3.07, p=.004].

Overall, the results show superior recognition of cues by the subjects who received relevant prequestions beforehand. This effect was substantially stronger on the items that would later fail in cued recall than it was on the items that would later succeed in cued recall. However, there is some potential that ceiling and floor effects influenced the ratings for successfully recalled items and artificially created the difference in recognition scores.

Cued Recall. Because of the potential for contamination of the cued recall results by prior presentation of targets in the target recognition task, the cued recall scores were analyzed only for the 52 subjects who began the test phase with cue recognition, followed by cued recall. The cued recall results are shown in Table 5.

Table 5 - Cued Recall, Experiment 4

PreQ Type	Studied	Unstudied
Relevant	.63	.15
Irrelevant	.48	.16
PreQ Effect	.15	-.01

Recall scores were submitted to a two-factor analysis of variance with study (studied, unstudied) as a within-subject factor, and prequestion type (relevant, irrelevant) as a between-subjects factor.

The analysis revealed a main effect of prequestion type, with subjects in the relevant prequestion group answering more questions than subjects in the irrelevant prequestion group [.39 vs.

.32, F(1,50)=4.00, MSe=0.03, p=.048]. In addition, there was a main effect of study. Subjects performed substantially better on questions they had studied than on questions that they had not studied [.55 vs. .15, F(1,50)=6544.47, MSe=0.01, p<.001]. These main effects were qualified by a significant interaction between study and prequestion type [F(1,50)=242.62, MSe=0.01, p<.001].

In order to understand the nature of the interaction, the studied items were analyzed separately from the unstudied items. Within the unstudied items, there was no effect of prequestion type [t(50)=0.41, ns]. However, within the studied items, there was a strong effect of prequestion type [t(50)=2.95, p=.005].

In summary, the cued recall results show the standard prequestion effect: higher recall of studied prequestioned items than of studied but unprequestioned items.

Target Recognition. Because of potential contamination of the target recognition results by differential recall results in prior cued recall, the target recognition ratings were analyzed only for the 53 subjects who began the test phase with target recognition.

For each subject, mean target recognition ratings were calculated, contingent on the presence or absence of the rated item during the study phase. The mean target recognition ratings are shown in Table 6.

Table 6 - Target Recognition Ratings, Experiment 4

PreQ Type	Studied	Unstudied
Relevant	4.397	1.402
Irrelevant	4.208	1.524
PreQ Effect	0.189	-0.122

The mean ratings were submitted to a two-factor analysis of variance with study (studied, unstudied) as a within-subject factor, and prequestion type (relevant, irrelevant) as a between-subjects factor.

The analysis revealed only a main effect of study, and a marginal interaction between prequestion type and study. Subjects reliably identified previously studied items as having been previously studied [studied vs. unstudied: 4.293 vs. 1.469, $F(1,51)=1032.50$, $MSe=0.18$, $p<.001$]. There was no main effect of prequestion type [$F(1,51)<1$]. The interaction between prequestion type and study was only marginally significant, but was in the direction of a standard mirror effect [$F(1,51)=3.095$, $MSe=0.21$, $p=.081$]. When the studied items were analyzed separately from the unstudied items, there was only a marginal effect of prequestion type in the studied items [$t(51)=1.53$. $p=.128$] and no effect of prequestion type in the unstudied items [t(51)=1.03, ns].[12]

The results of the target recognition test demonstrated only a weak effect of prequestioning on memory for the targets, which were not presented during the prequestion phase.

Discussion

The results of Experiment 4 do not completely rule out any of the potential explanations of the prequestion effect. However, given the weak prequestion effect in target recognition and the strong prequestion effect in both cue recognition and cued recall, it seems more probable that cue-specific information is important than that target-specific information is important.

EXPERIMENT 5

The results of Experiment 4 indicate that cue-specific theories of the prequestion effect are more likely to be correct than are target-specific theories. This stands in contrast to most of the generation effect literature, which pays scant attention to cue-specific effects. Within the generation effect literature, those explanatory theories that do not

---

[12] Because of the marginal nature of the target recognition results, the data from the subjects who received the opposite order of tests was also analyzed. These data are included in Appendix A.

solely emphasize target-specific information (e.g. Begg, Snider, Foley, & Goddard, 1989) rely upon cue-target relational information to explain cued recall results (e.g. Burns, 1990).

If cue identifiability is the driving force behind the prequestion effect in cued recall, then manipulations that result in enhanced cue identifiability should result in greater prequestioning effects. A commonly accepted method of improving memory for repeatedly studied information is to distribute study episodes rather than to mass them (Madigan, 1969). Because prequestioning results in two presentations of the question portion of an item, distributing those two presentations across an intervening time lag (as in standard prequestioning) should be more effective at establishing cue identifiability than having the two presentations massed (as in standard generation). It follows that if prequestion effects in cued recall rely on cue identifiability, as predicted by the Organization-Redintegration approach (Begg, 1982) or the Subsumption approach (Ausubel, 1968), then prequestioning should be more effective than generating.

Experimental support for the idea that cue-identifiability is a major influence in the prequestion effect, and that distributing the presentations of the prequestion and statement will enhance the effect, comes from a couple of sources. As discussed in Chapter 2, Berlyne (1966) found that memory performance was better for subjects who experienced a delay before being presented with the correct answer to a question than it was for subjects who received the correct answer immediately after receiving the question. However, Berlyne's experiment suffers from the problem that potential target candidates were presented during prequestioning, thus allowing some cue-target relational processing to occur even without certain knowledge of the correct answer.

In addition to Berlyne's (1966) experiment, evidence that supports the cue identifiability hypothesis comes from experiments performed by Glover and associates (Glover, Bullock, & Dietzer, 1990; Glover, Krug, Dietzer, George, & Hannon, 1990). Working with Ausubel's (1968) subsumption theory, Glover has found that presenting an advance organizer an hour in advance of a study passage results in better memory for material from the study passage than does presenting the advance organizer immediately prior to the study passage.

Experiment 5 sought to determine whether the preceding results would generalize to the current experimental paradigm. Subjects in the experiment were presented with an extended mixed list of prequestions and statements. Prequestions were presented either immediately before the relevant statement (generation), one intervening item in advance of the relevant statement (prequestion short lag) or many items in advance of the relevant statement (prequestion long lag). Cue identifiability hypotheses predict that the prequestion effect will increase from the generation condition to the prequestion short lag condition to the prequestion long lag condition. Cue-target relational hypotheses predict no such increase, because relational information cannot properly be encoded until the target is presented.

Method

Subjects. The subjects were 135 students from an introductory psychology course at McMaster University. They received course credit for their participation in the experiment. The experiment was conducted in group sessions of 10-16 subjects.

Materials. The materials used in Experiment 5 were derived from 84 general knowledge statements taken from the Nelson and Narens (1980) norms. Each of the 84 statements was rewritten so that it could be turned into a fill-in-the-blank question through the deletion of the final word of the statement, e.g. "The name of Roy Rogers' dog was B_______". The initial letter of the final word was always presented when the rest of the word was deleted. Sixty of the selected items were used as to-be-learned material, and were divided into five sets of twelve items. The average recall probabilities for the five sets were .236, .235, .236, .236, and .237, based on the Nelson and Narens norms. The other 24 statements were employed as fillers to facilitate construction of the study list.

In Experiment 5, the rating phase and the study phase were combined into a mixed-task phase. Lists that mixed questions with statements were therefore created for this phase. Within each list there were five item conditions, defined on the basis of the number of times an item appeared in the list, and the lag between multiple appearances. Twelve *Read Only* items were presented in each prequestion/study list, appearing only once, in statement form. Twelve *Generated* items were positioned in each mixed-task list such that the statement version of the item immediately followed the prequestion version. This is the presentation sequence used in some generation effect investigations (e.g. Kane & Anderson, 1978; Begg & Snider, 1987). Another twelve items were positioned such that the statement version followed the prequestion version with a single intervening display. These items were designated *Prequestion Short Lag* items. Twelve *Prequestion Long Lag* items were positioned such that there was a substantial number of intervening displays (mean=40.6, range=28-51) between the prequestion version and the statement version of the item. Finally, twelve *New* items were presented only at test, in order to measure prior knowledge.

Five mixed-task lists were generated by rotating the five item sets through the five item conditions. The order of the lists was random, with two restrictions. The first restriction was dictated by the various item conditions, e.g. that generated items have their statement version immediately follow their question version. The second restriction was that each block of 14 presentations contain two read only items, two generated items, two prequestion short lag items, two prequestion long lag statements, and two prequestion long lag questions.

The test booklet consisted of the question forms of each of the 60 non-filler items. The questions were presented in random order, with the restriction that statements in the first third of the study list appeared in question form in the first third of the test list, and similarly for the middle third and final third of the test list.

Procedure. The experiment was conducted in two phases: a mixed prequestion/study phase, and a recall phase.

In the mixed phase, subjects were instructed that they would be shown a list that contained a mix of statements and questions, and that sometimes the answer to a question would later be provided in the list. Subjects were asked to think of the answer to each question they saw, and to study all of the statements they saw. They were informed that they would be given a test on the statements later. The prequestion/study phase consisted of a video-taped presentation 120 items long. Statements and questions appeared one at a time on a television screen, each statement appearing on the screen for 7 seconds, with a 1-second blank interval between items. The entire study phase lasted about 16 minutes.

Following the prequestion/study phase, subjects were given test booklets and instructed that for each question they were to attempt to write down the correct answer, whether they had seen it on the computer screen or not. Subjects were given as much time as they needed for this task. This phase required approximately 9 minutes to complete.

Results

The results of Experiment 5 are shown in Table 7.

Table 7 - Cued Recall, Experiment 5

Item type	Recall	PreQ Effect
New	.16	
Read Only	.49	
Generated	.54	.05
Preq Short	.57	.08
Preq Long	.55	.06

Recall scores were submitted to a single-factor analysis of variance with item type (new, read only, generated, preq short, preq long) as a within-subject factor.

The analysis revealed a significant main effect of item type [$F(4,536)=165.91$, $MSe=0.02$, $p<.001$]. In a follow-up analysis, planned orthogonal contrasts revealed that the effect of item type was found because the new items were more poorly answered than the other item types [new vs. studied: $F(1,134)=626.28$, $MSe=0.02$, $p<.001$] and also because prequestioned or generated items were remembered better than read items [read only vs. (generated + prequestioned): $F(1,134)=17.90$, $MSe=0.03$, $p<.001$]. The generated items and the prequestioned items did not differ significantly [generated vs. prequestioned: $F(1,134)<1$] although there was a marginal drop in performance from the prequestion short lag items to the prequestion long lag items [preq short vs. preq long: $F(1,134)=2.51$, $MSe=0.02$, $p=.116$].

In summary, the results of Experiment 5 show benefits for prequestioning over reading. However, the lag between presentation of the prequestion and the presentation of the statement appears to be inconsequential.

Discussion

Contrary to expectations, distributing the presentations of the prequestion and statement version of an item did not result in enhanced cued recall. The prequestion long lag results were virtually identical to the generate results. Following the logic presented by way of introduction to Experiment 5, this result implies that cue identifiability is not the primary determinant of the prequestion effect. However, caution is required in accepting this conclusion. The results of Experiment 5 are in apparent contradiction to the findings from the work with advance organizers (Glover, Bullock, & Dietzer, 1990; Glover, Krug, Dietzer, George, & Hannon, 1990), and also to the findings from Berlyne's (1966) work with prequestions. It is possible that the results of Experiment 5 reflect a type II statistical error.

Another possibility is that the results of Experiment 5 were influenced by ceiling effects. If cue recognition was at ceiling for the generated items, then no further benefit could have been observed for prequestioned items. However, the failure to anticipate this problem, and to collect cue recognition ratings, leaves me unable to address the truth of this proposal.

EXPERIMENT 6

Because of the ambiguity surrounding the outcome of Experiment 5, I decided to approach the issue of cue identifiability from a different perspective. If cue identifiability is the only critical factor in the prequestion effect, then the semantic match between the cue and target should have little consequence for the size of the prequestion effect. That is, because increments to cue identifiability occur prior to the presentation of the target, it should be possible to find prequestion effects for "The capital of Australia is Pencil.", as well as for "The capital of Australia is Canberra."

In contrast to the cue identifiability theory of prequestion effects, other theories do not predict a prequestion effect for unrelated cue and target pairs. Schema theory, particularly as exemplified by Minsky's (1975) frame theory, predicts that there will be no benefit for unrelated targets. According to Minsky's theory, the slots that are instantiated by schema-relevant information have restrictions placed on the acceptable contents of a slot. Presumably "Pencil" would be unacceptable content for a "Capital of Australia" slot, and it is unlikely that a clearly incorrect city, such as "Paris", would be able to fill the slot either.

Theories based on partial target activation also predict no benefit for unrelated cue-target pairs. In explaining the effect of generation, Slamecka and Fevreiski (1983) argued that generation failures (analogous to unanswered prequestions) were "...instances of incomplete generation, that is, occasions where generation of the semantic attributes had not been followed by self-access to the proper lexical entry." (pg. 160) Because cues in unrelated or weakly related cue-target pairs are unlikely to provide partial generation or activation of their study-list target, the partial target activation hypothesis predicts no prequestion effect for either unrelated or weakly related pairs.

Experiment 6 tested prequestion effects for four degrees of relatedness between the prequestion cue and the target. These were: correct targets (The capital of Australia is Canberra), incorrect strongly related targets (Sydney), incorrect weakly related targets (Paris), and unrelated targets (Pencil). Subjects received either relevant or irrelevant prequestions prior to study. The study list consisted of a mix of statements from each of the four levels of relatedness.

Method

Subjects. The subjects were 108 students from an introductory psychology course at McMaster University. They received course credit for their participation in the experiment. The experiment was conducted in group sessions of 11-15 subjects. Fifty-four subjects participated in each of the between-subjects conditions.

Materials. The materials used in Experiment 6 were derived from 128 general knowledge statements taken from the Nelson and Narens (1980) norms. Each of the 128 statements was rewritten so that it could be turned into a fill-in-the-blank question through the deletion of the final word of the statement, e.g. "The capital of Australia is ______". Unlike Experiments 1-5, the initial letter of the final word was not present when the item appeared in question form. Sixty-four items were assigned to be study material, and the other sixty-four were assigned to serve as irrelevant prequestions. The pool of 64 to-be-studied items was divided into four sets of 16 items. The average recall probabilities for the four sets were .281, .279, .288, and .285, based on the Nelson and Narens norms.

Four possible completions were assigned to each to-be-studied item. *Correct* completions were the correct answer to the question, e.g. "The capital of Australia is Canberra." *Incorrect strong* completions were considered to be somewhat plausible or closely related to the correct answer - "The capital of Australia is Sydney." *Incorrect weak* completions came from the same category as the correct answer, but were more apparently incorrect - "The capital of Australia is Paris." *Unrelated* completions were drawn at random from a pool of 64 concrete, medium-frequency nouns (Concreteness 6.0-7.0, Paivio, Yuille, & Madigan, 1968; Frequency 24-49 per million, Thorndike & Lorge, 1944) - "The capital of Australia is Pencil." Incorrect strong and incorrect weak completions were chosen by the author. No norms exist to establish either mean recall or degree of association for these completions; therefore, all items were reviewed by colleagues prior to their use in the experiment. A complete list of stimuli is in Appendix B.

Four study lists were created for presentation on a video monitor. Each study list contained 64 statements, with each of the four completion conditions (correct, incorrect strong, incorrect weak, unrelated) contributing 16 statements. The four lists were generated by rotating each statement through each completion condition. The order of statements in the study lists was random, except that each block of four statements contained one statement from each of the four completion conditions.

Two booklets were created to collect learnability ratings from the subjects. Each booklet contained the question versions of 64 items. Below each question was a scale from 0 to 100 marked with numbers at intervals of 20. The relevant prequestion booklet contained the question versions of the 64 to-be-studied items. The irrelevant prequestion booklet contained the question versions of the 64 unstudied items. The order of the questions in the ratings booklets was random, with the exception that questions related to statements in the first half of the study list appeared in the first half of the prequestion booklet, and questions related to statements in the second half of the study list appeared in the second half of the prequestion booklet.

The test booklet consisted of the question forms of each of the 64 studied statements. The questions were presented in random order, with the same restriction as the ratings booklets.

Procedure. Experiment 6 was conducted in three phases: a rating phase, a study phase, and a recall phase.

In the rating phase, subjects received a learnability rating booklet and were asked to make learnability ratings for each of the 64 questions in the booklet. More precisely, they were asked to estimate, for each question, how likely they were to be able to answer that question on a final recall test, if they were given a chance to study the answer before being tested. The subjects were not warned that many of the studied answers would be incorrect. Half the subjects in each group were given the relevant prequestion booklet, and the other half were given the irrelevant prequestion booklet. The rating phase was self-paced, and required approximately 7-8 minutes to complete.

The study phase immediately followed the rating phase, and consisted of a video-taped presentation of each of the 64 statements in the study list. Statements were presented one at a time on a television screen. Each statement appeared on the screen for 7 seconds, with a 1-second blank interval between statements. Thus, the entire study phase lasted about 8.5 minutes.

Following the study phase, subjects received recall booklets and were instructed that for each question they were to attempt to write down the completion they had seen on the computer screen. Subjects were given as much time as they needed for this task. This phase required approximately 9-10 minutes to complete.

Results

The results of Experiment 6 are shown in Table 8.

Table 8 - Cued Recall, Experiment 6

PreQ Type	Correct	Strong	Weak	Unrelated
Relevant	.62	.68	.64	.22
Irrelevant	.55	.57	.58	.24
PreQ Effect	.07	.11	.06	-.02

Recall scores were submitted to a two-factor analysis of variance with prequestion type (relevant, irrelevant) as a between-subjects factor and completion type (correct, incorrect strong, incorrect weak, unrelated) as a within-subject factor.

The analysis revealed a significant main effect of completion type [$F(3,318)=259.80$, $MSe=0.01$, $p<.001$] and a marginal main effect of prequestion type [$F(1,106)=3.09$, $MSe=0.10$, $p=.082$]. In addition, the prequestion type by completion type interaction was significant [$F(3,318)=4.98$, $MSe=0.01$, $p=.002$].

The main effect of completion type is not surprising given the lack of cued recall norms for the false answers. In a follow-up analysis of the data using planned orthogonal contrasts, it was discovered that most of the main effect of completion type effect was caused by low performance on the unrelated completions compared with all other completion types [$F(1,106)=605.99$, $MSe=0.02$, $p<.001$], although there was also a difference found between the correct completions and the incorrect strong completions [$F(1,106)=7.30$, $MSe=0.01$, $p=.008$].

The interaction of completion type with prequestion type is the main result of interest. Planned orthogonal contrasts revealed that the interaction resulted from the fact that the correct, incorrect strong, and incorrect weak completion conditions showed statistically similar prequestion effects [correct vs. incorrect strong: $F(1,106)=1.83$, $p=.18$; (correct + incorrect strong) vs. incorrect weak: $F(1,106)<1$] whereas the unrelated completion condition resulted in a substantially different prequestion effect [(correct + incorrect strong + incorrect weak) vs. Unrelated: $F(1,106)=10.12$, $MSe=0.02$, $p=.002$]. In separate t-tests for each completion condition the effect of prequestion type was significant in the incorrect strong and incorrect weak conditions [$t(106)=2.61$, $p=.01$ and $t(106)=2.11$, $p=.04$ respectively], marginal in the correct condition [$t(106)=1.56$, $p=.123$], and non-significant in the opposite direction in the unrelated condition [$t(106)=0.57$, ns].

In summary, prequestion effects in cued recall exist for pretty much any level of relatedness between cue and target, with the exception of completely unrelated pairs. The magnitude of the prequestion effect does not appear to vary with the degree of relatedness of the cue and target.[13]

Discussion

The results of Experiment 6 clearly refute cue identifiability as the sole explanation of the prequestion effect in cued recall. If cue identifiability were the critical determinant of the prequestion effect, then completely unrelated cue-target pairs should have benefited from prequestioning as much as the related cue-target pairs. They did not. The results of Experiment 6 support the results of Experiment 5, which found no benefit for distributed practice of cue information on cued recall performance. Both experiments indicate a lack of influence of cue identifiability on prequestion effects in cued recall.

The results of Experiment 6 also provide information regarding other potential explanations of the prequestion effect. The fact that a strong

---

[13] There was some concern that the lack of a prequestion effect in the unrelated items might be attributable to the within-list manipulation of the four item types. A follow-up, Experiment 6b, is included in Appendix C to demonstrate that this is not the case.

prequestion effect was found for weakly related cue-target pairs such as "The capital of Australia is Paris." argues against a simple partial target activation hypothesis. At best, weakly related pairs would be expected to provide weak activation of the target. They would therefore be expected to derive only limited benefits from prequestioning. However, weakly related cue-target pairs showed prequestion benefits as large as those found for more strongly related cue-target pairs.

The findings for weakly related cue-target pairs also provide evidence inconsistent with schema-based explanations of prequestion effects, unless one accepts that the restrictions on slot contents in a frame are exceptionally weak or non-existent. Such weak restrictions run counter to the spirit of most schema theorization (e.g. Brewer & Nakamura, 1984).

EXPERIMENT 7

An aspect of prequestion effects that has not yet been investigated in this paper is whether or not the time that is spent attempting to answer prequestions might be better spent examining the material that is to be learned. Peeck (1970) employed two control groups in a prequestion experiment. Subjects in the prequestion conditions received four minutes to attempt to answer (or just read) prequestions, and were then given 15 minutes to read and study a related passage. The first control group was matched to the prequestion groups in that subjects received only 15 minutes to study the passage. However, subjects in a second control group received 19 minutes to study the passage, combining the time spent by prequestioned subjects on both the prequestions and the passage.

The results of Peeck's (1970) experiment were that the extended reading time group did not differ from the prequestioned groups in overall memory for the passage, as measured by a test that contained 50% prequestioned material and 50% unprequestioned material. The performance on prequestioned material was highest in the prequestioned groups, which differed significantly from the extended reading time control group. However, performance on unprequestioned material was highest in the extended reading time control group, which differed significantly from the prequestion (guess) group but not from the prequestion (no guess) group. The standard control group was significantly inferior to the extended reading time control group both in overall recall and in recall of prequestioned material. The two groups did not differ significantly in recall of unprequestioned material, but the numerical superiority of the extended reading time control group was preserved.

In his discussion of the above results, Peeck (1970) commented that

> "Results of [the extended reading time control group] show that regarding the total amount of knowledge acquired, time spent on prequestions might just as profitably be used for simply extending the reading time of the actual reading material." (pg. 245)

Of course, interpretation of Peeck's statement must be tempered by the knowledge that the pattern of results in the total recall scores masks offsetting effects on memory for prequestioned information and memory for unprequestioned information.

Applying Peeck's results to the current paradigm, in which all material is prequestioned, is difficult. Clearly the extra reading time was beneficial to Peeck's (1970) subjects. However, on prequestioned material, subjects performed significantly better if they actually received the prequestions than if they received extended reading time instead. Treated simplistically, that result would lead one to believe that in the current experimental paradigm, in which all material is prequestioned, subjects who are prequestioned should outperform subjects who receive extra study. However, the mapping of Peeck's results onto the current paradigm is hampered by the fact that selective attentional factors were likely operating in Peeck's experiments, whereas such factors are precluded in the current experiments.

Experiment 7 was conducted in order to investigate the effectiveness of prequestioning in comparison with the effectiveness of exposure to the material to be learned, and also in comparison with exposure to the targets to be produced (i.e. the answers alone). Because subjects in Experiment 7 were unable to artificially inflate the effectiveness of prequestions by strategically allocating attention, the results of the experiment should produce a purer estimate of the effectiveness of prequestions as aids to study in comparison with merely devoting more time to the studied material.

Experiment 7 repeats the basic design of Experiment 6. Each subject received a 64-statement mixed study list of correct answers, incorrect strong answers, incorrect weak answers, and unrelated answers. However, rather than using only two groups of subjects, as in Experiment 6, four groups of subjects were employed in Experiment 7. Each group was asked to make learnability ratings for 64 items prior to the study phase. One group rated relevant prequestions, another group rated relevant statements (i.e. both question and answer were present), a third group rated only the relevant targets, and the final group was a control group that rated 64 irrelevant statements.

Method

Subjects. The subjects were 217 students from an introductory psychology course at McMaster University. They received course credit for their participation in the experiment. The experiment was conducted in group sessions of 12-15 subjects. The number of subjects participating in the relevant statement, relevant prequestion, relevant target, and control conditions were 51, 54, 55, and 57 respectively.

Materials. The materials used in Experiment 7 were the same as those used in Experiment 6, with the following exception.

Learnability rating booklets existed in four different forms. Each rating booklet consisted of 64 items, below each of which was a scale from 0 to 100 marked with numbers at intervals of 20. Relevant statement booklets contained the statement versions of the 64 to-be-studied items. Because there were four possible completions for each item, this required that four relevant statement booklets be created. Relevant target booklets contained only the final word of the 64 to-be-studied items. There were four versions of this booklet as well. The relevant prequestion booklet contained the question version of the 64 to-be-studied items. Because no completions were presented, only one version of the relevant prequestion booklet existed. The control booklet contained the statement versions of the 64 unstudied items. Only one version of this booklet was created. The order of items in the booklets was random, with the same restriction as in Experiment 6.

Procedure. The experimental procedure was the same as that used in Experiment 6, save that all the subjects in each experimental session participated in the same experimental condition.

Results

The results of Experiment 7 are shown in Table 9.

Table 9 - Cued Recall, Experiment 7

RATING TYPE	CORRECT	STRONG	WEAK	UNRELATED
Rel State	.70	.73	.77	.52
Rel Preq	.54	.56	.61	.24
Rel Targ	.56	.55	.61	.24
Control	.52	.55	.55	.22

Recall scores were submitted to a two-factor analysis of variance with rating type (relevant statement, relevant prequestion, relevant target, control) as a between-subjects factor and completion type (correct, incorrect strong, incorrect weak, unrelated) as a within-subject factor.

The analysis revealed a significant main effect of completion type [F(3,639)=289.36, MSe=0.02, p<.001] and a significant main effect of rating type [F(3,213)=21.49, MSe=0.10, p<.001]. In addition, the rating by completion interaction was significant [F(9,639)=2.89, MSe=0.02, p=.002].

As in Experiment 6, the main effect of completion type is not surprising. A follow-up analysis of the data using planned orthogonal contrasts revealed that, once again, the majority of the main effect of completion type was caused by low performance on the unrelated completions compared with all other completion types [F(1,213)=594.04, MSe=0.02, p<.001], although there was also a difference found between the incorrect weak completions and the (correct + incorrect strong) completions [F(1,213)=23.10, MSe=0.01, p<.001]. The difference between the correct and incorrect strong completions found in Experiment 6 did not reach significance here [F(1,213)=1.70, MSe=0.01, ns].

The main effect of rating type in Experiment 7 appears to be entirely attributable to the relevant statement group. That group markedly outperformed the relevant prequestion group, the relevant target group, and the control group: the overall mean recall levels were .68, .49, .49, and .46 for the four groups respectively. The relevant statement group differed significantly from the other three conditions, but no other differences were significant using a Tukey's HSD test.

The interaction of rating type with completion type is once again of interest. Planned orthogonal contrasts revealed that the interaction was due to the fact that the correct, incorrect strong, and incorrect weak completion conditions responded similarly to the manipulation of rating type [$F(3,213)<1.00$ for both correct vs. incorrect strong and (correct + incorrect strong) vs. incorrect weak] whereas the unrelated completion condition responded differently to the manipulation of rating type [(correct + incorrect strong + incorrect weak) vs. unrelated: $F(3,213)=5.15$, $MSe=0.02$, $p=.002$]. In ANOVAs conducted for each completion condition, the effect of rating type was found to be consistently significant [smallest $F(3,213)=7.63$, $p<.001$]. Comparisons using Tukey's HSD showed that in each completion condition the relevant statement group significantly outperformed the other three groups, which did not statistically differ. The interaction between rating type and completion type appears to arise largely because the superiority of the relevant statement group over the mean of the other groups was larger in the unrelated completion condition (.29 advantage) than in the other completion conditions

(.18, .18, .16).

Thus, the results of Experiment 7 show that a significant benefit arises from dual presentation of to-be-studied material, particularly for unrelated items. The results in the prequestion condition of this experiment failed to replicate the pattern seen in Experiment 6.[14]

Discussion

The major question addressed by Experiment 7 has been firmly answered. Spending time by working on prequestions appears to be substantially less efficient than spending the same time with intact to-be-learned material. The experimental results are clouded by the failure to achieve a prequestion effect in the same conditions for which one was found in Experiment 6. However, the size of the prequestion effects in Experiments 6 were without exception below 12%, whereas in Experiment 7, with the same materials, the effect of double presentation of the to-be-learned material ranged between 18% and 34%.

With the above result in hand, we can state with a high degree of certainty that the memorial

---

[14] A separate analysis that examined only the relevant prequestion group and the irrelevant statement group was performed in order to make the analysis more comparable to that of Experiment 6. The results of the analysis did not differ from those of the analysis presented in the main text.

advantage of prequestions over extended reading time that Peeck (1970) found for prequestioned material was caused by selective-attentional processes at the time of study.

The implication of the results of Experiment 7 is that students are better served by spending time with to-be-learned material than by answering prequestions about the material. If an instructor wishes to provide learning assistance to students in the form of a pre-reading task, the students will benefit more from receiving a condensed summary of the points the instructor wishes the students to grasp than they will from receiving a set of prequestions to ponder. It must be pointed out, however, that this result is applicable only to the learning of factual material. Different results may be expected to obtain for conceptual application questions if one can safely generalize from post-question studies (Watts & Anderson, 1971) or problem-solving studies (Needham & Begg, 1991) to prequestion effects.

The failure to replicate the results of Experiment 6 is somewhat perplexing. Not only was the pattern of results changed, but the prequestion effect essentially disappeared (although the pattern of means is consistent with an overall prequestion effect). The only significant difference between Experiment 6 and

Experiment 7 is in the nature of the control group. The control group in Experiment 6 received irrelevant prequestions, whereas the control group in Experiment 7 received irrelevant statements. Comparison of the actual means for Experiment 6 and Experiment 7 leads to the conclusion that performance was lower overall in Experiment 7, but particularly so in the prequestion group. Given the general robustness of the prequestion effect so far, the lack of prequestion effect in Experiment 7 must be regarded as a puzzling anomaly.

CHAPTER 6 - GENERAL DISCUSSION

The experiments of Chapters 4 and 5 help delineate the scope and limitations of the effects of prequestioning. In this chapter, I will attempt to summarize the theoretical impact of the experiments and propose a potential explanation of the prequestion effect. A discussion of the practical implications of the experiments can be found at the end of the chapter.

Theoretical Implications - Inadequate Theories

The experiments described in Chapter 4 were designed to test the dominant explanation of the effect of prequestions. The conclusion to be drawn from the experiments is clear: any explanation of the prequestion effect that relies solely on the selective displacement of attention from unprequestioned material to prequestioned material is incomplete. If selective displacement of attention is prevented through the use of a between-groups manipulation of prequestioning, a significant prequestion effect still exists.

While asserting that explanations based on attention displacement are incomplete, I do not mean to imply that selective displacement of attention is unimportant. The results of Experiment 7, in combination with those of Peeck (1970), suggest that selective displacement of attention can have a powerful effect. In Peeck's experiment, subjects who were prequestioned outperformed subjects who were given extra study time in cued recall of the prequestioned material. In Experiment 7, this relationship was reversed. The major difference between the two experiments is that Peeck's subjects were able to selectively displace their attention from unprequestioned material to prequestioned material, whereas the subjects in Experiment 7 were not. Thus, Peeck's subjects benefited substantially from the ability to selectively displace attention.

Because selective displacement of attention cannot be the sole determinant of the prequestion effect, I designed the experiments in Chapter 5 to examine other potential explanations of the prequestion effect. These included schema theory, partial target activation, subsumption theory, and various forms of item-specific or relational information.

The results from the incorrect weak items in Experiment 6 (e.g. The Capital of Australia is Paris) argue against schema-theoretic explanations of the prequestion effect. The full-strength effect of prequestioning on memory for those items precludes the placement of limitations on the acceptable content introduced into a schema. Because such limitations are a feature of many versions of schema theory, those theories would have to be revised in order to account for the experimental results.

It could be argued that the experimental materials used in this paper, because of their simplicity, preclude the operation of schemata. This may be true. However, it should be pointed out that schema-based explanations have been offered for sets of experimental materials simpler than the materials used in the current experiments. For example, Peeck (1982) asked Dutch subjects to produce as many American presidents as they could, before studying a list containing a mix of presidents and states. Subjects who produced the names of US presidents prior to study showed better recall of presidents than did subjects in a control group. The results from the experiment have been interpreted as support for schema theory (Mannies et al., 1989).

Partial target activation is another explanation of the prequestion effect that seems to be ruled out by the experiments of Chapters 4 and 5. If partial target activation is a significant contributor to the prequestion effect, there should have been measurable differences between the prequestion effects for the correct answers, the incorrect strong answers, and the incorrect weak answers in Experiment 6. That result is predicted because the correct and incorrect strong answers should have been activated much more strongly by prequestioning than the incorrect weak answers. The absence of differences in the size of the prequestion effect between those three item conditions suggests that partial target activation is unimportant.

Results from Experiments 2 and 3 also suggest that partial target activation is unimportant. Those two experiments provide evidence that answers that very few people know before the experiment (i.e. rare knowledge items) benefit from prequestioning at least as much as answers that many people know before the experiment (i.e. common knowledge items). If partial target activation were important, one would expect the common knowledge items to benefit more from prequestioning than the rare knowledge items.

Theories that rely heavily upon cue-specific information also appear to be incapable of explaining the prequestion effect in cued recall. This can be seen most clearly in the results of Experiment 6, in which a prequestion advantage in cued recall was not found for unrelated question-answer pairs. Experiment 6 can be considered roughly analogous to the generation effect experiments by Graf (1980), in which subjects generated meaningful or non-meaningful sentences (see Chapter 3). Graf's results showed no generation effect for cued recall of the nouns of the non-meaningful sentences given the verbs as a cue; Experiment 6 showed no prequestion effect in cued recall of the answers to the unrelated study items, given the question as a cue. When the results with the unrelated items of Experiment 6 are combined with the results of Experiment 4, which demonstrated a prequestion advantage in cue identifiability, it can be seen that cue identifiability is not sufficient to establish a prequestion effect. Consequently, theories that place a heavy emphasis on cue identifiability (Subsumption Theory, Organization-Redintegration) do not provide a satisfactory framework for discussing the prequestion effect.

Theoretical Implications - Viable Candidates

Despite the inadequacy of cue-specific information to account for the prequestion effect, the distinction between item-specific and relational information (Hunt & Einstein, 1981) may still prove to be fruitful. For example, it is possible that target-specific information supports the prequestion effect. The results of Experiment 4 showed a weak effect of prequestioning on target recognition. Because recognition is principally sensitive to item-specific information (Hunt & Seta, 1984), the superior recognizability of prequestioned targets relative to unprequestioned targets is almost certainly the result of greater item-specific information about those targets. However, as noted in Chapter 5, the size of the effect in target recognition was quite small relative to the size of the prequestion effect in cued recall. Thus, it is unlikely that target-specific information alone is the determinant of the prequestion effect in cued recall.

The influence of prequestioning on cue-target relational information was not directly tested in any of the experiments of this thesis. However, the failure of cue-specific information to account for the prequestion effect, and the weakness of the prequestion effect in target recognition, suggest that cue-target relational information may be the major determinant of the prequestion effect. It must be noted, however, that prequestioning is not sufficient to establish relational information between a priori unrelated items. Prequestioning appears to be limited to drawing attention to pre-existing relations between items. Evidence supporting this claim comes from Experiments 6 and 7, in which no prequestion effect was found for unrelated cue-target pairs. Those results place prequestioning in a different class from such techniques as interactive mental imagery, which has been shown to be sufficient to establish relational information between a priori unrelated items (Begg, 1982).

It appears, then, that an adequate relational theory of prequestion effects must capitalize on pre-existing relationships between the prequestion and the answer without establishing new relationships. Furthermore, relational information (or at least the potential for it) must presumably be established during prequestioning, as the study phase in all experiments was identical for both prequestioned and unprequestioned subjects.

A potential explanation for the prequestion effect in cued recall is that prequestioning forces the subject to elaborate the prequestion (cue) in such a way that the presentation of a related answer (target) at study will result in the episodic encoding of pre-existing relations between the prequestion and the answer. This proposal is reasonable when one takes in to account that the prequestion will force the subject to search memory for prior knowledge of the answer to the prequestion. Such a search is likely to employ a variety of potential retrieval cues derived from the prequestion, which will then form part of the memorial record of the prequestion. If the presentation of the study statement results in contact with the memory trace of the prequestion, the derived retrieval cues may serve as relational mediators between the prequestion and the answer. However, if the answer is unrelated to the prequestion, the retrieval cues derived in response to the prequestion are unlikely to be capable of relating the prequestion and the target.[15,16]

---

[15] Note that this is not a schema theory. Although the theory proposes that prior knowledge is activated in order to support the prequestion effect, no reference to an abstracted organizing structure is required.

[16] The proposal that an unsuccessful retrieval attempt might nonetheless leave behind a useful residue was anticipated by Slamecka and Fevreiski (1983) in the generation effect literature. Slamecka and Fevreiski asked subjects to generate words according to an "opposite" rule (e.g. hot-c___). They chose relatively unfamiliar opposite pairs, and provided either few letters of the response word (pursue-a_____) or many letters (pursue-av__d) in order to manipulate the probability of succeeding with a particular word pair. After the attempt to generate was made, feedback was provided to insure that the correct completion was always processed. In free recall, subjects uniformly benefited from the attempt to generate the response word, regardless of whether the generation attempt failed or succeeded. In a second experiment, Slamecka and Fevreiski found that under recognition testing, words that subjects failed to generate were remembered better than read words, but were less memorable than words that the subjects had successfully generated. Slamecka and Fevreiski argued that failed generation attempts really represent partial generation successes, i.e. appropriate semantic attributes are generated, but lexical access to the word is not achieved until the word is presented.

Supporting Evidence

Evidence supporting the above proposal can be found in some of the original work on prequestioning. Berlyne's (1954) experiment involving familiar and exotic animals (e.g. the sea mat) showed that prequestioning subjects on familiar animals produced a greater prequestion effect than prequestioning subjects about less familiar animals (see Chapter 2). This result is predicted by the currently proposed theory, because the less familiar animals would not have been able to provide as many derived retrieval cues during prequestioning, and therefore would have provided fewer potential mediators between the cue and target at study.

Another result of Berlyne's (1954) experiment was that if a subject failed, at study, to recognize a particular statement as providing an answer to a previous prequestion, the subject was less likely to succeed with that item on the cued recall test. This result is also predicted by the cue-elaboration theory, because a study item that fails to make contact with the memory trace of the prequestion will also fail to gain access to the potential mediators the prequestion could have provided.

In addition to explaining the benefits of prequestioning in cued recall, the proposed theory has the additional benefit of explaining the cue recognition advantage enjoyed by prequestioned items. The elaboration of the cues provided by the prequestion phase of Experiment 4 would result in increased discriminability between the studied cues and the unstudied cues presented during the test phase of the experiment.

It is not sufficient merely to possess the potential for mediation; the mediators must be activated in order to be effective. This can be shown in two ways. First, given the random assignment of subjects to conditions in the experiments reported in this thesis, it must be assumed that the set of mediators available to subjects who studied prequestioned material was also available to subjects who studied the same material without prequestions. Thus, the difference lay in the activation of the mediators, rather than in their presence in the subject's knowledge base. A similar kind of evidence comes from the research by Pressley et al. (1990) that showed that merely judging a prequestion for comprehensibility did not result in a prequestion effect. The activation of semantic knowledge, rather than its mere existence, appears to be critical in the establishment of the prequestion effect.

It is not clear at this point how the

activation of mediators at study increases relational information in memory. It might be the case that the relational information is strictly episodic in nature; i.e. it is contained in the memory record of the study episode. On the other hand, it could also be the case that the simultaneous activation of the cue, the mediators, and the target results in the strengthening of pre-existing links in a semantic network. Perhaps both mechanisms play a role in increasing relational information. The current set of experiments do not provide a way of separating the semantic contributions to relational information from the episodic contributions.

Contradictory Evidence

Despite the success of the proposed explanation for the prequestion effect, a few contradictory results need to be explained.

The first experiments in this thesis to produce problematic results for the proposed explanation of the prequestion effect were Experiments 2 and 3. Both of these experiments showed that the prequestion effect for rare knowledge items was at least as big as the prequestion effect for common knowledge items. This apparently contradicts the results from Berlyne (1954), which showed that familiar animals enjoyed larger prequestion effects than less familiar animals, with respect to the learning of basic facts about the animal. However, it must be recognized that common and rare (or familiar and less familiar) were defined in slightly different ways in the experiments in question. In Berlyne's experiment, familiarity was determined on the basis of the animal itself, not on the basis of the subjects' ability to produce the fact about the animal. In contrast, in Experiments 2 and 3, common and rare items were defined strictly in terms of the normed probability of the subject being able to answer the question from prior knowledge. Thus, in Experiments 2 and 3, a subject could have encountered rare knowledge facts about very familiar topics, which, according to the cue-elaboration theory, would have resulted in large prequestion effects. Therefore, the results of Experiments 2 and 3 are not problematic for the currently proposed explanation.

Another problem arising from the experiments in this thesis is found in the results of Experiment 4. The presence of a weak target-recognition advantage because of prequestioning is not as easily explained as the cue-recognition advantage. According to Hunt and Einstein (1981), features that are shared by the cue and the target are relational. Therefore, the target-relevant mediators activated by prequestioning are relational information, and should have little consequence for target-recognition, which is sensitive to target-specific information (Hunt & Seta, 1984). This reasoning does not affect the cue-recognition result as much, because one might reasonably suppose there to be a set of derived retrieval cues during prequestioning that do not actually provide relational information to the target. Those derived retrieval cues would then be considered cue-specific information.

There are two potential solutions to the target-recognition problem. First, one might argue that the prior presentation of the prequestion without the target encourages subjects to focus their attention on the encoding of the target at study. This might be supposed to result in increased target-specific information. A second possibility is that relational and item-specific information cannot sensibly be removed from the context of the test situation. That is, one could argue that some of the mediators that enable the subject to relate the cue and the target might also allow the subject to discriminate the target from lures presented in a recognition test. Thus the mediators would be used for relational purposes in a cued recall test, but for item-specific purposes in a recognition test.

The results of Experiment 6 provide yet another challenge to the proposed account of the prequestion effect. It should be the case that the mediators provided by prequestioning are more likely to be useful for correct or incorrect strong answers than for incorrect weak answers. This, however, does not appear to be the case. I can provide no satisfactory explanation of that result, so it must be left as a challenge to the proposed theory.[17]

Summary

The proposed theoretical account of the prequestion effect is able to explain a number of important aspects of the effect: the benefit for cued recall of related targets, the lack of benefit for cued recall of unrelated targets, and the benefit for cue-recognition. In addition, most of the anomalous results can be explained in a way that does not contradict the proposed theory. However, the results of Experiment 6 that show incorrect weak completions to benefit as much from prequestioning as correct or incorrect strong completions awaits an explanation.

---

[17] Similar puzzling results are found in the generation effect literature. Despite the usual superiority of semantic encoding over rhyme encoding in many memory tests, generation on the basis of a rhyme rule appears to be just as effective at increasing cued recall or recognition memory as generation on the basis of a semantic rule, e.g. synonymy (Slamecka & Graf, 1978). In addition, experiments by Carroll and Nelson (1993) appear to show that the difficulty of generation has little effect on the size of the generation effect. (The "Read" conditions are actually easy generate conditions in the majority of their experiments.)

Practical Implications

The experiments presented in this thesis appear to indicate that factual prequestioning may be of limited practical value. Because prior research demonstrates that factual prequestions reduce memory performance for material that is not prequestioned, their use is restricted to situations in which the material that will be neglected is of limited importance, or in which all material can be prequestioned. For those situations however, the results of Experiment 7 suggest that giving students more time to digest the important information will be more effective than giving them prequestions on the information.

It must be pointed out that the above claims, which were made with respect to factual prequestions, may not generalize to conceptual prequestions. Work by Frase (1971) with conceptual prequestions suggests that carefully worded prequestions can increase the amount of material retained. However, an important feature of Frase's work was that the conceptual prequestions forced subjects to attend to several propositions in the text. It may be presumed that conceptual prequestions that did not require attention to many propositions in the text would have had selective attentional effects equivalent to factual prequestions.

Further work is required to examine the potential of conceptual prequestions, particularly in comparison with a summary. However, suggestive evidence comes from an experiment by Needham and Begg (1991, Experiment 5). Half the subjects in the experiment were given a problem, and instructed to attempt to solve the problem. The other half of the subjects were given the same problem, with the solution, and asked to memorize the problem. All subjects had the solution to the problem explained to them by the experimenter. Subsequently, subjects were given a recall test for the content of the problem, and were also given a new problem to solve. The new problem was an analogue of the first problem, i.e. the principles used to solve the first problem could be usefully transferred to solve the second problem. The experimental results are shown in Table 10.

Table 10 - Needham & Begg (1991), Experiment 5

	Solve Study Prob	Memorize Study Prob
Solution Rate for New Problem	.704	.440
Recall of Training Story	.239	.608

The subjects instructed to solve the first problem were much more successful at applying the relevant principles to the second problem than were the subjects instructed to memorize the first problem. However, their memory for the details of the problem was worse. If one considers the presentation of the training problem to be analogous to a conceptual prequestion, then conceptual prequestions might be expected to encourage retention of relevant concepts, without benefit for the supporting details. This prediction is supported by a meta-analysis performed by Hamaker (1986) that showed that higher-order prequestions lead to better memory for the higher-order concept than do factual prequestions, while factual prequestions lead to better memory for the factual material than do higher-order prequestions.

In short, the educational value of conceptual level prequestions remains to be investigated, particularly in comparison with other educational tools, such as summaries. However, factual prequestions do not appear to offer educational advantages that are not exceeded by other methods of instruction.

REFERENCES

Anderson, R.C., and Biddle, W.B. (1975) On asking people questions about what they are reading. In G.H. Bower (ed.) *The psychology of learning and motivation: advances in research and theory. Vol. 9.* Academic Press: New York.

Anderson, R.C., and Pearson, P.D. (1984) A schema-theoretic view of basic processes in reading comprehension. In P.D. Pearson (Ed.) *Handbook of Reading Research.* Longman Inc.: New York.

Ausubel, D.P. (1968) *Educational Psychology: A Cognitive View.* Holt, Rinehart, and Winston, Inc.: New York.

Begg, I., (1978) Similarity and contrast in memory for relations. *Memory & Cognition, 6*, 509-517.

Begg, I. (1982) Imagery, organization, and discriminative processes. *Canadian Journal of Psychology, 36*, 273-290.

Begg, I., Martin, L.A., and Needham, D.R. (1992) Memory monitoring: how useful is self-knowledge about memory? *European Journal of Cognitive Psychology, 4*, 195-218.

Begg, I.M., and Nicholson, S. (1994) Semantic and episodic relations are experimentally dissociable and stochastically independent. *Canadian Journal of Experimental Psychology, 48*, 399-417.

Begg, I., and Snider, A. (1987) The generation effect: evidence for generalized inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 553-563.

Begg, I., Snider, A., Foley, F., and Goddard, R. (1989) The generation effect is no artifact: generating makes words distinctive. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 977-989.

Begg, I., Vinski, E., Frankovich, L., and Holgate, B. (1991) Generating makes words memorable, but so does effective reading. *Memory & Cognition, 19*, 487-497.

Berlyne, D.E. (1954) An experimental study of human curiosity. *British Journal of Psychology, 45*, 256-265.

Berlyne, D.E. (1966) Conditions of prequestioning and retention of meaningful material. *Journal of Educational Psychology, 57*, 128-132.

Boker, J.R. (1974) Immediate and delayed retention effects of interspersing questions in written instructional passages. *Journal of Educational Psychology, 66*, 96-98.

Boyd, W.M. (1973) Repeating questions in prose learning. *Journal of Educational Psychology, 64*, 31-38.

Bransford, J.D., and Johnson, M.K. (1972) Contextual prerequisites for understanding: some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*, 717-726.

Brewer, W.F., and Nakamura, G.V. (1984) The nature and functions of schemas. In R.S. Wyer, Jr., and T.K. Srull (Eds.) *Handbook of Social Cognition, Vol. 1*. Lawrence Erlbaum Associates: Hillsdale, NJ.

Bull, S.G., and Dizney, H.F. (1973) Epistemic-curiosity-arousing prequestions: their effect on long-term retention. *Journal of Educational Psychology, 65*, 45-49.

Burns, D.J. (1990) The generation effect: a test between single- and multifactor theories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 1060-1067.

Burns, D.J. (1992) The consequences of generation. *Journal of Memory and Language, 31*, 615-633.

Carroll, M., and Nelson, T.O. (1993) Failure to obtain a generation effect during naturalistic learning. *Memory & Cognition, 21*, 361-366.

Craik, F.I.M., and Lockhart, R.S. (1972) Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684.

Donaldson, W., and Bass, M. (1980) Relational information and memory for problem solutions. *Journal of Verbal Learning and Verbal Behavior, 19*, 26-35.

Einstein, G.O., and Hunt, R.R. (1980) Additive effects of individual-item and relational processing. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 588-598.

Erdelyi, M.H., and Becker, J. (1974) Hypermnesia for pictures: incremental memory for pictures but not for words in multiple recall trials. *Cognitive Psychology, 6*, 159-171.

Frase, L.T. (1967) Learning from prose material: length of passage, knowledge of results, and position of questions. *Journal of Educational Psychology, 58*, 266-272.

Frase, L.T. (1968a) Questions as aids to reading: some research and theory. *American Educational Research Journal, 5*, 319-332.

Frase, L.T. (1968b) Effect of question location, pacing, and mode upon retention of prose material. *Journal of Educational Psychology, 59*, 244-249.

Frase, L.T. (1968c) Some unpredicted effects of different questions upon learning from connected discourse. *Journal of Educational Psychology, 59*, 197-201.

Frase, L.T. (1971) Effect of incentive variables and type of adjunct question upon text learning. *Journal of Educational Psychology, 62*, 371-375.

Gardiner, J.M., Gregg, H.G., and Hampton, J.A. (1988) Word frequency and generation effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 687-693.

Gardiner, J.M., and Hampton, J.A. (1985) Semantic memory and the generation effect: some tests of the lexical activation hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 732-741.

Gardiner, J.M., and Rowley, J.C. (1984) A generation effect with numbers rather than words. *Memory & Cognition, 12*, 443-445.

Glanzer, M., and Adams, J.K. (1985) The mirror effect in recognition memory. *Memory & Cognition, 13*, 8-20.

Glover, J.A., Bullock, R.G., and Dietzer, M.L. (1990) Advance organizers: delay hypotheses. *Journal of Educational Psychology, 82*, 291-297.

Glover, J.A., Krug, D., Dietzer, M., George, B.W., and Hannon, S.M. (1990) "Advance" advance organizers. *Bulletin of the Psychonomic Society, 28*, 4-6.

Graesser, A.C. (1981) *Prose comprehension beyond the word*. Springer-Verlag: New York.

Graf, P. (1980) Two consequences of generating: increased inter- and intraword organization of sentences. *Journal of Verbal Learning and Verbal Behavior, 19*, 316-327.

Hamaker, C. (1986) The effect of adjunct questions on prose learning. *Review of Educational Research, 56*, 212-242.

Hamilton, R.J. (1985) A framework for the evaluation of the effectiveness of adjunct questions and objectives. *Review of Educational Research, 55*, 47-85.

Hirshman, E., and Bjork, R.A. (1988) The generation effect: support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 484-494.

Hunt, R.R., and Einstein, G.O. (1981) Relational and Item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior, 20*, 497-514.

Hunt, R.R., and Seta, C.E. (1984) Category size effects in recall: the roles of relational and individual item information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 454-464.

Jacoby, L.L. (1978) On interpreting the effects of repetition: solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior, 17*, 649-667.

Johns, E.E., and Swanson, L.G. (1988) The generation effect with nonwords. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*, 180-190.

Kane, J.H., and Anderson, R.C. (1978) Depth of processing and interference effects in the learning and remembering of sentences. *Journal of Educational Psychology, 70*, 626-635.

Kolers, P.A. (1975) Specificity of operations in sentence recognition. *Cognitive Psychology*, 7, 289-306.

Koriat, A. (1995) Dissociating knowing and the feeling of knowing: further evidence for the accessibility model. *Journal of Experimental Psychology: General, 124*, 311-333.

Koriat, A., and Goldsmith, M. (1994) Memory in naturalistic and laboratory contexts: distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General, 123*, 297-315.

Madigan, S.A. (1969) Intraserial repetition and coding processes in free recall. *Journal of Verbal Learning and Verbal Behavior, 8*, 828-835.

Mannies, N.J., Gridley, B.E., Krug, D., and Glover, J.A. (1989) Knowledge mobilization: schema activation or generation effect? *The Journal of General Psychology, 116*, 121-132.

McDaniel, M.A., Riegler, G.L., and Waddill, P.J. (1990) Generation effects in free recall: further support for a three-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 789-798.

McDaniel, M.A., and Waddill, P.J. (1990) Generation effects for context words: implications for item-specific and multifactor theories. *Journal of Memory and Language, 29*, 201-211.

McDaniel, M.A., Waddill, P.J., and Einstein, G.O. (1988) A contextual account of the generation effect: a three-factor theory. *Journal of Memory and Language, 27*, 521-536.

McElroy, L.A., and Slamecka, N.J. (1982) Memorial consequences of generating nonwords: implications for semantic-memory interpretations of the generation effect. *Journal of Verbal Learning and Verbal Behavior, 21*, 249-259.

Minsky, M. (1975) A framework for representing knowledge. In P.H. Winston (Ed.) *The psychology of computer vision*. McGraw-Hill: New York.

Nairne, J.S., Pusen, C., and Widner, R.L.Jr. (1985) Representation in the mental lexicon: implications for theories of the generation effect. *Memory & Cognition, 13*, 183-191.

Nairne, J.S., Riegler, G.L., and Serra, M. (1991) Dissociative effects of generation on item and order retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 702-709.

Nairne, J.S., and Widner, R.L.Jr. (1987) Generation effects with nonwords: the role of test appropriateness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 164-171.

Nairne, J.S., and Widner, R.L.Jr. (1988) Familiarity and lexicality as determinants of the generation effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 694-699.

Needham, D.R., and Begg, I.M. (1991) Problem-oriented training promotes spontaneous analogical transfer: memory-oriented training promotes memory for training. *Memory & Cognition, 19*, 543-557.

Nelson, T.O., and Narens, L. (1980) Norms of 300 general-information questions: accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning and Verbal Behavior, 19*, 338-368.

Paivio, A., Yuille, J.C., and Madigan, S.A. (1968) Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, 76, Monograph Supplement*.

Parker, E.S., Birnbaum, I.M., and Noble, E.P. (1976) Alcohol and memory: Storage and state dependency. *Journal of Verbal Learning and Verbal Behavior, 15*, 691-702.

Peeck, J. (1970) Effect of prequestions on delayed retention of prose material. *Journal of Educational Psychology, 61*, 241-246.

Peeck, J. (1982) Effects of mobilization of prior knowledge on free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 608-612.

Peeck, J., van den Bosch, A.B., and Kreupeling, W.J. (1982) Effect of mobilizing prior knowledge on learning from text. *Journal of Educational Psychology, 74*, 771-777.

Pressey, S.L.A. (1926) A simple apparatus which gives tests and scores - and teaches. *School and Society, 23*, 373-376.

Pressley, M., Tanenbaum, R., McDaniel, M.A., and Wood, E. (1990) What happens when university students try to answer prequestions that accompany textbook material? *Contemporary Educational Psychology, 15*, 27-35.

Rickards, J.P. (1976) Interaction of position and conceptual level of adjunct questions on immediate and delayed retention of text. *Journal of Educational Psychology, 68*, 210-217.

Rickards, J.P., and Di Vesta, F.J. (1974) Type and frequency of questions in processing textual material. *Journal of Educational Psychology, 66*, 354-362.

Rothkopf, E.Z. (1963) Some conjectures about inspection behavior in learning from written sentences and the response mode problem in programed self-instruction. *Journal of Programed Instruction, 4*, 31-45.

Rothkopf, E.Z. (1966) Learning from written instructive materials: an exploration of the control of inspection behavior by test-like events. *American Educational Research Journal, 3*, 241-49.

Rothkopf, E.Z., and Bisbicos, E.E. (1967) Selective facilitation of interspersed questions on learning from written materials. *Journal of Educational Psychology, 58*, 56-61.

Sagaria, S.D., and Di Vesta, F.J. (1978) Learner expectations induced by adjunct questions and the retrieval of intentional and incidental information. *Journal of Educational Psychology, 70*, 280-288.

Sanders, J.R. (1973) Retention effects of adjunct questions in written and aural discourse. *Journal of Educational Psychology, 65*, 181-186.

Schmidt, S.R., and Cherry, K. (1989) The negative generation effect: delineation of a phenomenon. *Memory & Cognition, 17*, 359-369.

Serra, M., and Nairne, J.S. (1993) Design controversies and the generation effect: support for an item-order hypothesis. *Memory & Cognition, 21*, 34-40.

Slamecka, N.J., and Fevreiski, J. (1983) The generation effect when generation fails. *Journal of Verbal Learning and Verbal Behavior, 22*, 153-163.

Slamecka, N.J., and Graf, P. (1978) The generation effect: delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 592-604.

Slamecka, N.J., and Katsaiti, L.T. (1987) The generation effect as an artifact of selective displaced rehearsal. *Journal of Memory and Language, 26*, 1987.

Skinner, B.F. (1958) Teaching Machines. *Science, 128*, 969-977.

Swenson, I., and Kulhavy, R.W. (1974) Adjunct questions and the comprehension of prose by children. *Journal of Educational Psychology, 66*, 212-215.

Thorndike, E.L., and Lorge, I. *The teacher's word book of 30,000 words*. Teacher's College Bureau of Publications, Columbia University: New York.

Tierney, R.J., and Cunningham, J.W. (1984) Research on teaching reading comprehension. In Pearson, P.D. (ed) *Handbook of Reading Research*. Longman Inc.: New York.

Watkins, M.J., and Sechler, E.S. (1988) Generation effect with and incidental memorization procedure. *Journal of Memory and Language, 27*, 537-544.

Watts, G.H., and Anderson, R.C. (1971) Effects of three types of inserted questions on learning from prose. *Journal of Educational Psychology, 62*, 387-394.

APPENDIX A

Further Analysis of Experiment 4 Target Recognition

Because of the marginal nature of the target recognition results in Experiment 4, more data were analyzed. The original analysis examined only those subjects who began the test phase with target recognition. However, the target recognition test was also given to subjects that began the test phase with cue recognition. In the analysis that follows, data from all subjects are analyzed, treating the order of test presentation as a between-subjects variable. It must be kept in mind that half of the data were collected on target recognition test that followed a cued recall test on the same items. It is therefore possible that the superior cued recall performance of the relevant prequestion group compared to the irrelevant prequestion group could exert an influence on the target recognition results.

There were 104 subjects whose data could be used when all data were analyzed. Data from one subject was not analyzed because his target recognition form was improperly completed. For each subject, mean target recognition ratings were calculated, contingent on presence or absence of the rated item during the study phase. The mean scores were subjected to a three-factor analysis of variance with study (studied, unstudied) as a within-subject factor, and prequestion type (relevant, irrelevant) and order (target recognition first, target recognition last) as between-subjects factors. The average ratings are shown in Table A1.

Table A1 - Target Recognition Ratings, Experiment 4

	TARGET REC FIRST		TARGET REC LAST	
PreQ Type	Studied	Unstudied	Studied	Unstudied
Relevant	4.397	1.402	4.472	1.237
Irrelevant	4.208	1.524	4.000	1.574
PreQ Effect	0.189	-0.122	0.472	-.337

The results of the analysis showed neither a main effect of prequestion type, nor a main effect of order, nor an interaction between the two [highest $F(1,100)=1.21$, $MSe=0.163$, ns]. As in the earlier analysis with the target recognition first data alone, there was a main effect of study [studied vs. unstudied: 4.262 vs. 1.440, $F(1,100)=2003.42$, $MSe=0.21$, $p<.001$]. There was also a significant interaction between prequestion type and study [$F(1,100)=19.55$, $MSe=0.21$, $p<.001$] and a marginal triple interaction including order [$F(1,100)=3.88$, $MSe=0.21$, $p=.051$].

An analysis of the data from target recognition last condition reveals no main effect of prequestion type [F(1,49)<1]. However there is a main effect of study [studied vs. unstudied: 4.231 vs. 1.409, F(1,49)=971.72, MSe=0.21, p<.001] and also a significant interaction between study and prequestion type [F(1,49)=19.87, MSe=0.21, p<.001]. Separate analysis of the studied and unstudied items shows that the effect of prequestion type is significant in both [studied: t(49)=3.80, p<.001; unstudied t(49)=2.99, p=.005], but the effects are in opposite directions. This is in the direction of the standard mirror effect in recognition memory.

The results in the data from the target recognition last condition suggest that the significant interaction between prequestion type and study in the complete analysis is due to the fact that the marginal interaction in the target recognition first data is supplemented by a significant interaction in the target recognition last data. The triple interaction results from the fact that the prequestion type by study interaction is of different magnitudes in the different test orders.

The combined analysis suggests that the effect of prequestioning on target-specific information might be higher than suggested by the analysis of the target recognition first data alone. However, the presence of an interaction between order and the other two independent variables probably indicates that the prior cued recall test influenced the target recognition results in the target recognition last condition. Therefore the results of this analysis must be excluded from consideration.

APPENDIX B

Stimuli for Experiments 6 & 7

The name of the small Japanese stove used for outdoor cooking is
HIBACHI - WOK - MICROWAVE - SUNSET

The name of the short pleated skirt worn by men in Scotland is
KILT - TARTAN - TUNIC - TABLESPOON

The name of the ocean that is located between Africa and Australia is
INDIAN - PACIFIC - ARCTIC - GEM

The name of deer meat is
VENISON - JERKY - HAM - PAINTER

The longest river in South America is the
AMAZON - ORINOCO - MISSISSIPPI - COSTUME

The name of the chapel whose ceiling was painted by Michelangelo is
SISTINE - VATICAN - LOUVRE - FROG

The game that uses a doubling cube is
BACKGAMMON - CRAPS - CHESS - THORN

The last name of the singer who popularized a dance known as the "Twist" is
CHECKER - DOMINO - SINATRA - PRAIRIE

The last name of the man who assassinated President John F. Kennedy is
OSWALD - BOOTH - BUNDY - SPEAKER

The name for a six-foot depth of water is
FATHOM - KNOT - MILE - ELBOW

The organ that produces insulin is the
PANCREAS - LIVER - LUNG - FOX

Socrates' most famous student was
PLATO - ARISTOTLE - MARX - MAST

The last name of the man who wrote the "Star Spangled Banner" is
KEY - FOSTER - WEBER - GOLF

The name of the furry animal that attacks cobra snakes is
MONGOOSE - FERRET - GERBIL - ATTENDANT

The city in which the United States Naval Academy is located is
ANNAPOLIS - WEST POINT - LAS VEGAS - HOOF

The last name of the author who wrote "Oliver Twist" is
DICKENS - KIPLING - ATWOOD - PIANO

The last name of the author of the book "1984" is
ORWELL - GOLDING - FINDLAY - JURY

The name of the submarine in Jules Vernes' "20,000 Leagues Under The Sea" is
NAUTILUS - NEMO - BOUNTY - ROBBER

Nairobi is the capital of
KENYA - NIGERIA - GERMANY - FISHERMAN

The last name of the author who wrote under the pseudonym of Mark Twain is
CLEMENS - SAWYER - THOREAU - GEESE

The European city in which the Parthenon is located is
ATHENS - ROME - BONN - CHIN

The Italian city destroyed when Mount Vesuvius erupted in 79 A.D. was
POMPEII - NAPLES - HOUSTON - LIME

The last name of the author of "Little Women" is
ALCOTT - BRONTE - LEACOCK - SAUCE

The unit of electrical power that refers to a current of one ampere at one volt is
WATT - JOULE - KILOGRAM - PEPPER

The last name of the man who wrote "Canterbury Tales" is
CHAUCER - MILTON - MOWAT - GALLERY

The last name of the astronomer who theorized that the earth orbits the sun is
COPERNICUS - GALILEO - HAWKING - SALAD

The name of the collar bone is the
CLAVICLE - SCAPULA - FEMUR - LEMON

The capital of New York is
ALBANY - BUFFALO - PITTSBURGH - INN

The last name of the first signer of the "Declaration of Independence" is
HANCOCK - FRANKLIN - BONAPARTE - STEAMER

The city in which Heathrow airport is located is
LONDON - GLASGOW - CALCUTTA - KETTLE

The name of the palace in London where the monarch of England resides is
BUCKINGHAM - WINDSOR - VERSAILLES - THIEF

The last name of the scientist who discovered radium is
CURIE - RUTHERFORD - PASTEUR - STAIN

The name of the crime in which a person purposely betrays his country is
TREASON - ESPIONAGE - MURDER - RATTLE

The largest island in the world excluding Australia is
GREENLAND - NEWFOUNDLAND - HAWAII - BANKER

The animal which runs the fastest is the
CHEETAH - GREYHOUND - COW - PEACH

The name of the legendary one-eyed giant in Greek mythology is
CYCLOPS - LEVIATHAN - ACHILLES - SNAKE

The name of the constellation that looks like a flying horse is
PEGASUS - ICARUS - VIRGO - SWAMP

The kind of cat that spoke to Alice in "Alice in Wonderland" is called
CHESHIRE - PERSIAN - LION - OVEN

The capital of Delaware is
DOVER - NEWARK - SEATTLE - MULE

The name of the company that produces Baby Ruth candy bars is
CURTISS - CADBURY - HOSTESS - BEGGAR

The people who make maps are called
CARTOGRAPHERS - NAVIGATORS - PILOTS - HARNESS

The capital of Finland is
HELSINKI - OSLO - MADRID - CORD

The last name of the most popular pin-up girl of World War II is
GRABLE - MONROE - BRINKLEY - FORK

The last name of the artist who painted "Guernica" is
PICASSO - DALI - BATEMAN - CELLAR

The last name of the female star of the movie "Casablanca" was
BERGMAN - BACALL - PFEIFFER - BARREL

The river spanned by the George Washington Bridge is
HUDSON - DELAWARE - NILE - CANDY

The name of the mountain range that separates Asia from Europe is
URALS - ALPS - ANDES - HAMMER

The country for which the Drachma is the monetary unit is
GREECE - TURKEY - CHINA - ELEPHANT

The city in which the Baseball Hall of Fame is located is
COOPERSTOWN - HARRISBURG - ANCHORAGE - PHOTOGRAPH

The name of the brightest star in the sky excluding the sun is
SIRIUS - POLARIS - MARS - LAWN

The name of the island on which Napoleon was born is
CORSICA - ELBA - ICELAND - DAMSEL

The last name of the European author who wrote "The Trial" is
KAFKA - NIETZSCHE - SWIFT - LIMB

Angel Falls is located in
VENEZUELA - BRAZIL - HOLLAND - INSECT

The last name of the composer who wrote the opera "Don Giovanni" is
MOZART - PUCCINI - BERLIN - TOOL

The name of the play in which Elwood P. Dowd is a character is
HARVEY - DEATH OF A SALESMAN - HAMLET - GARMENT

The name of the goldfish in the story Pinocchio is
CLEO - GEPPETTO - HOOK - FOREHEAD

The last name of the actor who portrayed Dr. Watson in the Sherlock Holmes series was
BRUCE - RATHBONE - SELLERS - COTTAGE

The name of a number two wood in golf is
BRASSIE - DRIVER - PUTTER - PRIEST

The capital of Australia is
CANBERRA - SYDNEY - PARIS - PENCIL

The last name of the poet who originally wrote "Don Juan" is
BYRON - SHELLEY - POE - INSTRUMENT

The last name of the twenty-first president of the United States is
ARTHUR - EISENHOWER - QUAYLE - MEADOW

The name of Gene Autrey's horse is
CHAMPION - SILVER - SECRETARIAT - CASH

The name of the town through which Lady Godiva supposedly made her famous ride is
COVENTRY - STRATFORD - WINNIPEG - PHYSICIAN

The last name of the first flier to fly solo around the world is
POST - LINDBERGH - WRIGHT - WEAPON

APPENDIX C

EXPERIMENT 6b

This experiment was conducted in order to make sure that the lack of a prequestion effect for the unrelated items in Experiment 6 was not due to the within-list manipulation of relatedness.

Method

Subjects. Subjects were 50 students from an introductory psychology course at the College of New Caledonia. They received course credit for their participation in the experiment. The experiment was conducted in individual sessions. Half of the subjects participated in each of the two experimental conditions.

Materials & Procedure. The materials and procedure used in this experiment were the same as those used in Experiment 6, except in the following details. First, subjects in this experiment were presented with the study list on a computer monitor in individual sessions. The rate of study presentation was the same as in Experiment 6. Second, unrelated completions were the only completions presented in the study phase. Thus the subjects' task was to attempt cued recall for

64 random nouns that had been presented with 64 fragments of general knowledge statements. As in the previous experiment, no warning was given during the prequestion phase that anything other than the correct answer would be presented during study. All other aspects of this experiment were the same as in Experiment 6.

Results

The number of items correctly recalled by each subject was computed, and the results submitted to a t-test. Prequestion type (relevant, irrelevant) was the sole between-subjects factor in the analysis. The analysis revealed no significant difference between the relevant prequestion group and the irrelevant prequestion group [relevant vs. irrelevant: 12.36 vs. 11.60, $t(48)=0.25$, $MSe=7.22$].