1-1-2003

Vocabulary acquisition and the generation effect

Barbara Anne Badgett
University of Nevada, Las Vegas

VOCABULARY ACQUISITION AND THE GENERATION EFFECT

by

Barbara Anne Badgett

Bachelor of Science
University of Nevada, Las Vegas
2000

A thesis submitted in partial fulfillment
of the requirements for the

Master of Science Degree in Educational Psychology
Department of Educational Psychology
College of Education

Graduate College
University of Nevada, Las Vegas
December 2003

UNLV

Thesis Approval

The Graduate College
University of Nevada, Las Vegas

November 18, 2003

The Thesis prepared by

Barbara Anne Badgett

Entitled

Vocabulary Acquisition and the Generation Effect

is approved in partial fulfillment of the requirements for the degree of

Master of Science in Educational Psychology

Examination Committee Chair

Dean of the Graduate College

Examination Committee Member

Examination Committee Member

Graduate College Faculty Representative

1017-53

ABSTRACT

Vocabulary Acquisition and the Generation Effect

by

Barbara Anne Badgett

Dr. Alice J. Corkill, Examination Committee Chair
Associate Professor of Educational Psychology
University of Nevada, Las Vegas

This study investigated two methods of inducing the Generation Effect and how its principles might be incorporated in vocabulary acquisition. Subjects attempted to learn 22 unfamiliar vocabulary words under one of three conditions: 1) definition-only control subjects repeatedly wrote each word and definition; 2) sentence generation subjects wrote each word and definition and then wrote a meaningful sentence using that word; and 3) definition generation subjects read the words embedded in context sentences and extrapolated and wrote the word meanings. Subjects were tested following a distracter task, 48-hours later, and again 21-days later. Significant main effects were found for encoding condition and time of test, with no significant interaction between the two. Sentence generation subjects performed better than the other two groups of subjects and subjects performed best at immediate recall, followed by 48-hour and then 21-day delayed recall. The results are interpreted with respect to a levels-of-processing explanation.

TABLE OF CONTENTS

ABSRACT iii

ACKNOWLEDGMENTS v

CHAPTER 1 INTRODUCTION 1
- Review of Literature 3
- Current Study 66
- Hypotheses 73

CHAPTER 2 METHODOLOGY 75
- Participants 75
- Materials 75
- Procedure 78

CHAPTER 3RESULTS 81
- Scoring 81
- Vocabulary Acquisition Analysis 84
- Analysis of Specific Hypotheses 86
- Perceived Difficulty Analysis 87

CHAPTER 4 DISCUSSION 88
- Hypotheses 89
- General Discussion 93
- Recommendations for Future Research 101

APPENDICES 105

REFERENCES 127

VITA 130

ACKNOWLEDGMENTS

WOW! What an experience. I have several bruises, as I continue to pinch myself. It is hard to believe that I have successfully completed this process. I could never have done it alone; there are so many people to whom I owe great thanks.

I'm not sure how to express my gratitude to my parents. Your pride and faith in me served as driving forces during the many times when I lost faith in myself. To my Mom and Daddy J.: without your love and emotional and financial support this would not likely have been possible. If it were not for us leaches, to whom you so affectionately refer to as your children, you would probably be able to live in a fancy house on a big hill. I appreciate, from the bottom of my heart, all of the sacrifices you have made for me. Thank you for taking such great care of Schatzi while I was working. I know that when she comes home to me Dad will miss the gift of sand that she leaves in his bed for him. Mom, thank you for your help with my vocabulary terms and for being so understanding of my absenteeism. You have been so supportive throughout this entire process. Daddy J., thank you for helping me develop into a thinker; I wonder if you might regret that at times. Daddy D., knowing how proud you are of me puts a bounce in my step. Thank you for listening to me when I was down and doubtful of whether I would succeed in this process. I am one lucky girl!

I appreciate the guidance I received from all of my committee members. You were all very patient with me. To my advisor and dear friend, Alice Jane, I would like to extend my gratitude for so very many things: 1) all of the weekends you sacrificed in

order to help me; 2) the many meals you fed me during those long weekend hours; 3) helping me learn so much while we worked side-by-side; 4) helping me keep my head up when I ran into trouble; 5) all of your reassurance that I would succeed (boy, did I need that); 6) your love and support; 6) letting me keep an "aforementioned" or two and a few "effectuate's"; and 7) simply being you. I am so honored to be able to work with you. I have admired you and your work since I was an undergraduate. Sometimes it is still hard to believe that I have access to such an amazing mind. YOU ROCK!!

I extend much thanks to my fellow graduate students. To my fellow Goats: thank you for letting me barn with you and for dealing with my incessant whining—"And this other thing about me…and one time I…." Lil Goat, thank you for helping me enter all that dreadful data. Please be sure to thank your mom for me. It is now your turn, and if you need any help at all…don't hesitate to ask Mim. As for my Medium Goat (AKA my mini-mentor)—I don't know how I would have ever completed this darn thing without your serving as such a great example. In reality, I don't know if I would have lasted in graduate school without your guidance. Thank you for teaching me how to design and conduct experiments, deal with subjects, keep track of data, and create an IRB and thesis presentation. How do other graduate students complete theses without MIMs? Goat friends are nice. Seb, thank you for helping me with my sentences and for spending so much time creating all the illustrations I never used, Oops. I really appreciate your time. Thank you too, Matty. You were a great help, and oh so patient, when I was experiencing so much difficulty trying to understand the process.

Now that this is finished perhaps I will not be MIA to all those about whom I care so much. Don't worry… I won't require that you call me "Master" for too very long.

CHAPTER 1

INTRODUCTION

The purpose of this study was to investigate multiple facets of the Generation Effect. The accepted definition of the Generation Effect is that there is better retention for material that is self-generated as compared to material that is simply read (Slamecka & Graf, 1978; Graf, 1980). In most studies of the Generation Effect, researchers have interpreted this definition to mean that subjects should create examples that are related to the to-be-learned material. For example, a subject would be asked to learn the definition of the word "sprat." The subject would be given the word and the definition and would be instructed to "generate" a meaningful sentence using the word properly. An alternative interpretation of the definition of the Generation Effect might require that subjects generate the meaning of to-be-learned material instead. For example, rather than giving the subject a term and definition, the subject could be given the term and a sentence in which the term is used properly. The subject's "generation" task, then, would be to generate the definition of the term in question. Although this latter method is not precluded by the generally accepted definition of the Generation Effect, no studies in which this approach has been used seem to appear in the literature. Therefore, this study was designed to compare two separate methods of inducing the Generation Effect.

Even before Slamecka and Graf's (1978) delineation of the phenomena, numerous researchers attempted to effectuate the Generation Effect with different types of input

tasks and materials, including paired associates (e.g., Bobrow & Bower, 1969; Hirshman & Bjork, 1988; McElroy and Slamecka, 1982; Slamecka & Graf, 1978), unrelated word-word and word-nonword pairs (e.g., McNamara & Healy, 1995), sentence completion tasks (e.g., Anderson, Goldberg, & Hidde, 1971; & Ghatala, 1981), and sentence generation tasks (e.g., Dempster, 1989; & Graf, 1980). The Generation Effect has been obtained with encoding tasks (read or generate) as within-subject (e.g., Graf, 1980; Hirshman & Bjork, 1988; McElroy and Slamecka, 1982; & Slamecka & Graf, 1978) and between-subject variables (e.g., Anderson, Goldberg, & Hidde, 1971; Ghatala, 1981; Hirshman & Bjork, 1988; McNamara & Healy, 1995; & Slamecka & Graf, 1978). Several possible interpretations for the Generation Effect have been explored: 1) depth of processing, semantic or lexical activation (e.g., Anderson, Goldberg, & Hidde, 1971; Bobrow & Bower, 1969; Ghatala, 1981; Graf, 1980; Hirshman & Bjork, 1988; McElroy & Slamecka, 1982; Slamecka & Graf, 1978); 2) effort (e.g., Slamecka & Graf, 1978; McElroy & Slamecka, 1982); 3) cue-target relationship enhancement (e.g., Graf, 1980; & Slamecka & Graf, 1978) 4) cognitive procedures (e.g., McNamara & Healy, 1995; & Slamecka & Graf, 1978); and 5) multiple factors (e.g., Hirshman & Bjork, 1988).

The nature of the research conducted related to the Generation Effect is broad and complex. Some experiments may not appear to be directly related to the current study at first blush, but without exploring them the multi-facet nature of the Generation Effect would be less evident. "Multi-facet nature" in this context refers to the Generation Effect being evident with the use of various materials, tasks, and measures. Therefore, the following review of literature consists of five sections. First, experimental precursors to the Generation Effect are described. Second, the series of five experiments conducted by

Slamecka and Graf (1978), who coined the term "Generation Effect," are explained. Next, a number of studies designed to further elaborate the Generation Effect are discussed. The theoretical interpretation of the Generation Effect for each study described is presented along with the study itself. Following the descriptions of experiments related to the Generation Effect is a section on the effects of learning vocabulary from context. This literature is especially relevant to the aforementioned alternative definition of the Generation Effect offered in the current study where subjects were required to extract vocabulary word meanings from rich context sentences. The literature review concludes with a description and rationale for the current study. . The literature review is augmented by Appendix A, which presents an abbreviated description of all Generation Effect experiments included in the review.

Experimental Precursors to the Generation Effect

Considerable research supports the notion that generation tasks lead to better comprehension and retention than simply reading alone (Slamecka & Graf, 1978; Graf, 1980). One set of studies is particularly relevant. In this section, the work of Anderson, Goldberg, and Hidde (1971) is described in detail. This work is based on the work of Bobrow and Bower (1969), who, through a series of experiments using noun pairs, found evidence that generating, as compared to reading, lead to better comprehension which, in turn, facilitated recall. In another experiment, Bobrow and Bower (1969) found that creating sentences that were plausible continuations, elaborations, or implications of the action or state of affairs stated in an experimenter-provided sentence facilitated twice as much learning as did reading aloud the same experimenter-provided sentences three times.

With Bobrow and Bower's (1969) work in mind, Anderson et al. (1971), further investigated whether procedures that require the reader to comprehend words, rather than simply say them, would facilitate learning. In their first experiment, they compared the recall of two groups. The "no blank," or control, group was presented with 24 complete sentences. The "blank," or experimental group, was presented with the same 24 sentences with the last word left blank. Subjects in the "blank" group were asked to generate the last word of each sentence. Subjects in both groups read the sentences aloud during 4-second intervals in the study sessions. The blank group was asked to produce the missing word as they read. In the test sessions, all subjects were presented with the subject nouns of each sentence as retrieval cues and were asked to produce the last word in that particular sentence. The "blank" group outperformed the "no blank" group on all recall tests. With this evidence, Anderson et al. (1971), argued that a "blank" forces semantic encoding of the other words in the sentence and that semantic encoding is the precursor to learning.

In a second experiment, Anderson et al. (1971), added a backward association test to the experimental procedure. The backward association was included in order to ensure that greater recall by subjects in the "blank" condition was the result of last-word-in-the-sentence generation and not because of similarities between encoding and retrieval procedures. In the backward association test, subjects were asked to produce the subject noun of each sentence given the last word of the sentence.

Subjects worked with three lists of sentences. One third of the subjects started with each list. Half of the subjects received the backward test first and half received the forward test first. The "blank" group outperformed the "no blank" group regardless of

test format. Anderson et al. (1971) concluded that procedures that force meaningful processing of sentences facilitate learning.

Bobrow and Bower (1969) and Anderson et al. (1971), were likely the first to notice the Generation Effect. Subjects were required to produce (generate) information under controlled conditions. Subjects who generated outperformed subjects who did not. Slamecka and Graf (1978) further investigated this phenomenon. In the next section, Slamecka and Graf's (1978) work is described in detail.

The Generation Effect

In 1978, Slamecka and Graf published a series of five experiments that investigated whether a self-generated word would be better remembered than one that was externally presented. In this section, each of the five experiments is described in detail.

In Experiment 1, subjects were assigned to either a generate or a read-only experimental condition. Subjects in the generate condition viewed cards showing a stimulus word and the initial letter of a response word (e.g., rapid-f). Half of the generate subjects were self-paced and the other half viewed the cards at a 4-second presentation rate. Subjects were instructed to say the stimulus and response words once for each presentation. Subjects in the read-only condition viewed cards showing the stimulus and response words (e.g., rapid-fast). Half of the subjects were self-paced and the other half were timed at 4-second presentation rates. All subjects' memories of response words were tested with the same recognition test. Paired associates were presented under five encoding rules: 1) associate, for example, lamp-light; 2) category, for example, ruby-diamond; 3) opposite, for example, long-short; 4) synonym, for example, sea-ocean; and 5) rhyme, for example, save-cave. A list of 100 items, 20 per rule, was presented

individually on index cards. Cards were blocked by rule, whereby rules were presented together, with the order of rules varied across subjects. Subjects in the generate condition knew which rule they were supposed to use in order to complete each paired associate.

At recognition, for each of the 100-paired associates, subjects viewed the first word of the pair along with three potential associate words: one word was a response or appropriate target word and the other two were distracters. Using a mask to cover the other items, subjects were exposed to each set individually. They were instructed to circle the response word that was introduced or generated at the time of input and to rate their confidence in each forced choice by using a 5-point scale from 1 (no confidence) to 5 (high confidence).

Subjects in the generate condition performed significantly better than subjects in the read-only condition. There were no differences for timed versus self-paced and the magnitude of the generation effect did not vary across rules. Subjects who generated response words were able to more accurately and confidently recognize them across rules and at all paces.

In Experiment 2, Slamecka and Graf (1978) set out to investigate whether the Generation Effect would persist if subjects were exposed to both generate and read-only conditions. In this experiment, encoding task (generate, read-only) was a within-subject factor and all subjects viewed paired associates from the previously described five rules. The experiment also included a between-subjects factor: informed versus uninformed of the pending recognition test. Rule blocks of 20 were divided into subsets of 10 for generate and 10 for read-only presentations. At a 4-second presentation rate, all subjects were presented with half of the response words to read and half to generate. The same

recognition test was used for both groups, but the subjects were not asked to give a confidence rating; instead, they were asked to indicate, by a "G" or an "R", whether they had generated or read particular response words at the time of input.

The informed versus uninformed manipulation was inconsequential. In addition, there was no interaction between rules and generate versus read-only. In other words, the existence of the Generation Effect for items did not depend on the encoding rule used for the paired associates. Subjects were able to correctly recognize items with 74 percent accuracy for allocation of generate or read input. Thus, recognition and correct allocation to read or generate were related.

Slamecka and Graf (1978) explained that the results of Experiment 2 followed the same pattern as the findings in Experiment 1. Accurate recognition of items was related to the ability to recall the encoding condition. In other words, a subject's ability to correctly recognize target items was related to his/her ability to accurately remember the encoding task (read-only, generate) used for that item.

The aim of Experiment 3 was to investigate the effects of generation tasks upon stimulus words. Slamecka and Graf (1978) wished to examine the locus of the Generation Effect. They acknowledged that some might argue that the requirements of the generation task were such that they demanded a heightened level of attention to all aspects of the task. If that were true, subjects would need to more carefully attend to the stimulus word in order to elicit the intended generate word. By contrast, the read-only condition might not require more than superficial processing of the stimulus word by the subject. Therefore, the more elaborate processing involved in the generation condition would result in superior retention of both the stimulus and response words in the generate

condition. Under these circumstances, the term *Generation Effect* would not be an accurate one because the stimulus words had never been generated.

A 2 x 2 x 2 factorial design was used in Experiment 3, with generate versus read-only as a within-subject factor; stimulus versus response recognition as a between-subjects factor; and informed versus uninformed of test as a between-subjects factor. The stimulus materials included 66-rhyme paired associates (e.g., "save-cave"). Both groups were presented with half of the 66-rhyme items under the generate condition and the other half under the read-only condition. Half of the subjects were presented with the read-only condition first and the generate condition second. The presentation rate for all stimulus materials was four seconds. All subjects completed two recognition tests: one required recognition of stimulus words while the other required recognition of response words.

As in Experiments 1 and 2, subjects recognized generated response words significantly better than read response words. In fact, there was a large generation effect for response words, but not for stimulus words. Whether subjects knew of the impending recognition tests was irrelevant. The analysis of confidence ratings was conducted only on correctly recognized items. There was a significant effect of generate versus read-only and a significant interaction between generate versus read-only and stimulus versus response. The Generation Effect was more prevalent with response, and not stimulus, with respect to confidence ratings.

Slamecka and Graf (1978) concluded that responses do show a generation effect, but stimuli do not. This appeared to be true for recognition of stimuli and responses words and their corresponding confidence ratings. The results did not support the notion that a

generation situation leads to heightened attention to and processing of all elements involved. The memorial benefits appeared to apply only to items that were generated by the subject him or herself.

Through examination and interpretation of the results from Experiments 1, 2, and 3, the researchers were satisfied that the Generation Effect occurred with the use of recognition measures. Experiment 4 was designed to investigate whether the Generation Effect would occur with tasks requiring recall because recognition and recall are not always affected similarly (e.g., Brown, 1976; Tulving, 1976). Thus, Experiment 4 was designed to address the possibility that the Generation Effect was limited to situations where the copy cues were present and where the specific demands of recall tests might not bring out the Generation Effect.

In Experiment 4, generate versus read-only, rules, and trials were all included as within-subject factors. For each subject, half of the presentations were read and half were generated at a 4-second input rate. As this experiment required recall, rather than recognition, only three of the rule conditions remained appropriate: synonym, opposite, and rhyme. Subjects were presented with 20-items per rule (half read and half generated). All subjects were informed of the pending tests. Five alternating presentations and test trials were administered with 30-second distracter tasks between input and recall. At test, subjects were given 4-minutes to engage in free recall of as many of the response words as possible.

Subjects recalled significantly more of the generated items than the read items across trials. There was a reliable interaction of generate versus read-only across trials. In other words, on the first recall test, subjects recalled significantly more of the generated items

than the read items. By recall test number four, the difference in recall scores for generated versus read-only items had all but disappeared. With the results of this experiment, the researchers concluded that the Generation Effect manifests not only with recognition measures, but with free-recall measures as well. They asserted that this extended the Generation Effect's generality and showed that no cues were necessary in order to bring about the phenomenon. In addition, subjects appeared to learn more of the generated words from subsequent study trials as opposed to read words. Furthermore, subjects were less likely to forget generated words than read words.

Experiment 5 was designed as a counterpart to Experiment 3. Recall that in Experiment 3, subjects were tested to establish whether their stimulus word recognition would parallel their recognition of generate words. In Experiment 5, Slamecka and Graf (1978) intended to determine whether stimulus words involved in generation tasks would be better recalled than read stimulus words. In Slamecka and Graf's (1978) words, "The fact that stimulus recognition was not enhanced in that experiment [Experiment 3] cannot simply be assumed to hold for recall as well…."(p. 600).

Experiment 5 employed a 2 x 2 x 5 factorial design, with generate versus read-only as a within-subject variable, stimulus word versus response word recall as a between-subjects variable, and five study-test trials as a within-subject variable. The same stimulus materials as in Experiment 3 were used. Two types of cued-recall tests were used: one for stimuli and one for response.

The rules used were synonym, opposite, and rhyme with 10-generate and 10-read per rule for all subjects. All subjects were paced at 4 seconds, were presented with five alternating input and test trials, and were informed of the pending recall tests. Group one

was asked to recall stimulus words with the response words given as cues. Group two was asked to recall response words with the stimulus words given as cues. Both groups were given 5 minutes to complete each test.

Slamecka and Graf (1978) found a significant main effect for generate versus read-only conditions. Across trials, subjects were better able to recall response words from the generate conditions. The difference in recall of response words under generate conditions was more than three times greater than recall of stimulus words under generate conditions, although this difference was not supported statistically. Unlike the findings of Experiment 3, there was not a reliable interaction between generate versus read-only and stimulus versus response. The Generation Effect, however, persisted across all trials of the multitrial learning task. The results support the hypothesis that the Generation Effect extends to cued-recall situations. Importantly, although the data visually suggested otherwise, there was no significant interaction favoring recall of response words as opposed to stimulus words for generation pairs. This was inconsistent with the findings of Experiment 3.

In summary, Slamecka and Graf's (1978) five experiments clearly establish the existence of the "Generation Effect." The Generation Effect occurs: 1) when a word is generated under an encoding rule with the presence of a stimulus word; 2) with free and cued recall measures, cued and uncued recognition measures, and confidence ratings; 3) with associate, category, opposite, synonym, and rhyme encoding rules; 4) with paced or unpaced presentations; 5) with informed or uniformed pending memory tests; and 6) for generated words, but not stimulus words, when measured by cued recognition.

Slamecka and Graf (1978) provided seven possible explanations for the Generation Effect: 1) levels of processing; 2) stimulus-response relations; 3) previous recall from semantic memory; 4) increased effort; 5) excessive tagging of nodes, resulting in more access routes; and 7) response emission without copy prompts. They only discussed the first three possible explanations in detail. First, they addressed the qualitative principle that the deeper or more elaborate the processing, the better the memorial benefits (Craik & Lockhart 1972). That is, deeper processing is semantic in nature, whereas shallower processing of information is more concerned with the superficial features of input (e.g., acoustic or visual). Three problems with the plausibility of this explanation were identified: 1) differential attention to stimulus and response items; 2) performance on rhyme rule items; and 3) the idea that generation intrinsically entails deeper processing than reading. In the following section, each of these concerns is described briefly.

First, Slamecka and Graf (1978) suggest that if the levels of processing explanation were accurate, the stimulus words would receive at least as much processing attention as the response words. If the stimulus words were not processed sufficiently, then the appropriate response could not be obtained. As it is necessary for the stimulus to be encoded to at least the same depth as the response it evokes, it follows that the stimulus should enjoy memorial benefits similar to that of the response. Experiment 3 tested for the memorial benefits of the Generation Effect in the recognition of stimulus words. The results showed that the stimulus words did not experience the memorial benefit of the Generation Effect.

Second, Slamecka and Graf (1978) contend that the rhyme rule should have produced a relatively shallow level of processing as compared to the associate, category, opposite,

and synonym rules, because it deals with acoustic information. Under the levels of processing explanation, it would follow that the rhyme rule should not evoke the Generation Effect. In none of the experiments did the memorial benefits of the Generation Effect fail to surface for the rhyme rule. Slamecka and Graf (1978) explained, "The stability of the effect hardly invites a 'level' explanation" (p.602).

With respect to the third concern, Slamecka and Graf (1978) conjecture that the act of generation, regardless of encoding rule, might intrinsically necessitate a deeper level of processing than does the nearly automatic act of reading. They admitted that validating experiments would be necessary in order to give this idea substance, because they lacked any prior assessments of the processing depths characteristic of generation versus reading.

An additional explanation provided by Slamecka and Graf (1978), was that the generation task might have forced distinctive encoding of the relationship between stimulus and generated response words, whereas the act of reading the stimulus and response words would not necessitate any registration of that relation. Regardless of the fact that all subjects were informed of the operative rule that bound the words, it is plausible that the items that were read did not encourage the use of that information, as it was not necessary in order to complete the task of reading the paired-associates. Thus, the encoding of the paired-associates that were read might have lacked relational specificity. In Slamecka and Graf's (1978) words, "To the extent that such distinctiveness is a factor in memory, the Generation Effect might be accommodated" (p. 603). Slamecka and Graf (1978) further elaborated, "This notion is evidently a salient one, since it occurred independently to the authors and to the editorial reader." (p. 603).

The third explanation of the Generation Effect offered by the researchers was the idea that the initial recall, or generation of the response word, imparts memorial benefits of that same material on subsequent tests. They suggest that the act of generation is actually an instance of recall from semantic memory. That is, that in the experimental instances of generation, the subjects neither learned, nor created, anything new. By generating, subjects simply retrieved existing information from their knowledge base with the aid of encoding rules, stimulus words, and the first letters of response words-to-be-generated. The overt responses, or generations, then, are actually episodes that are later tested for retention. There were no recall-based episodes involved in the read-only tasks, as all responses were simply given. Thus, the memorial benefits enjoyed by subjects in the generation condition are due to their having been previously recalled. Eventually, Slamecka and Graf (1978) admit that this is hardly an explanation as it merely restates the phenomenon it is attempting to explain. "That is, that a generated word is better remembered than one that was read because it was generated (recalled)" (Slamecka & Graf, 1978, p. 603).

Slamecka and Graf (1978) established the existence of the Generation Effect with the use of paired associates. Subjects in read-only conditions were provided with stimulus and response words associated by experimenter provided rules (i.e., associate, category, opposite, synonym, and rhyme). Subjects in generate conditions were provided with stimulus words and the first letter of the response word and were to generate the response word associated with the stimulus word according to experimenter provided rules. They compared performance of subjects in read-only conditions with performance of subjects in generate conditions on measures of recognition of stimulus and response words, cued

recall of stimulus and response words, and free recall of response words. Generation advantages were found with encoding tasks as within-and between-subjects factors and by all measures employed. Slamecka and Graf (1978) explored several possible explanations for the Generation Effect advantage, (e.g., levels of processing, stimulus-response relations, and previous recall from semantic memory), but committed to none.

Slamecka and Graf's (1978) work led to further investigations of the phenomenon. In the next section, several studies designed to extend and elaborate the Generation Effect are described in detail.

Further Examination of the Generation Effect

Five sets of studies designed specifically to further explore the Generation Effect are included in this section of the literature review. For simplicity, the studies are arranged in chronological order. The studies described investigate the Generation Effect with a variety input tasks and materials, including: sentence construction tasks (Graf, 1980), sentence completion tasks (Ghatala, 1981), paired associates (McElroy & Slamecka, 1982; Hirshman and Bjork, 1988), and word-nonword pairs (McNamara & Healy, 1995). Several explanations for the Generation Effect phenomenon are explored in the following literature, including: 1) depth of processing, semantic or lexical activation (Ghatala, 1981; Graf, 1980; Hirshman & Bjork, 1988; McElroy & Slamecka, 1982); 2) cue-target relationship enhancement (Graf, 1980; Hirshman & Bjork, 1988); 3) cognitive procedures (McNamara & Healy, 1995); and 4) multiple factors (Hirshman & Bjork, 1988).

In 1980, Graf reported a series of five experiments designed to address a variation of Bobrow and Bower's (1969) comprehension interpretation of the Generation Effect. The comprehension interpretation is that generated material, as compared to material that is

simply read, is comprehended better and because comprehension is better, it is responsible for superior retention. Graf (1980) suggested that it was possible to view comprehension from an organizational perspective. The apparent memorial benefits of the Generation Effect might be a result of better comprehension, a matter of organization, or both.

Graf (1980) asserted that an organizational theory approach could account for the memorial benefits of the Generation Effect, because generating a sentence results in an enhancement in the integration or interword organization of the sentence. In Graf's (1980) words, "Comprehension of a sentence implies that the words of the sentence have come to form an integrated, well-organized, informational unit" (p. 317). Previous research had already established that better organization of to-be-remembered-information resulted in better recall of that information (e.g., Tulving, 1962; 1966). Therefore, it was likely that increased interword organization of generated sentences could account for the memorial benefits of the Generation Effect.

If this organizational explanation of the Generation Effect held true, then the effects of generating would be influenced by factors that affect the integration or interword organization of sentences. That is, because meaning is an important organizational dimension for words, it would follow that meaning would also affect the interword organization of sentences. Therefore, meaningful sentences would be easier to integrate then nonmeaningful sentences.

If generating a sentence results in increased integration or interword organization of that sentence, then the amount of this integration would depend on the ease with which the sentence is integrated. A meaningful sentence would be more easily integrated than a

nonmeaningful sentence. As a result, generating a meaningful sentence would lead to better performance than reading a meaningful sentence. In contrast, generating or reading a nonmeaningful sentence, which does not offer such a basis for integration or interword organization, would not produce memorial benefits. Graf (1980) presupposed that if a test that was sensitive to integration were used, then the size of the Generation Effect would interact with the meaningfulness of the input sentence, because meaningful, as compared to nonmeaningful, material would be better integrated during input. The five experiments conducted by Graf (1980) designed to test this explanation of the Generation Effect are described next.

The five experiments were similar and all used the same stimulus materials. Four sets of sentences were used: 1) meaningful read, 2) meaningful generate, 3) anomalous read, and 4) anomalous generate. The basic design included the presentation format or processing condition (read-only versus generate), as a within-subject factor and input material (meaningful or anomalous) as a between-subjects factor.

In the read-only presentation format, subjects were presented with meaningful or anomalous six-word sentences on a computer screen. The grammar of the sentences was: article (the), adjective, noun, verb (-ed), article (the), and noun. The anomalous sentences consisted of a random arrangement of the content words from the meaningful sentences; however, no meaningful arrangements of the words were formed. The two sets of sentences were used in read-only conditions. Each sentence appeared on the screen for eight seconds and subjects were instructed to "simply read the sentence out loud, loudly and clearly, exactly once without errors" (Graf, 1980, p. 319).

For the generate presentation format, a list of randomly arranged content words, with the subject noun clearly delineated, was shown on the screen for eight seconds. Subjects were instructed to generate the sentence in their heads, using the specified grammar, and then to say the sentence out loud. The generated anomalous sentences consisted of a random arrangement of the content words of the meaningful sentences, where no meaningful arrangements of the words were formed.

In all experiments, subjects were given practice trials. Subjects who were unable to complete input tasks without errors were excused from the experiment. All subjects in all experiments were informed of pending memory tests.

In Experiment 1, the input consisted of 16 critical sentences. Subjects assigned to the anomalous and meaningful conditions both received 8-read presentations and 8-generate presentations that were blocked in fours. The order of presentation was alternated between subjects. Following the input, subjects were given 6 minutes to complete cued recall tests that consisted of random listings of the verbs used in the 16 critical sentences. They were instructed to write the critical sentence or any part of the sentence that had contained each verb.

Significant main effects were found for materials (meaningfulness of input sentences) and processing conditions, with a significant interaction between the two. Specifically, subjects in the anomalous sentence condition performed similarly regardless of processing condition (read-only, generate). With respect to the meaningful sentence condition, generate subjects significantly outperformed read-only subjects. Subjects in the meaningful sentence condition, regardless of processing condition, outperformed subjects in the anomalous sentence condition, regardless of processing condition.

Experiment 2 was conducted in order to determine whether the absence of a generation effect for the anomalous condition was due to the extremely low level of recall.

Experiment 2 was identical to Experiment 1, except that subjects viewed all input sentences three times instead of just once. The Generation Effect appeared for subjects in the meaningful input condition, but not for subjects subjected to anomalous input. There was a main effect for materials, processing conditions, and an interaction between the two. The findings replicated those in Experiment 1 and extended to multiple presentations. The withstanding pattern of results and the substantial increase of performance after three study trials attested to the robustness of the Generation Effect. Although not identified by the author, this "robustness" might have been an artifact of spaced presentations of the input material.

A comparison of the data from Experiment 1, which employed single study trials, and Experiment 2, which offered three consecutive study trials before testing, presented an opportunity to examine the effects of study trials on recall in the meaningful and anomalous conditions. Graf (1980) suggested that the additional trials in Experiment 2 would result in a greater benefit to subjects in that experiment when compared to the performance of subjects in Experiment 1. A *t*-test between the size of the effect found in Experiment 1 and the size of the effect found in Experiment 2 showed only nonsignificant marginal benefits. Graf (1980) suggested that subjects in the meaningful sentence condition continued to benefit from generating, as opposed to reading, over study trials. Subjects in the anomalous condition also benefited from additional study trials. In order to more clearly assess the effects of processing conditions as a function of

study trials and to examine the influence of multiple tests, Graf (1980) conducted a third experiment.

Experiment 3 was similar to Experiment 2, except that subjects were tested after each of the three study trials. Graf (1980) used a 2 x 3 x 2 design with processing condition (generate or read-only) and trials as within-subject factors and materials (meaningful or anomalous) as a between-subjects factor. Analysis of the results revealed: 1) that subjects were better able to recall in the meaningful than in the anomalous condition; 2) increased recall performance over study trials; 3) a generation effect in the meaningful condition, but not the anomalous condition; and 4) the endurance of the Generation Effect over study trials in the meaningful condition. A repeated measures analysis of variance showed a significant main effect for material and trials, as well as interaction effects for material with trials, material with processing condition, and trials with processing condition. No other analyses reached significance.

Graf (1980) concluded that there was faster acquisition of meaningful material, as opposed to anomalous material, over learning trials. This was thought to be due to the ceiling effect experienced by subjects in the generate condition. In other words, subjects' combined recall performance for each material type was greater across trials for the read-only condition than the generate condition. The results of Experiment 3 essentially correspond to those of Experiments 1 and 2. There was no evidence for the Generation Effect in the anomalous condition over a range of performance levels. In the meaningful sentence condition, the Generation Effect endured over study trials and only diminished when recall neared the ceiling in the meaningful-generate condition.

One result of Experiment 3 was not consistent with the results of Experiments 1 and 2. In Experiments 1 and 2, there was a marginal increase in the magnitude of the Generation Effect over study trials. This was not the case in Experiment 3. Graf (1980) argued that the multiple test experiences awarded the subjects in Experiment 3 an opportunity to monitor recall performance and, perhaps, realize the ineffectiveness of reading as a learning strategy. This may have motivated the subjects to put forth more effort in the learning of the material, hence the differences between Experiments 1, 2, and 3.

Experiment 4 differed slightly from Experiments 1, 2, and 3. Here, Graf (1980) used two groups of 24 subjects, with half of each group receiving anomalous input and half receiving meaningful input. Half of each input type was under generate conditions and the other half was under read-only conditions. One group received one exposure to the input list and the other group received two consecutive inputs. Another change in Experiment 4 was the use of a subject-paced, word-pair recognition test, rather than a timed, cued recall test.

The recognition test required subjects to view a list of noun-pairs that all originated from the input materials. Some of the nouns pairs were identical to the input materials (intact pairs); others were not (broken pairs). Subjects viewed one pair at a time and were asked to indicate whether they had seen the pair in the input list. Graf (1980) expected that a subject's ability to recognize intact pairs and reject broken pairs would depend on whether the words were well organized during input. Thus, he expected performance to be better for subjects in the generate than in the read-only condition with

meaningful material. In addition, he predicted similar levels of performance between read-only and generate conditions for anomalous material.

For each subject a difference score was obtained for generate and read conditions. Analysis of the mean difference scores indicated: 1) better performance for meaningful pairs than anomalous pairs; 2) better overall performance for the group with two presentations; and 3) a generation effect on meaningful material for both groups, but no sign of the Generation Effect for anomalous materials. Subjects in the two-presentation group displayed better, though not significantly, performance for the read-only versus generate pairs. Analysis of variance with material and group as between-subjects factors and processing condition as a within-subject factor supported the difference score findings. Hence, there was a main effect for material and group as well as interaction effects for material with group and for material with processing condition.

Graf's (1980) findings for Experiment 4 supported his assumptions. The disadvantage in grouping a string of unrelated words was demonstrated in the main effect for material. The group effect illustrated an overall increase in performance with two exposures compared to one. The interaction of the materials with group suggested that subject performance was superior when the input material was meaningful. Graf (1980) explained that the interaction of material with processing conditions implies that generated sentences are more likely to be integrated than read-only sentences. This appeared to be true only for meaningful material, however, as there was little difference in performance between processing conditions with anomalous material.

The overall findings of the word-pair recognition tests in Experiment 4 concurred with those of the cued-recall tests in Experiments 1, 2, and 3. Graf (1980) interpreted the

consistent findings of the Generation Effect in meaningful conditions coupled with its absence under anomalous conditions to mean that meaningfulness of material is crucial to the Generation Effect. With the above four experiments he established that generating, as apposed to reading, tended to result in better semantic interword integration of words in sentences. In his fifth experiment, he investigated a different type of organization with respect to the Generation Effect.

In Experiment 5, Graf (1980) sought to establish whether the Generation Effect was solely localized to interword organization or whether it also influenced intraword organization. Graf (1980) defines intraword organization as the sensory and perceptual integration of a word, apart from its relation to other words. He hypothesized that generating might require more attention to individual words in anomalous sentences than in meaningful sentences, because anomalous sentences are not semantically constrained. He surmised that it might be inherent in generating that individual words are examined more closely and more often than in reading. With this, he presupposed that just as generating amplifies the interword organization of sentences, the significant amount of attention awarded to individual words under generate conditions might result in an increase in intraword integration. Reading, which does not necessitate as great an amount of attention to individual words, might have less of an effect on intraword organization.

The same input materials were used in Experiment 5 as in the previous four experiments. Subjects were randomly assigned to meaningful or anomalous sentence conditions and were presented with 32 sentences. Half of the sentences for each group were shown in generate format and half were shown in read-only format. After input, all

subjects were given a Yes-No recognition test. One- hundred-twenty-eight distracter nouns were used on the test. The distracters and all of the nouns that were included in the input sentences were presented on a computer screen one at a time. Subjects were instructed to indicate whether they recognized each word from the input materials by depressing the Yes button or the No button on the computer keyboard. The test was subject-paced and required approximately 10 minutes to complete.

A difference score was obtained for each subject in both conditions. A generation effect for both meaningful and anomalous input material was present. There was a main effect for processing condition, with no other significant effects.

These results were used to support the idea that the beneficial effect of generating, as opposed to reading, is not unique to interword organization of input sentences. In order for a subject to generate a sentence with experimenter provided words, in an experimenter specified sentence frame, the subject must closely attend to the given words and decide where to place them. Simply reading experimenter provided sentences did not demand such close attention to individual words. Graf (1980) explained that the data supported the idea that close examination or attention to individual words results in an increase in intraword organization.

Graf (1980) suggests that two patterns emerge from the results of the five experiments. First, there was an interaction between the input processing condition (read-only versus generate) and the meaningfulness of input sentences. Graf (1980) interprets this finding of the Generation Effect with meaningful input material, and its absence with anomalous input material, to illustrate the essential nature of meaningfulness in the Generation Effect. He contends that one consequence of

generating, versus reading, is a higher level of interword organization of input sentences. The apparent critical nature of meaningfulness in the Generation Effect led him to surmise that the interword organization involved was semantically based. When anomalous materials were used, the generate and read-only condition performance scores failed to diverge.

Second, a generation effect was found that was independent of meaningful input materials. Graf (1980) contends that the beneficial memorial consequences of the Generation Effect are not unique to interword organization of words in sentences. The act of generating a sentence requires that subjects pay close attention to individual words in order to properly place each of them into the experimenter-specified sentence frame. It is this close attention that produces an increase in intraword organization.

In summary, Graf (1980) concludes that generation offers memorial benefits by increasing the degree of interword and intraword recognizability of individual words. That is, generating likely involves more semantic processing than reading with both sentences and individual words—levels of processing.

Ghatala (1981) sought to investigate whether a non-generation task that induced meaningful representation and/or organization of to-be-remembered material would result in a retention advantage over read-only tasks and be comparable to generation tasks. Specifically, she sought to test whether cognitive operations involved in generating had special mnemonic value. If this were correct, then a non-generation task that induced meaningful representation and/or organization would not lead to a retention advantage similar to that of a generate task. In order to test these hypotheses, Ghatala (1981) compared the retention of material learned in read-only, generation, and judgment tasks.

The subjects in Ghatala's (1981) study were sixty 6th and 7th grade students (mean age = 12.5 years) who attended middle school in a middle-class urban neighborhood. An equal number of 6th and 7th graders were randomly assigned to each of the three conditions. Subjects were tested individually and were presented with sentences on 5 x 8 inch index cards at 5-second presentation intervals. Each of the sixty subjects was assigned to one of the three conditions: generate, judgment, or read-only. The subjects in the generate condition were presented with each of the 27 sentences with a blank in place of the last word. The subjects were asked to read the sentence aloud and offer the word that most obviously fit the blank as they read. Subjects in the judgment condition were provided with intact sentences with the last words underlined. They were to read the intact sentences and then judge, by saying "yes" or "no", whether the last word correctly completed the sentence. Subjects in the read-only condition were instructed to read each sentence aloud twice. All subjects were informed of the retention test to follow.

The materials used were taken from Anderson, et al. (1971) and consisted of 24 sentences. In addition to these sentences, Ghatala (1981) added three filler sentences. The three filler sentences were not tested, but were used in an attempt to maintain the attention of subjects in the judgment condition. The last word of a filler sentence was not determined by the preceding words of the sentence. Rather, the last word was a sentence low probability word (e.g., "Mothers usually make delicious apples").

Forward and backward cued recall tests were given to all subjects following the presentation of the sentences. For both tests, subjects were presented with 5 x 8 inch index cards, each showing a single cue. On the forward test, subjects were offered subject nouns from the input sentences as cues and were asked to say the last word of the

sentence from which the subject noun came. On the backward test, subjects were offered the last word of the input sentences as cues and were asked to respond orally with the subject nouns from the appropriate sentences. The order of the tests was counterbalanced across subjects in order to insure that the generation group did not enjoy memorial benefits from positive transfer from study task to test, since the two activities were more similar in the generate condition than in the read-only condition (Kane & Anderson, 1978).

Data were analyzed using a 3 x 2 analysis of variance with encoding condition (read-only, generate, judgment) as a between-subjects variable and type of test (forward, backward) as a within-subject variable. Newman-Keuls comparisons between encoding conditions showed that subjects in the generate and judgment conditions outperformed subjects in the read-only condition, but did not differ from one another.

The results supported Ghatala's (1981) hypothesis that a task that induces meaningful representation and/or organization of to-be-remembered material would lead to memorial benefits over a read-only task and be comparable to a generate task. Ghatala (1981) did not find support for the hypothesis that the cognitive operations involved in generating information from semantic memory have special mnemonic value beyond optimal processing of the material. Generating the last words of the sentences did not lead to better retention than judging the correctness of the last words of intact sentences. Ghatala (1981) claimed that the judgment task likely induces the same type of meaningful representation and/or organization of the input material as the generation task. Judgment and generation tasks did produce better recall than read-only tasks, presumably because

the read-only tasks could be completed with superficial analysis of the sentences—levels of processing.

Ghatala (1981) investigated whether a non-generation task that induced meaningful representation and/or organization of to-be-remembered material (judgment task) would result in a retention advantage over read-only tasks and be comparable to generation tasks. She found that the judgment task did result in this advantage over the read-only task with performance comparable to that resulting from the generation task. Thus, she concluded that cognitive operations involved in generating had no special mnemonic value, but instead claimed that the judgment task and generation task likely induced the same type of meaningful representation and/or organization of the input material. She essentially attributed the superior performance of subjects assigned to judgment and generation tasks over the performance of subjects assigned to read-only tasks to the read-only tasks having been completed with superficial analysis of the sentences—levels of processing. This is in direct contrast to what would be found by McNamara and Healy (1995) some 13-years in the future. McNamara and Healy's (1995) set of studies will be discussed later in this review. Next, McElroy and Slamecka's (1982) attempt to clarify the cognitive mechanisms responsible for the Generation Effect will be discussed.

With Graf's (1980) work, as well as other research, in mind, McElroy and Slamecka (1982) suggested two mutually exclusive interpretations for the Generation Effect. One category includes all interpretations that implicate semantic memory or a person's existing general knowledge (Graf, 1980, Anderson, Goldberg, & Hidde, 1978; Slamecka & Graf 1978). The other category does not implicate semantic memory as the locus of the effect. Instead it postulates generation superiority to be due to the intrinsic

differences between the tasks of generating and reading (Jacoby, 1978; Slamecka & Graf, 1978).

Seeking to investigate these two interpretations of the Generation Effect, McElroy and Slamecka (1982) ran a series of three experiments that examined the memorial benefits of generating material with which subjects could not have had prior knowledge bases or semantic histories: nonwords. In the following section, each of the three experiments is described.

A list of paired-associates, including word pairs and nonword pairs, was used in Experiment 1. Nonwords pairs were related by a formal letter transposition rule; word pairs were related by an opposites rule. Read-only and generated words could be elaborated semantically, but processing was required only of the generated words. In contrast, it was impossible to process nonwords semantically, regardless of read-only or generate conditions. Thus, McElroy and Slamecka (1982) claimed that a generation effect with nonwords would support the hypothesis that the generation act in and of itself somehow facilitates better retention. If, on the other hand, the magnitude of the Generation Effect were less with nonwords, then semantic memory would be implicated.

For Experiment 1, two sets of input material were used: 1) read material consisted of 60 items, 30-word pairs and 30-nonword pairs and 2) generate material consisted of 30-stimulus words with the first letter of the response words and 30-stimulus nonwords with the first letter of the response nonword. A 2 x 2 x 2 factorial design was employed, with item type (word, nonword) and task (read-only, generate) as within-subject factors, and presentation rate (timed, self-paced) as a between-subjects factor. A yes-no recognition test was used to measure retention. This test consisted of random presentation of 60

responses and 60 distracters. The retention of encoding task was measured for read-only versus generate and word versus nonword item types.

Subjects were presented with a deck of 60 cards containing all 60 pairs in random order. For generated items, subjects were to provide the opposites for stimulus words and to use a letter transposition rule to provide responses to nonword stimuli. The letter-transposition rule required the first three letters of the stimulus to be put in backward order after the consonant provided as the first letter (e.g., "dand—snad"). Subjects spoke the stimulus and response words aloud. Timed subjects were allotted 6 seconds to complete the task for each pair, only then could they move on to the next pair. Self-paced subjects were allowed to move to the next pair when they uttered the stimulus and response words on the present card.

After input, a 30-item recognition test was given in which subjects were to indicate whether a response word had occurred at input by circling a Y or N. At the same time, subjects used a scale from 1 (no confidence) to 5 (high confidence) to indicate their level of confidence as to the accuracy of their response. After the recognition test, subjects completed task allocation tests. Subjects were presented with two lists: words and nonwords. They were to indicate, by writing a "G" or an "R," whether they had generated or read the word at input. They also gave a confidence rating as to the accuracy of their response. Subjects were required to mark 15 items as "G" and 15 items as "R" in order to control for response biases.

Analysis of variance was performed with presentation rate (timed, self-paced) as a between-subjects factor, and task (read-only, generate) and item (word, nonword) as within-subject factors. There were significant main effects for all factors. Timed

subjects performed better than self-paced subjects. McElroy and Slamecka (1982) suggested that this was likely because the timed subjects were forced to spend more time on a pair than self-paced subjects. Recognition was better for words than for nonwords and better for generate than for read-only items. There was as significant interaction between task (read-only, generate) and item type (word, nonword). Specifically, a generation effect surfaced only for words.

Analysis of confidence ratings showed that timed subjects were more confident than self-paced subjects. In addition, all subjects were more confident on words and generate items. There was not a significant interaction between item type and task. Analysis of variance of the allocation data showed that timed subjects were more accurate than self-paced subjects and that for both presentation rates, judgment accuracy was better on words than on nonwords.

Results from analysis of the confidence rating for correctly allocated items were in concurrence with the above findings. Again, timed subjects were more confident than self-paced subjects and all subjects were more confident with words than with nonwords. Confidence was higher with generated items than with read-only items. The interaction of item type (word, nonword) and task (read-only, generate) was also significant. Specifically, subjects reported higher confidence levels for generated words.

McElroy and Slamecka (1982) subjected allocation data (whether subjects had read or generated an item at input) to an analysis of variance, which revealed that accuracy was significantly better for timed than self-paced conditions. Both timed and self-paced condition subjects were significantly more accurate in their task allocations with words than nonwords. Similar results were found when confidence ratings for correctly

allocated items were analyzed. Timed subjects were more confident than self-paced subjects, all subjects were more confident with words than nonwords, and confidence was higher for generated items. A significant interaction was found for item type (word, nonword) and task (read-only, generate). Specifically, subjects displayed a higher level of confidence for generated words.

McElroy and Slamecka (1982) interpreted the results of this experiment to mean that the Generation Effect is semantically based. Although they were able to support the semantic memory interpretation of the effect, they explained that the results were not sufficient to completely reject the alternative hypothesis—that generation superiority is due to the intrinsic differences between the tasks of generating and reading. In their words, "It is possible that the Generation Effect may have been found with nonwords if the overall level of performance had been higher, or if a different testing procedure had been used" (McElroy & Slamecka, 1982, p. 254). Based on this conjecture, they designed a second experiment to determine whether the results of Experiment 1 would generalize to a multi-trial free recall situation.

In Experiment 2, twelve subjects completed five study-test trials and the list of 30 words was omitted, as the focus of the experiment was on nonwords. McElroy and Slamecka (1982) used a 2 x 5 factorial design with task (read-only, generate) and trials as within-subject factors. The same nonword pairs and procedures were used as in the first experiment. Subjects were informed of the pending free-recall tests that were given after each of the five trials. Read-only and generate items were arranged in random order and occurred equally as often. Subjects were timed at 6 seconds per presentation. Subjects

counted backwards for 30 seconds after input and then were allocated 4 minutes to complete free-recall tests.

No generation effect was found for nonwords. Surprisingly, the read-only condition performed marginally better on all five tests. The effect for trials was significant, indicating that learning had taken place over the five learning trials. Trials did not interact with task. Therefore, there was no generation effect with nonwords on any of the trials or across trials.

The hypothesis that generation always results in better retention was not supported. McElroy and Slamecka (1982) explained two reservations they had regarding these findings. First, they questioned whether something particular to the letter transposition-rule they used, and not the nonsemantic nature of the nonwords, may have been responsible for their inability to find the Generation Effect with nonwords. Second, generated nonwords were not familiar to subjects: generated words were. Subjects were visually and acoustically familiar with generated words and generated words had actual meanings. Conversely, subjects had never previously seen or heard the nonwords. Though generate and read-only nonwords had equal acoustic exposure, the same cannot be said for visual exposure. Subjects never actually viewed generated nonwords during input. The researchers further explained that this inequity might have negated any benefit the generation task may have otherwise yielded. A third experiment was designed to address these reservations.

In Experiment 3, in order to address concerns about the letter-transposition rule employed in Experiments 1 and 2, both a letter-transposition rule and a rhyme rule were used. McElroy and Slamecka (1982) rationalized their use of the rhyme rule by the fact

that rhyme rules had effectuated the Generation Effect in previous experiments (Slamecka & Graf, 1978). In order to address the inequity of visual exposure of words and nonwords in Experiments 1 and 2, the researchers modified the input procedure. All input items were presented to subjects twice in succession: first, in a read-only or generate format, and then in a read-only format for all subjects. This procedure guaranteed visual exposure to the nonwords.

Sixteen subjects participated in Experiment 3. A 2 x 2 x 5 design was used with rules (letter-transposition, rhyme), tasks (read-only, generate), and trials (five) as within-subject factors. There were two lists of the 24 nonword pairs. The first time subjects encountered the stimulus materials, the 24 nonword pairs were a mixture of read-only and generate. The second time subjects encountered stimulus materials, all 24-nonword pairs were presented in the read-only format. Twelve pairs on each list were associated by rhyme rule and 12 pairs were associated by letter-transposition rule.

Subjects studied the first list of 24 nonwords blocked by rule (the 12 letter-transposition rule items were chunked together and the 12 rhyme rule items were chunked together), with read-only and generate items occurring randomly within the list. Items were exposed at 5 seconds per presentation. The second presentation of the list immediately followed the first and all stimulus materials were in the read-only format. In both cases, the subjects uttered the stimulus and response words aloud. After input, subjects counted backwards for 30 seconds and then completed a 3-minute written free recall test of the response words. The test cycle was completed five times with the same lists but the order of items varied across trials.

No generation effect was found with rhyme or letter-transposition rules on any trial or across trials. The only effect found was for trials, indicating that learning had taken place.

McElroy and Slamecka (1982) thought that their inability to find a generation effect could have been a result of the input procedures. Specifically, they suggested that having subjects read the response immediately after generating it may have in some way negated any potential benefits that might have otherwise been gained by the act of generation. In order to address this concern, they conducted an extension to Experiment 3. In this study, they tested 8 subjects with the same extended procedure, but with words under rhyme and opposite rules and a singe study-test trial.

An overall generation effect was obtained. There was a significant interaction for task and rule. A simple effects test revealed that there was a generation effect for rhyme pairs only. The means for generate opposites were higher than those for read-only opposites, but the difference did not reach significance. McElroy and Slamecka (1982) concluded that it was the nonsemantic nature of the material used in Experiment 3, and not the exposure conditions, that led to the absence of the Generation Effect with nonwords.

In McElroy and Slamecka's (1982) three-plus experiments, they were repeatedly unable to invoke a generation effect with nonwords. In contrast, a generation effect was easily found with words, but when the stimulus materials were not units in semantic memory, no generation advantage was obtained regardless of variations in testing procedures, encoding rules, and presentation procedures.

Ruling out several artifactual explanations for this outcome, McElroy and Slamecka (1982) concluded "with some confidence" that the memorial benefits that are usually associated with the generating are not a direct consequence of the act of generating. Thus, the advantages cannot be attributed to any intrinsic aspect of the generation task such as increased effort or greater congruity between study and test conditions. Instead, the researchers concluded that the lexical status of the stimulus materials was the crucial factor resulting in a generation advantage.

McElroy and Slamecka (1982) explained that, "The generation task may activate more of the attributes associated in semantic memory for generated responses than would be activated by the read task" (p.258). This would better enable a learner to access traces of generated than read-only responses, because generated traces would be more elaborate in the amount of information encoded during input—levels of processing. In order to further explain the locus of semantic-memory involvement in the Generation Effect, McElroy and Slamecka (1982) provided the following anecdotal report.

McElroy and Slamecka (1982) apparently found a generation effect for opposites and rhyme-ruled word-word pairs, but not for word-nonword rhyme-ruled pairs (e.g., jate-late). They failed to report any specifics of the experiment that produced these results, but nonetheless concluded that this finding was inconstant with the lexical activation view, "...unless one assumes that the presence of a nonword stimulus inhibits activation of the response word's node ... in semantic memory" (p.258).

In sum, McElroy and Slamecka (1982) concluded that memorial benefits of the Generation Effect are not an automatic consequence of the act of generating. Rather the benefits appear to rely on the involvement, at some level, of semantic memory.

Finding the interpretations of the Generation Effect incomplete, Hirshman and Bjork (1988), through a series of four experiments, investigated possible explanations for the Generation Effect. They concentrated on relational and item-specific theories. Based on research by Glisky and Rabinowitz (1985) and Gardner and Hampton (1985) they asserted that the relational explanation was not sufficient to explain the Generation Effect. Their first research question was whether a one-factor, item-specific explanation, such as the lexical-activation hypothesis, was sufficient to explain the effect. Their second research question was whether cued- and free-recall measures would be differentially affected by relational and item-specific factors.

Undergraduates served as subjects in all four experiments and the input materials were the same for Experiments 1, 2, and 3. Two lists of 14-word pairs were devised and served as the to-be-remembered materials. The 14 stimulus words were the same and were presented in the same order on both lists. For one list, the odd numbered stimulus words were paired with their first associates and the even numbered stimulus words were paired with their third associates. For the other list, the odd numbered stimulus words were paired with their third associates and the even numbered stimulus words were paired with their first associates. A 14-page booklet was constructed for each list with one pair typed on each page. The read-only groups received booklets that had the stimulus and response words printed on each page. The generate groups received booklets that had the stimulus words followed by fragments of the response words printed on each page. For both conditions, a line was printed below the stimulus-response pair for a written response. Cued and free recall measures were used. The cued recall test offered a randomized reordering of stimulus words with spaces provided for a

written response next to each word. For the free-recall measure, subjects were given a piece of paper and were asked to offer the response words they had studied at input. Word-search puzzles were used as distracter tasks.

Experiment 1 was specifically designed to ascertain the relative sizes of the Generation Effect as measured by free and cued recall. In a 2 x 2 x 2 mixed factorial design they included encoding task (read-only, generate) and type of test (free recall, cued recall) as between-subjects factors and associative strength (first, third) as a within-subject factor. The associative strength factor was included in order to determine if cued recall was more sensitive to relational factors than free recall. Hirshman and Bjork (1988) hypothesized that if cued recall was more sensitive, there would be better performance for first associates than third associates. In tests of free recall there would be a smaller advantage, if any, for recall of first associates compared to third associates. Eighty subjects participated in Experiment 1.

Subjects in the generate condition were told that they would be given a word and some letter cues of a related word. They were to figure out the second word and write both words in the space provided. The read-only subjects were told that they were to read the two related words presented and then to write both words in the space provided. All subjects were given 10 seconds per word pair and were informed that their memory would be tested, but details were not provided.

Subjects in the free-recall condition were given word search puzzles to work on for five-minutes. Subjects in the cued-recall condition were required to work on word puzzles for a total of 20-minutes. This discrepancy was included in order to avoid

potential ceiling effects on the cued-recall test. After subjects finished their condition specific word puzzles, they completed the appropriate recall test.

The first and last pairs studied at input were not used in analysis and served as primacy and recency buffers. Thus, the analyses were conducted on six first associate and six third associate pairs. There were significant main effects for encoding task (read-only, generate) and type of test (free recall, cued recall). More pertinent to the research question, the Generation Effect was larger in the cued-recall measure than in the free-recall measure. In other words, there was a significant interaction between encoding task (read-only, generate) and test type (free recall, cued recall). Planned comparisons were conducted and revealed that the Generation Effect was significant in cued recall, but not in free recall.

More interestingly, there was a significant interaction between test type and associative strength (first, third). Third associates were better recalled than first associates when measured by free recall, but first associates were better recalled than third associates when measured by cued recall. Planned comparisons were conducted and revealed that both differences were significant.

Experiment 1 was specifically designed to ascertain the relative sizes of the Generation Effect measured by free and cued recall. The analyses revealed that the cued-recall measures showed a significantly larger generation effect. The associative strength factor was included in order to determine if cued recall was more sensitive to relational factors than free recall. The analyses revealed that cued-recall measures were more sensitive to relational factors than free recall. Hirshman and Bjork (1988) did not believe that an item-specific explanation under the lexical-activation hypothesis was sufficient to

explain these findings. Therefore, they designed a second experiment to address their concerns about the larger generation effect obtained in cued-recall measures, as opposed to free-recall measures in Experiment 1. In Hirshman and Bjork's (1988) words, "If the size of the Generation Effect were to be sensitive to the level of performance, the Type of Test x Encoding Task interaction found in Experiment 1 might be due to the different performance levels of free recall and cued recall" (p. 487).

In order to address the aforementioned concerns, the retention interval was lengthened to 48-hours for the cued-recall measure in Experiment 2. This was done in an attempt to reduce the level of performance of the cued-recall group to the level of performance of the free-recall group. Thus, the methods in the Experiment 2 were nearly identical to those of Experiment 1, with one exception: the cue-recall group was dismissed immediately following the completion of the input task. They returned 48-hours later to complete the cued-recall measure.

Again, analyses revealed that the Generation Effect was larger on measures of cued recall as opposed to free recall, resulting in a significant interaction between type of test and encoding task (read-only, generate). Planned comparisons revealed a significant generation effect for subjects who completed the cued-recall measure, but not for subjects who completed the free-recall measure. Once again, as in Experiment 1, there was a significant interaction between type of test and associative strength. Planned comparisons revealed that subjects who completed the free-recall test better recalled third associates as compared to first associates. First associates were significantly better recalled than third associates by subjects who completed cued-recall tests.

There were significant main effects for test type and encoding task. Specifically, subjects performed better on measures of cued versus free recall. In addition, subjects in the generate condition recalled significantly more items than subjects in the read-only condition. Furthermore, there was an advantage of generate over read-only in free-recall measures for third associates. Results from Experiments 1 and 2 were then analyzed together.

Because one purpose of Experiment 2 was to recreate the free-recall condition in Experiment 1 in an attempt to determine if the modest nonsignificant advantage of generate over read-only tasks should have been interpreted as a real difference, a single analysis of variance was conducted on the free recall data from Experiments 1 and 2. Experiment (1, 2) and tasks were used as between-subjects factors and associative strength was used as a within-subject factor. A significant three-way interaction was observed. Significant main effects for encoding task and associative strength were also uncovered.

The three-way interaction between experiment, task, and associative strength was examined separately for free recall conditions in Experiments 1 and 2. Planned comparisons were conducted and revealed that the encoding task interacted with associative strength in Experiment 2, but not in Experiment 1. In Experiment 2, generate tasks resulted in a numerical advantage over read-only tasks for third associates only. There was a numerical advantage for both first and third associates in Experiment 1. Generated items were better free recalled than read-only items, as evidenced by the main effect of encoding task. Third associates were better recalled than first associates, as evidenced by the effect of associative strength.

Hirshman and Bjork (1988) concluded that the Generation Effect when measured via free recall was reliable. In an attempt to replicate and extend this conclusion, an extension of Experiment 2 was devised. The free-recall conditions from Experiment 1 and 2 were replicated, except that prior to input subjects were informed that they would be completing a free-recall test of the response items after input. The results replicated those from previous free-recall conditions. There was a nonsignificant advantage for generate over read-only conditions for both first and third associates. Results from the extension experiment and those from the free-recall conditions in Experiments 1 and 2 were then compared. For this analysis, type of experiment and encoding task were between-subjects factors and associative strength was a within-subject factor. Recall was significantly better in the extension experiment than recall in Experiments 1 and 2, but type of experiment did not interact with encoding task. In addition, for all three experiments, there was a significant generation effect and third associates were significantly better free recalled than first associates.

The researchers claimed that the combined results for free recall in Experiments 1, 2, and the extension, offer evidence that there are small generation advantages in free recall when a between-subjects design is employed. They interpreted the results from the two experiments and extension as evidence that a one-factor, item-specific theory, such as lexical activation, should be ruled out as sufficient explanation for the Generation Effect, and that:

> …any theory that accounts for the Generation Effect solely in terms of the assertion that the generation condition yields more of something vis-à-vis the to-be-recalled response term of a studied paired associate is untenable. It does not

matter whether the 'something' is lexical features, number of stored operations or whatever. (p. 489).

In other words, subjects in the generate condition may benefit from item-specific advantages inherent in item content, but generating must also result in some sort of memorial strengthening that does not occur through reading alone. Hirshman and Bjork (1988) suggest that this *something extra* is the relationship between the stimulus and response words.

According to this two-factor hypothesis, Hirshman and Bjork (1988) expected that both relational and item-specific factors contributed to the advantages of the act of generation for cued recall and that only item-specific factors contributed to the advantages of the act of generating in free recall. They explained that because third associates were better recalled than first associates in Experiments 1 and 2, this supported the idea that stronger relations between items in a pair does not necessarily facilitate free recall of response items. Indeed, it may even inhibit it. As a result, they predicted that it would be possible to eliminate the Generation Effect in free recall without removing the effect in cued recall. Experiment 3 was designed to test this hypothesis.

Experiment 3 differed from Experiments 1 and 2 in one way, instead of writing the stimulus and response words at input, subjects in both read-only and generate conditions were instructed to write only the response words. Hirshman and Bjork (1988) contended that this change in procedure would focus subjects' attention on response words. They hypothesized that this would benefit the read-only condition more than the generate condition on measures of free recall. In the generate condition, the response word is generated immediately before it is written. As a consequence, the actual writing of the

word should be of little benefit. By contrast, for subjects in the read-only condition, the act of writing the response word should substantially enhance the activation of response-word-specific features in memory. This manipulation should not result in the elimination of the Generation Effect in cued recall because it is dependent on both relational and item-specific factors.

For Experiment 3, 156 undergraduate students served as research subjects. The methods were the same as in Experiment 1, except during encoding subjects wrote only the response, not both the stimulus and the response. Significant effects for test type and encoding task were found. More pertinent to the specific hypotheses of Experiment 3, the Generation Effect surfaced in cued recall, but not in free recall. There was a significant interaction between encoding task and test type. Planned comparisons revealed that there was a significant effect of generation in cued recall, but no effect of generation in free recall. Interestingly, there was a slight, though nonsignificant, numerical advantage for the read-only group over the generate group on measures of free recall. As anticipated, test type interacted with associative strength. Third associates were significantly better free recalled than first associates; but on cued recall, first associates were significantly better recalled than third associates.

Finally, there was a significant interaction between encoding task (read-only, generate) and associative strength (first, third). Subsequent analyses revealed a nonsignificant effect between encoding task and associative strength in cued recall. The interaction between encoding task and associative strength, however, was significant in free recall. Planned comparisons revealed that subjects in the read-only condition better free recalled first associates, as compared to subjects in the generate condition. The same

was not true for free recall of third associates. Subjects in the generate condition free recalled greater numbers of third associates, which replicated the results of Experiments 1 and 2. This pattern of results appears to support the hypothesis that the Generation Effect may be eliminated under free-recall conditions. One limitation not addressed by Hirshman and Bjork (1988), however, was the restrictive nature of the stimulus materials. Nevertheless, Hirshman and Bjork (1988) conclude that the results of Experiment 3 support their two-factor theory of the Generation Effect.

Hirshman and Bjork (1988) admit that a problem with their account is that it is contrary to previous research findings published on the Generation Effect. Several experiments employing similar materials and generation tasks have reported large generation effects with free-recall measures. These studies relied heavily on within-subject designs in which generate and read-only tasks were mixed on one stimulus list. Hirshman and Bjork (1988) contend that when a whole-list approach (mixing generate and read-only tasks within one list) is incorporated, read-only and generate items compete for attentional resources which might interfere with one another at recall. Thus, effects revealed in favor of a generation task might actually be second order effects due to the whole-list design. Paired-specific effects of generation are those effects ascribable to generating a response word from a related stimulus, as opposed to reading a stimulus-response pair. Not accounting for whole-list effects, then, could result in the overestimation of the paired-specific effects of generation. Hirshman and Bjork (1988) explained, "...the use of within-subject designs can overestimate the size of the Generation Effect in free recall, and we contend this overestimation has obscured the critical interaction between type of test and encoding task reported in this article" (p.

491). Therefore, they conducted a fourth experiment designed to investigate this concern.

Materials and procedures for Experiment 4 were the same as for Experiment 1, with one major exception: encoding task (read-only, generate) was a within-subject factor. The difference between Experiment 4 and Experiments 1 and 2 allowed the researchers to directly test their suppositions that: 1) the Generation Effect is greater in within-subject designs than in between-subjects designs; 2) the crucial difference between Experiments 1 and 2 and previous free-recall experiments in the literature was the use of between-subjects designs; and 3) because of their confusion between pair-specific and whole-list effects of generation, previous authors systematically overestimated the average generation advantage as measured by free recall and thus concealed a theoretically important test type by encoding task interaction.

Forty undergraduate students participated in this experiment. Hirshman and Bjork (1988) employed a 2 x 2 mixed factorial design with test type (free recall, cued recall) as a between-subjects factor and encoding task (read-only, generate) as a within-subject factor. The within-subject design of the encoding task required the manipulation of the input lists. This was done by exchanging items between read-only and generate versions within the same list. Even-numbered items were switched between read-only and generate versions on each test. Four lists were created. On two of the lists, read-only items were in the odd-numbered list positions and generate items were in the even-numbered list positions. The opposite was the case for the other two lists. On two of the lists, read-only items were third associates and generate items were first associates with the opposite being true for the other two lists. Subjects were randomly assigned to work

with one of the four lists. This set-up resulted in the counterbalance of assignment of read-only versus generate task to list position, item type, and associative relation of that item across all subjects. In this experiment, unlike Experiments 1, 2, and 3, associative strength was not factorially crossed with encoding task and type of test.

The procedures for Experiment 4 mirrored those from Experiment 1, with three exceptions: 1) instead of being assigned to read-only or generate, subjects were assigned to one of four lists; 2) subjects were given examples of read-only and generate encoding tasks prior to the presentation of the input list; and 3) only 20 (instead of 40) subjects received free-recall tests, and only 20 (instead of 40) subjects received cued-recall tests. Retention intervals were the same as in Experiment 1: a five-minute delay for subjects in the free recall condition and a 20-minute delay for subjects in the cued recall condition.

Once again, there were significant effects for type of test and encoding task. Contrary to the results of Experiments 1, 2, and 3, there was not a significant interaction between encoding task and type of test. The difference between the generate and read-only conditions means remained larger when measured by cued recall, as opposed to free recall measures. Planned comparisons revealed that the effect of generation was significant in both cued-recall and free-recall measures.

Hirshman and Bjork (1988) concluded that within-subject designs effectuate larger generation effects that between-subjects designs. When measured by free recall, the difference between generate and read-only condition means was .28 in within-subject designs and .07 in between-subjects designs. When measured by cued recall, the difference between generate and read-only condition means was .38 in within-subject designs and .22 in between-subjects designs. In addition, consistent with results reported

by other investigators, the materials and procedures employed resulted in a large, reliable generation effect in free recall when a within-subject design was incorporated.

When comparing the results from Experiments 1 and 4 it is evident that the Generation Effect was larger in Experiment 4 (within subject) than in Experiment 1 (between subjects). This was likely a result of the means of the read-only group in Experiment 4 being much lower than the means of the read-only group in Experiment 1. This was true of both free- and cued-recall measures. The free-recall mean fell by .23 from Experiment 1 to Experiment 4, while the cued-recall mean fell by .18 from Experiment 1 to Experiment 4. This pattern was interpreted to mean that the larger generation effect found in the within-subject design was likely a result of decline in performance on read-only items.

The critical finding from Hirshman and Bjork's (1988) four experiments is the following: a one-factor theory is not sufficient to account for the pair-specific effects of generation. They viewed these results to be consistent with their two-factor theory.

Their results effectively demonstrated that the Generation Effect was much smaller when measured by free recall versus cued recall. In addition, first associates were better recalled when measured by cued recall. Furthermore, first associates were more poorly recalled with free-recall measures.

Hirshman and Bjork (1988) claimed that the above empirical facts could be explained by a theory that views generating as superior to reading as a study condition in two regards. The act of generation better activates features of the response term in memory (lexical activation, item-specific theory) and it strengthens the stimulus response relation (relational theory) in memory. They made two additional suppositions. First, they assert

that cued recall is sensitive to stimulus response strength, but is also facilitated by response activation. They find this to be consistent with the classic view that paired-associate learning involves response learning in addition to stimulus-response association. Therefore, by their account, the Generation Effect in cued recall could be attributed to differences in stimulus-response strength and response activation between read-only and generate items.

Second, they assumed that the free recall of responses was sensitive chiefly to response strength or activation and was constrained, rather than facilitated, by stimulus-response strength. This assumption is based, in part, on prior evidence that increased association reduces access to components of the unit. This was evidenced by the results of the analyses of free recall performance for first and third associates. In effect, the Generation Effect found in free-recall measures of response words is attributable to response strength or activation while it is simultaneously inhibited by stimulus-response strength. That is, in order to access a third associate, first associates must be effortfully inhibited. At free recall, third associate response strength would be greater than first associate stimulus response strength. This would result in increased likelihood of recall for the third associate. In other words, because the first associate was inhibited at input, it would be much more difficult to access at time of recall.

In summary, free recall in read-only and generate conditions was better for third associates than for first associates and the contrary was true for cued recall. This is because stimulus-response strength facilitates cued recall, but inhibits free recall. In addition, the Generation Effect is larger in cued-recall measures than in free-recall measures. This occurs because stimulus-response activation and response activation

advantages of generating over reading contribute to the Generation Effect in cued recall. Only response activation contributes to the Generation Effect in free recall. Furthermore, modifying the input task by directing read and generate subjects to write only the response words, instead of the stimulus-response pair, removes the Generation Effect in free recall, but not cued recall. According to the foregoing arguments, this procedural change should increase the response activation more in the read-only condition than in the generate condition. Based on Hirshman and Bjork's (1988) theory, this anticipated removal of the Generation Effect in free recall is because the effect depends exclusively on response activation. The Generation Effect should remain for cued recall, as the effect in generate conditions depends on both response activation and stimulus-response relation advantages. These particular results were obtained in Experiment 3.

Finally, Hirshman and Bjork (1988) suggest that results from within-subject designs are affected by whole-list factors that decrease performance on read-only conditions, thus increasing the appearance of the Generation Effect. Whole-list interactions of conditions on attentional and output processes can effectively be studied through a within-subject design. If, on the other hand, the focus of a study is on the pair-specific effects of generation, a within-subject design is not reliable, as it confounds the whole-list processes with pair-specific processes of interest. This can obscure the hypothetically important pair-specific effects of generation, such as the aforementioned interactions between encoding task and type of test.

Despite Hirshman and Bjork's (1988) painstaking explanation for the Generation Effect, researchers continue to search for the optimal rationale for the Generation Effect. McNamara and Healy (1995), for example, offer a procedural account (Crutcher &

Healy, 1989) of the Generation Effect. In this account, the Generation Effect works because subjects engage in cognitive operations that serve to connect a question to a target answer. Cognitive operations or procedures are defined as mental operations that link a cue, or question, to a target answer. They refute that the important factor is the subject's actual generation or production of target items at time of input. This procedural account differs from other accounts in that, as the name infers, it is process oriented, as opposed to item oriented. It does not focus on the nature of the generated items or the relationship between such items, but instead focuses on the actual process of generating. This explanation does not concern the effort or difficulty involved in such processes, but the nature of the cognitive operations applied or the interrelationship between the cognitive operations performed at study and test.

McNamara and Healy (1995) conducted two experiments designed to test this theoretical explanation for the Generation Effect as well as to extend the generation paradigm to the investigation of skill and knowledge acquisition. The first experiment required that subjects engage in math-oriented processing. Therefore, it will not be discussed. The second experiment employed word-nonword pairs as stimulus materials. In the following section it is described in detail.

In Experiment 2, thirty word-nonword pairs were constructed and presented under the guise that the subjects were to learn foreign vocabulary equivalents of English words. A 2 x 3 design, with training condition (read-only or generate) as a between-subjects factor and test (pretest, posttest, retention test) as a within-subject factor, was employed. Subjects were presented with the word-nonword pairs in a 10-minute initial study period.

Subjects were allowed to study the pairs in any manner they wished, but were not allowed to use paper or pencil.

Immediately following exposure to the lists, subjects were allotted 10 minutes to complete a pretest. Subjects were given a sheet of paper with the English word followed by a blank line and were instructed to write down as many foreign words as they could remember next to the English equivalent. The same procedures were used for the posttest, which followed training, and the retention test, which was administered one week later.

During the training period subjects in both conditions were situated in front of individual computer screens and given answer sheets with two or three blanks per line (depending on training condition). For the read-only condition, the English words were presented to the subjects on the computer screen, at which time they were to write that word in the first blank on the answer sheet. They were then to press the space bar, which caused the foreign word to then appear on the screen, at which time they were to write the foreign word next to the English equivalent. After writing the foreign word, they pressed the space bar again, and the next English word appeared and the process was repeated. Once subjects completed this process with all 30 word-nonword pairs, they were given a short break. After the break they were instructed to press the space bar twice and the list was presented again. Subjects repeated the process with the same list.

Generate condition subjects received slightly different instructions. After writing the English word presented on the screen, subjects were to attempt to remember and write, in the second blank, the foreign counterpart (which had not yet been shown to them on the screen). Subjects were required to write something on the second line, even if they were

not certain of their accuracy. After writing the English and foreign words, the subjects were to press the space bar, at which time the appropriate foreign word appeared. Subjects were to write that word in the third blank on their answer sheet if it was different from the answer they provided on the second blank. They, too, were given a break after completion of the entire list and then began the process again. Subjects in both training conditions were exposed to 14 blocks of training.

Immediately following training, subjects in the read-only and generate conditions completed a posttest. Subjects were given a sheet of paper with the English word followed by a blank line and were instructed to write down as many foreign words as they could remember next to the English equivalent.

Subjects returned one week after training and completed a retention test (identical to the posttest). Following this retention test, the experimenters explained to the subjects the concept of mnemonics and offered an example. Subjects were given sheet of paper with the list of word-nonword pairs in the same order as shown to subjects during the initial study period (before the pretest). Subjects were instructed to indicate, by writing "yes" or "no" next to each pair, whether they had used a mnemonic to try to learn the pair. If they indicated that they had used a mnemonic, they were to describe it. If they had used a mnemonic but could not remember the details, they were to indicate that they did not remember the mnemonic.

An analysis of variance was conducted with condition (read-only or generate) as a between-subjects factor and test (pretest, posttest, retention) as a within-subject factor. An orthogonal test using a contrast code reflecting learning was conducted between pretest and both successive tests. A second orthogonal test using a contrast code

reflecting forgetting was conducted between posttest and retention test. McNamara and Healy (1995) expected to find an interaction between training condition and test for the contrast between pretest and the subsequent tests.

There was no generation advantage for the pretest, but there was a sizable generation advantage for the posttest and retention test. Analysis of variance showed a main effect for test, which reflected an improvement in performance from pretest to posttest and forgetting from posttest to retention test. Single-degree of freedom tests showed that difference between pretest and each of the following tests (reflecting learning) was significant. The difference between the posttest and retention test (reflecting forgetting) was also significant. The interaction of training condition and test was significant. Simply put, a generation advantage was observed for the learning of nonwords.

Through exit interviews, the researchers discovered that many of the read-only subjects had tried to recall the foreign words before they appeared on the screen. Therefore, they conducted a separate analysis to ascertain whether there was a main effect of internally generating for the read-only condition. Via a median split, six high and six low internally generate subjects were identified in the read-only condition. The six high internally generate subjects reported internally generating an average of 97% of the time and the six low internally generate subjects reported internally generating 49% of the time.

There was a main effect for test. All subjects appeared to have learned between pretest and posttest. In addition, subjects appeared to have forgotten between posttest and retention test. The main effect for internally generating was not significant. Mean differences between high and low internally generating subjects, however, were in the

predicted direction: means for high internal generating subjects were greater than means for low internal generating subjects. There was no interaction between test and internally generating. Although there was no statistically significant difference between performance for high versus low internally generate subjects, McNamara and Healy (1995) suggest that the read-only subjects who most often internally generated the foreign words during training may have also used better memorization strategies than the read-only subjects who internally generated less often during training during initial list exposure.

With respect to the after-the-fact, self-report mnemonic use, McNamara and Healy (1995) report no significant differences in use of self-selected mnemonics between read-only and generate groups. Using a median split, McNamara and Healy (1995) uncovered a statistically significant difference between the mnemonic scores of the high and low internal generate read-only subjects. In essence, subjects in the read-only group who reported that they internally generated, also reported more mnemonic use, and tended to recall more nonwords.

To further investigate the locus of the generation advantage, McNamara and Healy (1995) used each subject's mnemonic score as a covariate. The interaction of training condition (read-only, generate) and test (comparing pretest and both successive tests) remained statistically significant, indicating greater learning by generate subjects regardless of mnemonic coding. An analysis of variance was then conducted with training condition as a covariate. There was a significant interaction between mnemonic score and test, indicating better performance for subjects who used mnemonic coding.

A mnemonic score variable (high or low mnemonic) was created via median split. A generation advantage surfaced for posttest and retention tests for subjects with low mnemonic scores. The generation advantage did not surface in any test for subjects with high mnemonic scores. This means that read-only subjects with high mnemonic scores (who also tended to internally generate) learned and retained comparably to all generate subjects. The act of generating appeared to improve performance only for subjects who did not automatically use mnemonic strategies.

In sum, there was a generation advantage for learning non-words in association with English words at posttest and a one-week retention test. Subjects in the read-only condition, who, in training, reported high internal generation of foreign words before they were exposed to them, recalled 27% more nonwords than subjects in the read-only condition who reported internally generating less often.

McNamara and Healy (1995) were able to find a generation advantage for the learning of nonwords, thus extending the generation advantage from instances of episodic memory to the learning of new material. They also found superior learning for subjects who used mnemonic strategies to learn the word-nonword pairs. They found some support for the prediction that the generation advantage would be eliminated for subjects who used mnemonic strategies.

The findings were interpreted to support a procedural account for the Generation Effect, "...according to which the critical factor leading to a generation advantage for learning new facts or skills is that cognitive procedures be developed during the learning process and that these procedures be reinstated at test" (McNamara & Healy, 1995, p. 162). McNamara and Healy (1995) define a cognitive procedure as a mental operation

linking a stimulus to a response. In the case of Experiment 2 in which subjects were to learn word-nonword pairs, the cognitive operation was the subject's process used to associate the two. By this definition, mnemonic links used to associate the word-nonword pairs would constitute a cognitive procedure.

Table 1: Competing Explanations for the Generation Effect

Explanations	Descriptions
Depth of processing:	Generated items are processed at a deeper level and thus enjoy superior retention.
Increased Effort:	The GE is due to an increased amount of effort or arousal for generated stimulus relative to stimulus that is read.
Cue-target relationship enhancement:	The connection from the stimulus term to the response term is strengthened by the act of generating.
Cognitive procedures:	The critical factor leading to a generation advantage is that mental operation linking a stimulus to a response be developed during learning and reinstated at test.
Multiple factors:	The GE is due to more than a single factor. Two or more of the above explanations may be responsible for generation advantages.

The above section discussed several explorations of the Generation Effect with multiple input tasks and measures. The research reviewed provided several explanations

for the phenomenon. These explanations are summarized in Table 1. Based on these studies, it seems that the Generation Effect could be related to the learning of new vocabulary. That is, the acquisition of new vocabulary is not dissimilar to the previously mentioned paired associate learning tasks. A new word is learned in the way that it relates to old, or already understood words. Thus, this final section briefly discusses the learning of vocabulary and the learning of vocabulary through context.

Learning Vocabulary from Context

One way in which people likely learn unfamiliar vocabulary words is based on the context in which the unfamiliar word is used (e.g., Crist & Petrone, 1977; Gipe, 1979; Jenkins, Stein, & Wysocki, 1984; Nagy, Herman & Anderson, 1985,). Research that explores this hypothesis has returned mixed results. Whether research subjects were specifically instructed to extract word meanings from context may play a significant role (e.g., Nagy, Herman & Anderson, 1985). In addition, the richness of the context in which the unfamiliar words appear may also influence a subject's ability to determine the meaning of the word (e.g., Nagy, Herman & Anderson, 1985). In several instances, in order to produce a more "authentic" learning-from-context environment (e.g., Gipe, 1979; Nagy, Herman & Anderson, 1985; Jenkins, Stein, & Wysocki, 1984), researchers have employed unnecessarily complex designs, which result in complicated interactions that even the researchers found difficult to explain. Nevertheless, three sets of studies, one of which employs a generation effect-type experimental condition, are worth mentioning.

In order to investigate the effectiveness of learning vocabulary from context, Crist and Petrone (1977) compared the effects of two different methods of teaching the

conceptual meaning of unfamiliar words: 1) offering the unfamiliar word and definition or 2) offering a context sentence with the unfamiliar word missing.

In their study, 70 undergraduates were to learn the conceptual meanings of 15 vocabulary words: 35 subjects were assigned to the definition condition and 35 subjects were assigned to the context condition. All subjects were exposed to training materials five times under spaced conditions. The definition group was presented with 75 cards, one at a time. Each card had a to-be-learned word on one side and its definition on the other. Subjects were to read the definition and then turn the card over to reveal the corresponding word.

Context condition subjects were also presented with 75 cards. A different context sentence appeared on each card, totaling 5 different contexts for each of the 15 to-be-learned words. Each context sentence was missing the vocabulary word, which appeared on the reverse side of the card (e.g., "A process so _____ that men would spit on it." "heinous"). Subjects were instructed to read the context sentences and to try to determine, from that context, what word went in the blank space. They were then to turn the card over to reveal the target word.

After completing the pack of 75 cards, subjects completed a context test and a recall test. The context test was given first and offered 15-new contexts similar to those given to the context group (the vocabulary words were deleted), but which they had not previously seen. Subjects were to supply the missing word. This test was followed by a recall test, in which subjects were presented with the 15 definitions. Subjects were to provide the word that corresponded with the given definition.

The data were subjected to a 2 x 2 analysis of variance with study condition (definition, context) as a between-subjects factor and test type (context, definition) as a within-subject factor. The analysis yielded a significant interaction between study condition and test scores. Subjects in the context condition performed significantly better than subjects in the definition condition on the context test. The two study condition groups performed equally on the definition test.

It was not surprising that the context condition subjects performed significantly better on the context test, as they had prior experience with contexts. The important finding was that the two groups had performed equally on the definition test. The definition group had been exposed the definitions during training and the context condition group had not previously seen the definitions.

Crist and Petrone (1977) interpreted these findings to indicate that definitions of unfamiliar words could be just as effectively learned through context exposure as exposure to actual definitions. In addition, studying contexts rather than definitions alone appeared to result in greater conceptual understanding of word meanings as evidenced by subjects in the context condition recognizing sentences in which the newly learned vocabulary words could be used.

In 1987, Dempster reported a series of five experiments designed to investigate, among other things, the effectiveness of learning unfamiliar vocabulary from context. All five experiments used 38 unfamiliar vocabulary words as the to-be-learned materials. These words were selected because they were judged unlikely to be known by the sample. The list consisted of 26 nouns, 9 adjectives, 1 adverb, 1 verb, and 1 preposition.

In Experiment 1, Dempster (1987) compared free recall of the 38 to-be-learned vocabulary words between three groups who were presented with each word one time at 20 seconds per presentation. The control group was presented with the vocabulary words and their one to four word definitions. Group two was presented with the same words and definitions, but was also presented with a context sentence for each word. Group three was presented with the same words and definitions along with three context sentences. All groups were given ten minutes to complete a free recall test after a distracter task. Dempster (1987) found no significant differences in favor of the context groups. Indeed, subjects in the control group performed slightly better than those in the two context groups.

Failing to find support for the hypothesis that multiple-sentence contexts lead to better vocabulary learning than single-sentence context or no context, Dempster (1987) tested whether multiple presentations would make the advantage of context sentences more apparent. In Experiment 2, he offered three spaced presentations of each of the 38-vocabulary words to subjects assigned to the same conditions as in Experiment 1. Subjects viewed each word with condition appropriate materials for 7 seconds.

The no-context control booklet presented subjects only the words and their definitions. The single-context control condition booklet provided subjects with the words, their definitions, and the same sentence for each repetition. The three-sentence context condition booklet offered each word, its definition, and a different sentence for each repetition of each word. Subjects completed the same free recall test as in Experiment 1. No significant differences between group means were observed.

Experiment 3 was conducted in order to examine, "...the possibility that spaced presentations per se have a facilitating effect on vocabulary learning" (Dempster, 1987, p. 165). Dempster (1987) included two no-context control conditions: one with massed presentation and one with spaced presentation and two three-sentence context conditions: one with massed presentation and one with spaced presentation. For the massed presentation, the booklet consisted of 114 pages with each vocabulary word presented three times in a row on its own page. The spaced presentation booklets were the same as in Experiment 3. The subjects were administered the same free-recall tests used in previous experiments.

A nonsignificant performance advantage was observed for subjects in both of the control groups. Subjects working under spaced presentation conditions significantly outperformed subjects in the massed presentation groups. These findings provided support for spaced presentations, but not variable encoding.

In Experiment 4, Dempster (1987) retained the experimental conditions from Experiment 3, but altered the booklets such that the retention interval for some of the 38-vocabulary words would be consistent between the massed and spaced presentation conditions. Again, no significant effect for variable encoding was observed. The results did, however, provide additional support for the hypothesis that spaced presentations lead to better vocabulary learning than do massed presentations.

In his fifth experiment, Dempster (1987) included three experimental conditions: 1) definition-only control; 2) definition plus context sentences; and 3) a no-presentation baseline group. For subjects in the first two treatment conditions, each vocabulary word was presented 3 times with 37 pages between each presentation (spaced presentation).

Presentation time was 7 seconds per page for both of the above conditions. Subjects were instructed to study the information provided to them in order to attempt to learn the vocabulary words and definitions. The no-presentation baseline group was not exposed to the vocabulary words prior to testing.

Two tests were administered after input and a distracter task: a sentence cued recall test for 19 of the words and a cued-recall test for all of the 38 words. For the sentence-cued-recall test, each subject was given a 2-page test booklet with 19 randomly ordered sentences, each with a blank in the place of the target word. Subjects were instructed to fill in the blanks with the appropriate words that they had learned at input. For the recall test, each subject was given a 2-page test booklet that included the 38 vocabulary words followed by blanks for definitions. Subjects were instructed to write the corresponding definitions in the spaces provided or, if they were unable to recall the definition, they could write a meaningful sentence using the target word. The sentence-cued-recall test was given first and subjects were given 10 minutes to complete each of the tests.

As anticipated, the no-presentation group (baseline) performed very poorly with cued- and free-recall means of .25 and 2.67 respectively. This data was excluded from subsequent analyses. A nonsignificant advantage for the no-sentence context group was observed on both recall measures.

In Experiment 5, Dempster (1987) found that recall was not significantly affected by manipulations designed to affect the number of retrieval routes to the to-be-learned information. Three context sentences did not lead to better recall than the definition only.

In summary, under single, massed, and spaced presentation conditions Dempster (1987) found that offering subjects single or multiple context sentences in conjunction

with words and definitions did not facilitate vocabulary learning over offering only words and definitions. These findings do not support variable encoding hypothesis, that is, offering multiple retrieval routes results in better retention and recall of input material.

Despite unsuccessful attempts to find a learning from context effect in his 1987 experiments, Dempster (1989), in a series of three unpublished experiments, attempted to invoke the Generation Effect to assist subjects in the acquisition of unfamiliar vocabulary. Each of the three experiments required that subjects generate their own sentences while attempting to learn new vocabulary. Each of the experiments is described in turn in the following section.

Dempster (1987) used 38 unfamiliar vocabulary words as the to-be-learned materials. These words were selected because they were judged unlikely to be known by the sample. The list consisted of 26 nouns, 9 adjectives, 1 adverb, 1 verb, and 1 preposition.

In Experiment 1, the 38 to-be-learned vocabulary words (from Dempster, 1987) were presented in a 114-page booklet. Each vocabulary word was presented 3 times with 37 pages between each presentation (spaced presentation). The booklets included one vocabulary word and definition per page. Presentation time was 11 seconds per page. Subjects were assigned to one of two conditions: definition-only and context (generation). Subjects in the definition-only condition were instructed to copy the word and its definition repeatedly. Subjects in the context condition were asked to use a portion of their presentation time to covertly generate one or more sentences for each presented word. Two tests were administered after input and a distracter task: a sentence cued recall test for 19 of the words and a recall test for all 38 words. The two tests were identical to those used in Experiment 5 in Dempster's 1987 paper.

There were no significant differences between groups on either dependent measure, although subjects in the context group slightly outperformed those in the control group. Through exit interviews, Dempster (1989) learned that several experimental-group subjects felt that they did not have sufficient time to generate sentences. The subject-reported insufficient time allocation and the trend favoring the context condition led to the changes for a second experiment.

In Experiment 2, presentation intervals were increased to 15 seconds and the number of repetitions was reduced to two. All other variables remained the same. There was a nonsignificant advantage for subjects in the control condition. Once again, participants in the context condition complained about the pacing rate.

For Experiment 3, Dempster (1989) increased presentation time to 30 seconds for both groups and decreased the number of presentations to one. This was done in order to keep total study times equal for Experiments 1, 2, and 3. Subjects in both groups were given a 38-page booklet of blank paper to turn in synchrony with a 38-page vocabulary booklet. Control-condition subjects were instructed to use the entire 30 seconds to read and copy the word and definition several times. The context-condition subjects were asked to copy the word and definition and to write one to three self-generated sentences using the words. After a distracter task, subjects were asked to complete the same context-cued recall and free-recall tests with allocations of 10 minutes per test. A nonsignificant advantage for the control condition subjects appeared on both cued- and free- recall test scores.

Based on these three experiments, Dempster (1989) concluded that recall was not significantly affected by manipulations designed to provide contextual support in addition

to definitions. Neither the covert nor written generation of context sentences by subjects lead to better recall of the to-be-learned words. These results led Dempster to question the practical usefulness of the learning-from-context approach to vocabulary acquisition. Specifically, he interpreted the results of Experiment 3 to mean that requiring students to generate and write sentences containing new vocabulary words may actually interfere with the learning of said words.

Despite the apparent failure to obtain a generation advantage for subjects in Dempster's (1989) experiments, other researchers have found learning from context effects (e.g., Crist & Petrone, 1977; Gipe, 1979; Nagy, Herman & Anderson, 1985; Jenkins, Stein, & Wysocki, 1984). Careful inspection of Dempster's (1989) procedures revealed two potential flaws: insufficient time allocation for experimental subjects and the possibility that spaced presentations may have overwhelmed Dempster's (1989) ability to find a generation advantage. It was partially due to these perceived shortcomings that the present study was designed.

The Current Study

The purpose of the current study was to investigate multiple facets of the Generation Effect and how its principles might effectively be incorporated in the acquisition of vocabulary. The accepted definition of the Generation Effect is that there is better retention for material that is self-generated as compared to material that is simply read (Slamecka & Graf, 1978; Graf, 1980). In most studies of the Generation Effect, researchers have interpreted this definition to mean that subjects should create examples or elaborations that are related to the to-be-learned material. For example, perhaps a subject would be asked to learn the definition of the word "sprat." The subject would be

given the word and the definition and would be instructed to "generate" a meaningful sentence using the word properly. An alternative interpretation of the definition of the Generation Effect might require that subjects generate the meaning of to-be-learned material instead. For example, rather than giving the subject a term and definition, the subject could be given the term and a sentence in which the term is used properly. The subject's "generation" task, then, would be to generate the definition of the term in question. Although this latter method is not precluded by the generally accepted definition of the Generation Effect, no studies in which this approach has been used appear in the literature. Therefore, this study was designed to compare two separate methods of inducing the Generation Effect.

Two alternative methods of inducing the Generation Effect were investigated in a 3 x 3 design with encoding task (read, generate sentence, generate definition) as a between-subjects variable and time of test (immediate, 48-hour delay, 21-day delay) as a within-subject variable.

Subjects in the definition-only control group attempted to learn 22 unfamiliar vocabulary words by repeatedly writing each word and its definition during the presentation time. Subjects in the sentence generation group attempted to learn the 22 unfamiliar vocabulary words under the traditional method of effectuating the Generation Effect: they wrote each word and its definition and then wrote a meaningful sentence using that word. Subjects in the definition generation group attempted to learn the 22 unfamiliar vocabulary words under the previously mentioned alternative method of producing the Generation Effect: they read each target word embedded in a rich context sentence and extrapolated and wrote a definition for the target word. Subjects in all three

groups were tested immediately following a distracter task, after a 48-hour delay, and again 21-days later.

The impetus for the current study was the puzzling findings of Dempster (1987, 1989). Recall that he was unable to effectuate better recall of definitions from subjects who were offered multiple context sentences to support definitions (Dempster, 1987) or from subjects who were offered the opportunity to covertly or overtly generate their own context sentences in support of experimenter provided definitions (Dempster, 1989). Dempster (1987, 1989) found no differences in performance between the subjects in experimental groups and subjects in the read-only groups. Admittedly, these findings were inconsistent with previous research related to the Generation Effect and learning vocabulary from context.

Even before Slamecka and Graf's (1978) delineation of the phenomenon, numerous studies have shown the emergence of the Generation Effect with different types of input tasks and materials, including paired associates (Bobrow & Bower, 1969; Hirshman & Bjork, 1988; McElroy and Slamecka, 1982; Slamecka & Graf, 1978), unrelated word-word and word-nonword pairs (McNamara & Healy, 1995), and sentence completion tasks (Anderson, Goldberg, & Hidde, 1971; & Ghatala, 1981). The Generation Effect has been obtained with encoding tasks as (read or generate) within-subject (Graf, 1980; Hirshman & Bjork, 1988; McElroy and Slamecka, 1982; & Slamecka & Graf, 1978) and between-subject variables (Anderson, Goldberg, & Hidde, 1971; Ghatala, 1981; Hirshman & Bjork, 1988; McNamara & Healy, 1995; & Slamecka & Graf, 1978). The difference between these studies and Dempster's (1987, 1989) studies, aside from the difference in findings, was the practicality of the tasks, materials, and implications.

Dempster (1989) tested the practical usefulness of generation by applying it to a real-life learning situation. In his 1989 study, he sought to assist learners by affording them the opportunity to generate their own context sentences in attempts to facilitate the learning of unfamiliar vocabulary words. Interpretation of the results from Dempster's (1989) study would indicate that a student's creation of context sentences including previously unfamiliar vocabulary words would not facilitate the learning of those words. Through analysis of that study and pilot studies conducted by the present researcher in 2002, it was discovered that failure to facilitate learning (absence of the Generation Effect) was likely due to time constraints imposed on the subjects. Subjects in Dempster's (1989) experimental conditions were simply not offered sufficient time to complete and conceptualize their assigned tasks (i.e., read and copy definition, then create a meaningful sentence, or sentences, using that word). Conversely, subjects in read-only conditions had ample time to complete and conceptualize their task—reading and copying the unfamiliar word and definition several times in an attempt to learn the material.

An additional shortcoming in Dempster's (1987, 1989) studies may have been the use of spaced presentations of to-be-learned vocabulary. Specifically, the effect spaced presentation had on subject learning might have overwhelmed Dempster's (1989) ability to find a generation effect. The spacing effect is one of the most enduring psychological phenomena (e.g., Bloom & Shuell, 1981; Landauer & Bjork, 1978; Rea & Modigliani, 1985; Reder & Anderson, 1982; Reynolds & Glaser, 1964; Smith & Rothkopf, 1984). Therefore, the spacing effect might have confounded Dempster's (1987, 1989) encoding variability component. That is, subjects in read-only conditions may have benefited

tremendously from spaced presentations because their only task was to focus on the words and their definitions while subjects in sentence generation groups had insufficient time to thoroughly process the word and its definitions because they were compelled to write a meaningful sentence using the to-be-learned vocabulary word. Regardless, Dempster's (1989) attempt to apply the principles of the Generation Effect to educationally relevant settings was not paralleled by other reviewed research literature.

In much of the research related to the Generation Effect, the investigators were primarily interested in theoretical interpretations of this phenomenon. That is, why does the act of generation facilitate learning? To that end, they used sentence completion tasks (Anderson, Goldberg, & Hidde, 1971; & Ghatala, 1981), sentence construction tasks (Graf, 1980), and the learning of paired associates (Bobrow & Bower, 1969; Hirshman & Bjork, 1988; McElroy and Slamecka, 1982; Slamecka & Graf, 1978). Paired associate learning, and its relevance in the learning of vocabulary, is not without value. Consider, for example, the learning of foreign languages (e.g., target word coupled with synonym). In much generation effect research, the paired associates (word-word, word-nonword, nonword-nonword) used were restricted by researcher-imposed rules (i.e., first associate, third associate, [Hirshman & Bjork, 1988]; associate, category, opposite, synonym, rhyme, [Slamecka & Graf, 1978]; letter transposition, rhyme, [McElroy & Slamecka, 1982]) leading to little in the way of practical utility. Sentence completion tasks and sentence construction from experimenter provided frames, likewise, offer little, if any, educational value.

Little attention has been afforded to testing the Generation Effect with tasks requiring the learning or enhancement of knowledge. Of the literature reviewed, the only study

(other than Dempster, 1989) that focused on acquisition of new knowledge was McNamara and Healy's (1995) second experiment. In that experiment, subjects were to learn word-nonword pairs under the guise that they were learning the foreign equivalents of English words. Although this offered a sounder basis for the exploration of potential educational implications of the Generation Effect, the subjects were not actually acquiring "real world" knowledge, although they thought they were. The difference between that investigation and the current study is that in the current study, subjects were, in fact, attempting to acquire "real world" knowledge. That is, they were attempting to acquire new vocabulary.

In addition, in the current study, an alternative method of inducing the Generation Effect was investigated. This alternative method may be a more authentic generation task. Recall that the commonly accepted definition of the Generation Effect is that there is better retention for material that is self-generated as compared to material that is simply read (Slamecka & Graf, 1978). Here, in the definition generation condition subjects were presented with unfamiliar vocabulary words embedded in rich context sentences. The generation task for the subjects was to extrapolate, or generate, the meaning of the target word from that context sentence. This was interpreted to be of more practical usefulness in real life settings in that college students routinely encounter unfamiliar vocabulary words in written and spoken prose. It is not common for them to be awarded the opportunity to stop and use a dictionary in order to determine the meanings of those unfamiliar words. Rather, they might attempt to extrapolate the meaning from surrounding context clues. Without direct tests of retention, however, it is impossible to know whether they retain the meaning of those words. Previous research has

demonstrated that undergraduate college students (Crist & Petrone, 1977) and elementary school children (Gipe, 1979; Jenkins, Stein, & Wysocki, 1984; Nagy, Herman, & Anderson, 1985) effectively obtain and retain meanings of words from context in written prose. This has been effectuated when subjects have received explicit instructions to do so (Crist & Petrone, 1977; Gipe, 1979) and under more natural reading conditions (Jenkins, Stein, & Wysocki, 1984; Nagy, Herman, & Anderson, 1985). It has been demonstrated with almost unnaturally informative contexts (Gipe, 1979; Jenkins, Stein, & Wysocki, 1984) and with contexts specifically designed to be more natural (Nagy, Herman, & Anderson, 1985). Inclusion of the definition generation condition (learning from context) provided the opportunity to further investigate whether this method of inducing the Generation Effect had practical usefulness in the learning of vocabulary.

In the current study, the better-established method of evoking the Generation Effect was also investigated in the acquisition of new vocabulary knowledge. Using the more traditional method of effectuating a generation effect, sentence generation subjects were presented with unfamiliar vocabulary words and definitions. Their task was to generate a meaningful sentence including the unfamiliar vocabulary word. Subjects were instructed that a "meaningful" sentence should offer a reader who is unfamiliar with the word the ability to extrapolate the meaning. Subjects in a definition-only control group were presented with the words and definitions and were instructed to repeatedly copy them during the presentation time. Subjects were allocated sufficient time to complete assigned input tasks (55 seconds per word) and only one presentation of input material was offered. This was done to determine whether the perceived flaws in Dempster's (1987, 1989) studies could be corrected. The performance of the three groups at

immediate cued recall, 48-hour delayed cued recall, and 21-day delayed cued recall was then compared.

In summary, the purpose of the current study was threefold. First, it was designed to investigate the practical educational application of the Generation Effect. Second, two alternative methods that might effectively produce the Generation Effect in the learning of unfamiliar vocabulary were investigated: 1) given a rich context sentence including the target word, generating a definition of the target word from that sentence (definition generation condition) and 2) given the unfamiliar vocabulary word and definition, generating a meaningful sentence including that word (sentence generation condition). Third, the design of the study was intended to determine whether the Generation Effect could make a contribution to learning from context approaches to vocabulary acquisition. That is, although several studies support the practice of learning vocabulary from context it does not appear that researchers have addressed the notion that with this type of vocabulary learning learners are actually generating word meanings based on the information presented in the context. By examining the learning from context approach to vocabulary acquisition from a Generation Effect perspective it would follow that the learners' generation of word meanings should result in strong retention.

Specific Hypotheses

It was expected that there would be an overall main effect for treatment group. Specifically, consistent with previous research, it was expected that both generation groups (sentence generation, definition generation) would outperform the control group on tests of immediate, 48-hour delayed, and 21-day delayed cued recall (Anderson, Goldberg, & Hidde, 1971; Bobrow & Bower, 1969; Ghatala, 1981; Graf, 1980; Hirshman

& Bjork, 1988; McElroy & Slamecka, 1982; McNamara & Healy, 1995). Also, based on previous research, it was expected that the definition generation group and the sentence generation groups would perform similarly on recall tests (Ghatala, 1981; Graf, 1980). Based on pilot studies conducted in 2002, it was anticipated that the sentence generation group and the control group would show similar rates of decay across time. Lastly, it was hypothesized that the definition generation group and sentence generation groups' performance on tests of delayed recall (48-hour and 21-day delay) would differ. Glover, Bruning, and Plake (1982) found that subjects who completed tasks requiring "more difficult" processing recalled at a higher rate than subjects completing "less difficult" processing. If there were a difference in the level of difficulty between the tasks required of the sentence generation group and the definition generation group, then their performance on tests of delayed recall might differ.

To summarize, the hypotheses put forth in this study are as follows:

1. Both generation groups (group one: sentence generation; group two: definition generation) would outperform the control group (definition-only) on tests of immediate and delayed recall.
2. The definition generation group and the sentence generation group would perform similarly on recall tests.
3. The sentence generation group would continue to outperform the control group on both tests of delayed recall (as stated in hypothesis one), but the groups' performance would decay at similar rates.
4. The definition generation group and sentence generation groups' performance on tests of delayed recall (48-hour and 21-days) would differ.

CHAPTER 2

METHOD

Participants

Participants were 61 University of Nevada, Las Vegas, undergraduate students accessed from the Educational Psychology Department subject pool during the Spring 2003 and Summer 2003 semesters by way of voluntary sign-up for research credit. This research credit served as partial fulfillment of course requirements.

Materials

This study was designed to investigate multiple facets of the Generation Effect. It was designed specifically to investigate how subject generation of sentences or definitions could facilitate vocabulary acquisition at both immediate and delayed tests of recall and how these different types of generation tasks might produce differential effects in recall. Subjects attempted to learn unfamiliar vocabulary words and definitions under one of three conditions: 1) read words and their definitions and copied them several times; 2) read and copied the words and their definitions one time and wrote one meaningful sentence per word; or 3) read meaningful sentences that included the vocabulary word and attempted to define the word. A description of all tasks used in the experiment follows.

Verbal Intelligence Assessment

The verbal ability instrument was adapted from the Kit of Factor Referenced Tests (Ekstrom, French, Harman, & Dermen, 1976). The test consisted of 36 multiple-choice items requiring subjects to select a synonym (from four alternatives) for each of the 36 vocabulary words. Subjects were instructed to select the word that had the same, or nearly the same, meaning as the vocabulary word (see appendix B). This version of the verbal test is considered to be quite difficult with students rarely achieving a score of 20 or greater (LayIng, 1994).

Vocabulary Words

Twenty-two uncommon English nouns were chosen as the to-be-learned vocabulary. These words were selected from the thirty-eight words used by Dempster (1987). Dempster (1987) selected a word primarily if it was judged unlikely to be known by the subjects. Dempster's list consisted of 26 nouns, 9 adjectives, 1 adverb, 1 verb, and 1 preposition. Based on the low scores of the subject's in Dempster's 1987 and unpublished follow-up studies and time constraints, the size of the vocabulary list was reduced. Only the nouns were used because using terms from one part of speech was considered to likely be less confusing to the learners. In addition, there were a sufficient number of nouns from Dempster's (1987) list to allow for this. The original list consisted of 26 nouns, however, four pairs of nouns were very similar to each other (e.g., "tarn-mountain lake: corniche-mountain road"). As a result, one noun from each pair was omitted in order to mitigate the potential for confusion by the subjects.

Vocabulary Acquisition Booklets

Subjects in all conditions were provided with a 23-page, 8.5 x 5.5 inch lined paper booklet. Booklets were used by the subjects to complete the vocabulary acquisition task.

Puzzles

Hexagon match word puzzles were used as distracter tasks between acquisition and the immediate recall test (see appendix C). Subjects were to place seven presented words into hexagons so that each letter would match the letter in the adjacent hexagon. All words were to read in a clockwise direction.

Perceived Difficulty Assessment

Sentence generation and definition generation subjects completed a perceived task difficulty assessment. Immediately following the completion of the input task for the 22 words, subjects used a 5-point Likert scale to indicate their perceived difficulty of the acquisition task. Subjects were instructed to assign score of "1" if they found the acquisition task to be "very difficult;" a "2" indicated that they found the task "difficult;" a "3" indicated "neutral;" a "4" indicated "easy;" and a "5" indicated "very easy."

Recall test

Three tests were used to measure the subjects' retention of word meanings (see appendix D). In these tests, nouns were offered as cues for recall of the definitions. The first test, an immediate recall measure, was given after the vocabulary acquisition task and a three-minute distracter task (puzzles). The second test, delayed recall I, was given exactly forty-eight hours after the acquisition phase. The third test, delayed recall II, was given twenty-one days after the acquisition phase. The 22 nouns offered as cues for definition recall were assembled in a different order on each of the three recall tests.

Procedure

Data was collected by the author. Approval for research involving human subjects was obtained from the University of Nevada, Las Vegas Institutional Review Board. Subjects participated in one one-hour session and two half-hour sessions.

The sixty-one subjects were randomly assigned to one of the three treatment conditions: twenty-one subjects were assigned to condition one, definition-only control; nineteen subjects were assigned to condition two, sentence generation; and twenty-one subjects were assigned to condition three, definition generation. All tasks were administered in a classroom setting with no more than twenty-one subjects participating at one time. Subjects in all conditions were instructed: 1) that their objective was to learn the definition of the vocabulary words using the experimenter provided method; 2) not to use their own methods to learn the words; 3) not to turn pages in their booklets forward or backward during the study except at the appropriate times; and 4) to move along with the rest of the group even if they had not completed the assigned task for the current word. During the acquisition task the researcher stood at the front of the room and during the recall tasks the researcher stood at the front of the room and roamed the room.

Subjects viewed the unfamiliar nouns with the group appropriate auxiliary materials via an overhead projector. Subjects viewed each unfamiliar noun and auxiliary materials for 55-seconds. The conditions were as follows:

Condition One:

Definition-only control: During the 55-second time interval subjects in the definition-only control viewed the 22 unfamiliar nouns with one to two word definitions via an overhead projector. They were instructed to

repeatedly write the unfamiliar noun and its definition. A timer buzzed every 55-seconds and the subjects were instructed to turn to the next page in their booklets and to repeat the task for the next word. At the same time, the researcher changed the overhead to the next vocabulary word and definition.

Condition Two:

Sentence generation: During the 55-second time interval subjects in the sentence generation group viewed the 22 unfamiliar nouns and their definitions via an overhead projector. Subjects were instructed to write the unfamiliar noun and its definition once and to then create a "meaningful" sentence which exemplified the meaning of the unfamiliar noun. A timer buzzed every 55-seconds and the subjects were instructed to turn to the next page in their booklets and to repeat the task for the next word. At this time, the researcher changed the overhead to the next vocabulary word and definition. Upon completion of the acquisition task, subjects rated their perceived difficulty of the acquisition task using a rubric presented on the overhead projector. Subjects wrote the appropriate number in the upper left-hand corner of their test booklets.

Condition Three:

Definition generation: During the 55-second time interval subjects in the definition generation group viewed the 22 unfamiliar nouns and a corresponding sentence exemplifying the meaning of the word via an overhead projector. Subjects were instructed to read the unfamiliar noun

and the accompanying sentence and to then generate a brief definition for the unfamiliar noun. A timer buzzed every 55-seconds and the subjects were instructed to turn to the next page in their booklets and repeat the task for the next word. At this time, the researcher changed the overhead to the next sentence. Upon completion of the acquisition task, subjects rated their perceived difficulty of the acquisition task using a rubric presented on the overhead projector. Subjects wrote the appropriate number in the upper left-hand corner of their test booklets.

Following the vocabulary acquisition phase, all subjects worked on word puzzles for 3 minutes (see appendix C). Immediately after working on the word puzzles, subjects completed the immediate noun-cued definition recall test. Subjects returned 48-hours after initial testing to complete their first delayed recall test. Subjects returned 21-days after initial testing to complete their second delayed recall test.

After the vocabulary acquisition task, but before the immediate test, subjects in the sentence generation group and definition generation group were presented with a perceived difficulty assessment task via the overhead projector. Subjects were instructed to judge their acquisition task according to the following scale: 1 indicated very difficult; 2 indicated difficult; 3 indicated neutral; 4 indicated easy; 5 indicated very easy. Subjects were instructed to write their rating on the upper left hand corner of their test booklet.

CHAPTER 3

RESULTS

Subjects participated in an experiment consisting of a vocabulary acquisition task and three noun-cued recall tests measuring acquisition of the vocabulary. The subjects were tested immediately after the acquisition task and a three-minute distracter task, forty-eight hours after the acquisition task, and 21-days after the acquisition task.

Scoring

Noun-cued definition-recall tests for the definition-only control and the sentence generation groups were scored using the following rubric: subjects received full credit (one point) for a complete definition (e.g., "mountain lake" for the unfamiliar noun "tarn"); partial credit (one half point) for a partial definition (e.g. "lake" for the unfamiliar noun "tarn"); no credit for an incorrect response.

Noun-cued definition-recall tests for subjects in the definition generation group were scored by comparing a subject's response on the test to the definition generated by that subject. For example, if a subject generated the definition "office assistant" for "amanuensis" rather than the more correct definition, "secretary," that subject's free-recall response for "amanuensis" was considered completely correct when they responded "office assistant," partially correct if they responded "assistant," or incorrect for some unrelated response. Noun-cued definition-recall tests for the definition generation group were scored according to the following rubric: subjects received full

credit (one point) for a complete definition (e.g., "amanuensis" means "secretary" which would be a correct response, however, "office assistant" would be equally correct); partial credit (one half point) for a partial definition; no credit for an incorrect response. Therefore, the maximum score on the recall test regardless of experimental group was 22; minimum score was 0.

The meaningfulness of the sentences created by the sentence generation group was perceived to be pertinent to the authenticity and value of the generation task. In previous studies, findings indicated that the Generation Effect did not occur with meaningless items such as nonwords (e.g. McElroy & Slamecka, 1982) or with anomalous sentences (Graf, 1980). Therefore, a rubric was established to rate the sentences created by the sentence generation group subjects.

Sentences were rated on a scale of 0 to 2. A score of 0 was awarded if: 1) there was no sentence written; 2) the subjects did not attempt to write an appropriate sentence, but instead tried to be humorous (e.g. "I have been buskin my butt to write these words over and over." "Tarn! I can't swim through that!"); or 3) the subject, sometimes knowingly and sometimes not, did not understand the definition, and thus wrote a completely inappropriate sentence or one that was entirely ambiguous (e.g. "I had an éclat." "The enigma of 'Hey Diddle Diddle, the Cat and the Fiddle' is my favorite one." "My boyfriend was sick and used it as a proselyte so I would come over instead of going out with my friends.") A score of 1 was awarded if: 1) the subject used the word appropriately and in a way that was possibly meaningful to him/her, but a reader who was unfamiliar with the word would not likely be able to extrapolate the meaning from the sentence (e.g. "My boss just hired a new amanuensis." "In my family, I am the

farceur."), or 2) the subject used the word in a fairly meaningful sentence, but appeared to be confused as to the exact meaning of the word (seemingly because some of them did not know exactly what the one or two word definition meant) (e.g. "Earthquakes usually don't take place on foibles, only bigger faults." "My brother was given the cognomen John, after his father."). A sentence was awarded a score of 2 if the subject used the word in a meaningful sentence that clearly illustrated the meaning of the word (e.g. "He was a Catholic who became a Mormon, so we call him a proselyte." "Robin Williams is a farceur.") thus, a reader unfamiliar with the word could likely extrapolate the meaning from the sentence.

Twenty-two vocabulary words were presented to all subjects. The sentence generation group was to generate a meaningful sentence for each word. The maximum possible total score for a set of sentences was 44 and the minimum possible total score was 0. The data for subjects that scored 22 (half of the possible total) or less on the total sentence ratings were considered suspect. These subjects did not follow directions and did not properly complete the assigned task. On these grounds, the data for 4 subjects from the sentence generation group was eliminated from further analyses. All subjects in the definition-only control and sentence generation group appeared to follow the directions. Therefore, all subject data from those two groups was retained for analysis.

Due to the critical nature of the definition extraction task, it was important to determine the level of accuracy of subject-generated definitions. Thus, the definitions offered by subjects in the definition generation condition were analyzed and scored as to their accuracy. Completely accurate definitions and definitions that were near the meaning of the target word were given a score of 1 and definitions that were completely

inaccurate were given a 0. Post hoc evaluation of accurate meaning extraction in this condition revealed a mean score of 19.24. This means that subject were able to accurately extrapolate the intended meanings of the unfamiliar vocabulary words on the average of 87% of the time.

Vocabulary Acquisition Analyses

Means and standard deviations of vocabulary test scores (dependent measure) are reported in Table 2. Means and standard deviations for the Verbal Intelligence Test are pictured in Table 3. The vocabulary test scores were submitted to a 3 x 3 mixed model analysis of co-variance with time of test (immediate, 48-hour delay, 21-day delay) as the within-subject factor, treatment condition (definition-only control, sentence generation, definition generation) as the between-subjects factor, and score on the Verbal Intelligence test as a covariate.

Table 2: Means and Standard Deviations for All Groups on Dependent Measure

	Immediate		48-hour		21-day		
	Mean S. D.	Predicted Mean	Mean S. D.	Predicted Mean	Mean S. D.	Predicted Mean	n
Definition-only control	8.67 4.88	8.74	7.29 5.30	7.37	5.71 5.72	5.80	21
Sentence generation	12.90 4.08	12.79	11.30 3.95	11.17	8.97 3.60	8.84	15
Definition generation	10.60 6.23	10.60	9.12 5.75	9.13	7.23 5.24	7.25	21

Figure 1: Estimated Marginal Means for Dependent Measures

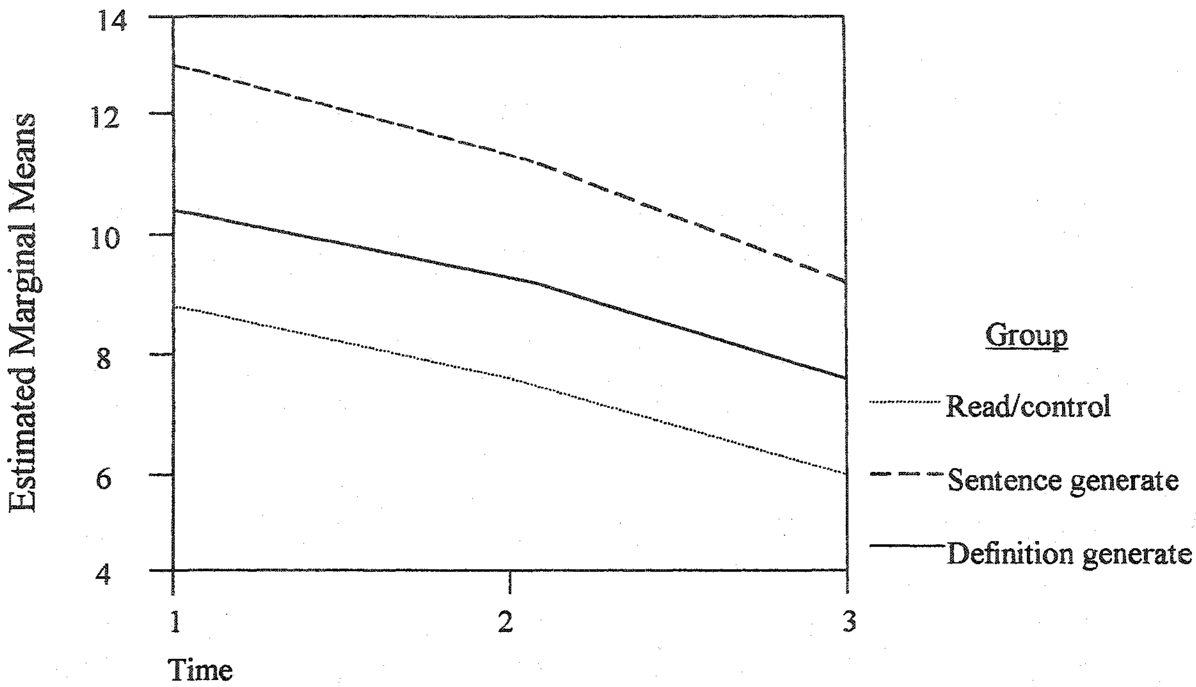


Table 3: Means and Standard Deviations for All Groups on Verbal Intelligence Measure

	Means	SD
Definition-only control	19.14	5.59
Sentence generation	19.53	4.78
Definition generation	19.29	5.02
Overall	19.30	5.09

The correlations between the covariate, verbal intelligence (as measured by the Verbal Intelligence test) and scores on the immediate, 48-hour delay, and 21-day delay tests were $r = .46$, $p < .01$; $r = .54, p < .01$; and $r = .56, p < .01$ respectively. The

covariate, score on the Verbal Intelligence test, was significant $F_{(1, 53)}=22.945$, $p<.001$. The interaction between time of test and treatment condition was not significant, $F_{(3.131, 82.963)} = .477, p > .05$. There was a significant main effect for time of test, $F_{(3.131, 82.963)} = 8.203, p < .001$. There was a significant main effect for treatment condition, $F_{(2, 53)} = 3.255, p < .05$.

The significant main effect for treatment condition was further explored via the Tukey HSD post-hoc comparison procedure. Subjects in the sentence generation group outperformed subjects in both of the other treatment conditions. Performance of subjects in the definition only control group and definition generation group did not differ.

The significant main effect for time of test was also further explored via the Tukey HSD post-hoc comparison procedure. Subjects performed significantly better on the immediate test than on the 48-hour delay and 21-day delay tests. Subjects performed significantly better on the 48-hour delay test than on the 21-day delay test.

Analysis of Specific Hypotheses

Hypothesis 1 was that both generation groups (group one: sentence generation; group two: definition generation) would outperform the control group (definition-only) on tests of immediate and delayed recall. This hypothesis was provided with partial support. Then main effect for treatment condition indicates that the sentence generation group outperformed both other treatment conditions. Although the definition generation group did not outperform subjects in the definition only condition, their overall performance was better.

Hypothesis 2 was that the definition generation group and the sentence generation group would perform similarly on recall tests. This hypothesis was partially supported.

As there was no time-by-treatment interaction, the performance scores for these two groups is considered equivalent.

Hypothesis 3 was that the sentence generation group would continue to outperform the control group on both tests of delayed recall (as stated in hypothesis one), but the groups' performance would decay at similar rates. This hypothesis was not supported. As there was no time-by-treatment interaction, the performance scores for these two groups across time is considered equal.

Hypothesis 4 was that the definition generation group and sentence generation groups' performance on tests of delayed recall (48-hour and 21-days) would differ. This hypothesis was not supported. As there was no time-by-treatment interaction, the performance scores for these two groups across time is considered equal.

Perceived Difficulty Analysis

Perceived difficulty ratings from subjects in the sentence generation group and the definition generation group were analyzed via an independent *t*-test. There was no significant difference in perceived difficulty ratings between sentence generation (M = 3.0, SD = 1.13) and definition generation [M = 3.33, SD = .91], $t_{(34)} = -.976$, $p = .34$] treatment conditions. The magnitude of the difference in the means was small (eta squared = .03).

CHAPTER 4

DISCUSSION

Previous research has clearly established the existence of the Generation Effect (Slamecka & Graf, 1978) and the ability to induce it with: 1) encoding tasks (read, generate) as between-subjects (e.g., Anderson, Goldberg, & Hidde, 1971; Ghatala, 1981; Hirshman & Bjork, 1988; McNamara & Healy, 1995; & Slamecka & Graf, 1978) and within-subject (e.g., Graf, 1980; Hirshman & Bjork, 1988; McElroy and Slamecka, 1982; & Slamecka & Graf, 1978;) variables; 2) paired associate materials (e.g., Bobrow & Bower, 1969; Hirshman & Bjork, 1988; McElroy and Slamecka, 1982; Slamecka & Graf, 1978); 3) unrelated word-word and word-nonword pairs (e.g., McNamara & Healy, 1995); 4) sentence completion tasks (e.g., Anderson, Goldberg, & Hidde, 1971; & Ghatala, 1981); and 5) sentence construction tasks (e.g., Graf, 1980). The current study differs from the aforementioned research in that it was, in part, designed to investigate practical educational applications of the Generation Effect. It was designed specifically to investigate how subject generation of sentences or definitions might facilitate vocabulary acquisition at both immediate and delayed tests of recall and how these different types of generation tasks might produce differential effects in recall.

The discussion is organized as follows. Results of the analysis of the specific research hypotheses are discussed in turn. Partial explanations of the results are provided with the corresponding hypothesis. A general discussion of the results follows and

includes a detailed discussion related to how the current findings contribute to the literature on the Generation Effect and learning from context. Next, the limitations of the current study are addressed. In the final section, suggestions are presented for future research on the Generation Effect and how its principles might be incorporated in vocabulary learning.

Hypotheses

Hypothesis 1 was that both generation groups (sentence generation, definition generation) would outperform the control group (definition only) on tests of immediate and delayed recall. Subjects in the sentence generation group significantly outperformed subjects in both the definition generation and the definition only control groups when recall measures were collapsed across time (main effect for treatment condition). Subjects in the definition generation group performed better than subjects in the definition only control group, but this difference failed to reach significance.

These findings indicate that offering subjects vocabulary words and definitions and the opportunity to generate meaningful sentences facilitated learning over: 1) being offered only the target words and definitions and 2) being offered rich context sentences and generating target word meanings from them. This is likely because the act of generating personally meaningful sentences is a type of elaborative rehearsal. This type of rehearsal involves deep semantic processing of input information which results in strong and stable memories.

Hypothesis 2 was that the definition generation group and the sentence generation group would perform similarly on recall tests. As there was no significant time-by-treatment interaction, this hypothesis is partially supported. The performance of subjects

in the sentence generation and definition generation groups at recall did not differ statistically. In addition, the subjects in the definition-only control group performed as well as subjects in both experimental groups. This means that the Generation Effect failed to appear. In other words, if the Generation Effect had appeared, then the control group would have a predicted mean recall level significantly lower than either, or both, of the treatment conditions.

This failure to effectuate a generation effect at immediate recall may have been due to immediate recall tests being potentially poor indicators of whether learning has actually taken place (Bjork, 2001). In a review of acquisition (immediate recall testing) versus retention (delayed recall testing) of a variety of to-be-learned materials and skills, Bjork (2001) clearly demonstrates that at acquisition subjects in control conditions may appear to have "learned" as much as subjects in experimental conditions. When performance between acquisition and retention of to-be-learned material is considered, however, experimental condition subjects appear to have an advantage at retention. That is, control subject performance at delayed recall testing no longer remains equivalent to performance of subjects in experimental conditions. The subjects in the experimental conditions now (at delayed testing) show a tremendous advantage (Bahrick, 1984). Therefore, control group subjects may have performed similarly to experimental groups at immediate test, but that does not necessarily mean that they actually learned the same amount of information.

Another possible reason for the absence of the Generation Effect on the recall tests could have been differential levels of subject motivation. The subjects that participated in the current study were obtained through the Educational Psychology Department's

subject pool during the 2003 spring and summer sessions. During these sessions, educational psychology students appeared to be reluctant to participate in research. Most of them opted to complete research summaries, as articles to-be-summarized were available online (A. J. Corkill, personal communication, September, 2003). Several other researchers in the Educational Psychology Department also had difficulty with subjects from this pool during the same time span.

In addition, extremely low power was observed for the time-by-group interaction (.145). This lack of power may have been partially responsible for the apparent absence of the Generation Effect. If the sample size per condition had been larger, the effect might have appeared.

Hypothesis 3 was that the sentence generation group would continue to outperform the control group on both tests of delayed recall (as stated in hypothesis 1), but the groups' performance would decay at similar rates. There was no time-by-treatment interaction, indicating that performance scores for these two groups were equal across time. Thus, hypothesis 3 was not supported.

Interpretation of the analysis revealed that the control condition and sentence generation subjects' performed similarly on tests of delayed recall. This may have been due, in part, to the extremely low power for the time-by-treatment interaction (.145). If the number of subjects had been larger, the effect might have appeared. Although the performance on tests of delayed recall between the definition-only control and sentence generation groups was not significantly different, the observed differences were in the predicted direction. That is, subjects in the sentence generation group performed consistently better across time, but without significance. The predicted mean difference

between theses two groups was 3.8 for the 48-hour delay test, and 3.04 for the 21-day delay test. The sentence generation group subjects also performed consistently, though not significantly, better than the definition generation group subjects across time. Between the sentence generation subjects and definition generation subjects, there was a predicted mean difference of 2.04 for the 48-hour delay test and 1.59 for the 21-day delay test (in favor of the sentence generation group's subjects).

Hypothesis 4 was that the performance by subjects in the definition generation group and sentence generation group on tests of delayed recall (48-hour, 21-day) would differ. This hypothesis was not supported. As there was no time-by-treatment interaction, the performance scores for these two groups across time were considered equal.

Failure to find a difference between the performance of the definition generation group subjects and the sentence generation group subjects on tests of delayed recall may have been a result of extremely low power (.145) for the time-by-treatment interaction. If the number of subjects had been greater, a difference may likely have been found. There may have also been a lack of motivation for subjects participating in the study. The subjects were from 2003 spring and summer sessions. During these sessions, other students and faculty found it difficult to find research participants. Many students opted to complete article summaries, instead of participating in research studies. This was likely because the articles to be reviewed were available online (A. J. Corkill, personal communication, September, 2003).

Although there were no significant differences found in level of recall between these two experimental groups, the subjects in the sentence generation group consistently performed better across time. The difference between the predicted means was 2.04 for

the 48-hour delayed recall test and 1.59 for the 21-day delayed recall test (in favor of sentence generation group's subjects).

General Discussion

The following section addresses the general findings of present study with possible explanations for those findings. First, the superior performance of subjects in the sentence generation condition will be discussed in comparison to Dempster's (1989) results. Possible reasons for differential results between the studies will be explained in detail. Second, possible theoretical explanations for differences in performance (some significant, some not) between the groups in the present study will be given attention. Third, the performance of the subjects in the definition generation group will be discussed in comparison to the performance of subjects in the definition only control group. Specifically, the failure to find a significant difference between the two groups will be discussed in relation to the learning from context approach to vocabulary acquisition.

The results indicate that when collapsed across time subjects in the sentence generation condition subjects performed significantly better. This is interpreted to mean that offering subjects the opportunity to generate meaningful sentences which include the target words, facilitates vocabulary learning over offering them the opportunity to copy target words and definitions repeatedly or offering them rich context sentences (including the target words) from which to generate the definitions of the target words. This finding is contrary to Dempster's (1989) study. Recall that in Dempster's (1989) study, in Experiments 1, 2, and 3 subjects in the read-only conditions and subjects in context sentence generation conditions did not differ in performance. Interpretation of the results

from Dempster's (1989) study would indicate that a learner's creation of context sentences including previously unfamiliar vocabulary words would not facilitate the learning of those words. This failure to find differences in performance between control and experimental groups might have been due to two potential flaws in his methodology.

One potential flaw was that he apparently did not offer subjects in context conditions sufficient time to complete and conceptualize their assigned tasks (i.e., read and copy definition, then create meaningful sentence, or sentences, using that word). In contrast, subjects in the read-only conditions were offered ample time to complete their assigned task—reading and copying the unfamiliar word and definition several times in an attempt to learn the material.

Another potential drawback of Dempster's (1987, 1989) methodology may have been the use of spaced presentations of to-be-learned vocabulary. In particular, the effect spaced presentation may have had on subjects' learning might have overwhelmed Dempster's (1989) ability to find the Generation Effect. That is, subjects in the read-only conditions may have benefited tremendously from spaced presentations, as their only task was to focus on the words and definitions. On the other hand, subjects in sentence generation conditions reported having insufficient time to thoroughly process the word and its definition because they were required to write meaningful sentence(s) (one to three) using those words with the same time allocation as the read-only subjects. Both of the aforementioned concerns about Dempster's (1987) methodology were addressed, to a greater or lesser degree, in the current study.

Presentation times in the current study were sufficient for subjects in both the sentence generation and definition generation conditions. The pacing of presentation of

the to-be-learned materials was determined through pilot studies. Subjects in Dempster's (1989) study were eventually afforded 30-second presentations of the learning material. Through pilot studies, the current researcher discovered that 30-seconds was insufficient. The present study offered subjects 55-second presentation rates. Inspection of subjects' input products revealed that, by and large, they had sufficient time to complete their tasks. In addition, subjects were offered only one presentation of each target word and definition.

With the above concerns addressed, it could be concluded that the generation of meaningful sentences facilitates vocabulary learning. At the same time, it is unlikely that this difference is due to increased effort. Recall that subjects in both of the experimental groups were asked to rate their perceived difficulty of their assigned input task. The level of perceived difficulty did not differ between subjects in the sentence generation and definition generation groups as evidenced by a *t*-test between group means of ratings.

The differences in performance of the groups in the current study might be explained through a levels-of-processing approach (Craik & Lockhart, 1972). According to this theory, information can be processed in a number of ways and the depth of that processing influences the durability and strength of the memory trace. This depth of processing is believed to fall on a continuum from shallow to deep. Shallow processing leads to weak memory traces while deep processing (e.g., semantic based or meaning based processing) leads to more stable and enduring memory traces. Craik and Lockhart (1972) explained two types of rehearsal that lead to differential depths of processing: maintenance rehearsal and elaborative rehearsal.

Maintenance rehearsal involves rote repetition of an item's auditory representations. This results in shallow processing, which in turn, results in fragile memory traces. Subjects in the definition only control condition of the current study engaged in rote rehearsal. Thus, they likely processed the input information on a shallow level. That may explain their inferior performance on the three cued recall tests.

Elaborative rehearsal, on the other hand, requires the learner to engage in a deeper, more meaningful analysis of the stimulus. This type of rehearsal results in deeper processing of the input information and, in turn, stronger and more stable memory traces (Craik, 1979). Subjects in the definition generation condition of the current study were presented with meaningful sentences (including target words) from which they were to extract meanings of target words. This required them to engage in a meaningful analysis of the presented material in order to perform the assigned task. The more meaning a person extracts from a stimulus, the greater the depth of processing. Thus, it would follow that the task completed by subjects in the definition generation condition likely resulted in a deeper processing of the input material. Recall that although significant differences were not revealed between these two groups of subjects, when compared to subjects in the definition only control condition, definition generation condition subjects consistently performed better across time.

Subjects in the sentence generation group likely engaged in even deeper processing of the input information. In order for subjects in the sentence generation group to perform their assigned task (i.e., generate a meaningful sentence including the target word) they were required to engage in a meaningful analysis of the input information (i.e., target word and definition) that was somewhat different than the task required of subjects in the

definition generation group. The generation of personally meaningful sentences required these subjects to tie the new vocabulary to personally relevant and meaningful information. Thus, subjects were putting the information into their prior knowledge stores in ways that were meaningful to them. That is, by providing their own semantically rich context sentences subjects in the sentence generation condition may have almost automatically made the target words and definitions meaningful and relevant to them. According to the self-reference effect, a topic of the levels-of-processing approach, this would result in deeper processing and thus, superior retention (Rogers, Kuiper, & Kirker, 1977). The self-reference effect indicates that people better recall information when they have related that information to themselves. This helps to explain the superior performance of the sentence generation subjects on the three cued-recall tests.

Definition generation subjects consistently, though not significantly, outperformed definition-only control subjects across time. This might be interpreted to mean that the learning from context approach to vocabulary acquisition is better than providing words and their definitions alone (Crist & Petrone, 1977; Gipe, 1979). Due to the critical nature of the definition extraction task, it was important to determine the level of accuracy of subject-generated definitions. Thus, the definitions offered by subjects in the definition generation condition were analyzed and scored as to their accuracy. Completely accurate definitions and definitions that were near the meaning of the target word were given a score of 1 and definitions that were completely inaccurate were given a 0. Post hoc evaluation of accurate meaning extraction in this condition revealed a mean score of 19.24. This means that the unfamiliar vocabulary words were correctly defined by

subjects 87% of the time on average. Thus, subjects were able to accurately extrapolate intended word meanings from the context rich sentences.

The fact that the definition generation subjects did not significantly outperform definition only control subjects should not discount the potential benefits of a learning-from-context approach to vocabulary acquisition. Granted, the differences in performance were not significant, but they were in the predicted direction. The value of the learning from context approach in and of itself has been well established (e.g., Jenkins, Stein, & Wysocki, 1984; Nagy, Herman, & Anderson, 1985). Learning vocabulary from context has repeatedly been shown to be superior to learning vocabulary by being offered the word and its definition (e.g., Crist & Petrone, 1977; Gipe, 1979). This has been demonstrated in direct tests of the two methods with college students (Crist & Petrone, 1977) and elementary school children (Gipe, 1979).

In summary, there was significantly superior performance of sentence generation condition subjects over definition generation condition subjects and definition only control condition subjects. Subjects in the definition generation condition consistently outperformed definition-only control condition subjects, though without significance. Superior performance of subjects in the sentence generation group was likely due to the deep level of processing required by their experimental task. Their assigned task of creating personally meaningful sentences enabled them to tie target word meanings to prior knowledge. This is interpreted to mean that offering learners the opportunity to write personally relevant sentences, which include to-be-learned vocabulary words, significantly enhances vocabulary acquisition.

The nonsignificant performance advantage of definition generation condition subjects over definition-only control condition subjects was likely due to the fact that extrapolating word meanings required deeper processing than repeatedly writing the target words and definitions. The more meaning a person extracts from a stimulus, the greater the depth of processing. The task of repeatedly writing the target word and definition was a case of rote rehearsal. Maintenance rehearsal involves rote repetition and results in shallow processing, which in turn results in weak memory traces. This would explain the inferior recall of the subjects in the definition only control group.

Though definition generation subjects did not perform significantly better than control subjects, they consistently outperformed them across time. If the power for the time-by-treatment had been higher, significant differences control and experimental groups might have been observed. Thus, the learning from context approach to vocabulary learning is interpreted to have value and deserves further attention in research associated with the principles of the Generation Effect.

Limitations

The current study had three obvious limitations: 1) limited power, 2) subject pool peculiarities, and 3) considerable variability on the dependent measure scores. The absence of a significant time-by-treatment interaction was not surprising, as the power for that interaction was incredibly low (.145). This was likely due to the subject group sizes. This may have affected the potential emergence of the Generation Effect. Peculiarities in the subject pool were related to the sample size, and thus, power. In addition, subject pool issues may have contributed to problems beyond the problem of power.

Subjects employed for the current study were accessed through the Department of Educational Psychology during the 2003 spring and summer sessions. During these sessions, graduate students and faculty had difficulty finding subjects to participate in research. Study after study was posted, but few subjects signed up. Some researchers even directly recruited in educational psychology classes. It is likely that students opted for the alternative to participating in research—completing article summaries. During the semesters in question, the articles to summarize were available online (e-reserve). This made it possible for student to view the articles from any Internet capable computer. Thus, they were apparently less likely to participate in posted research projects (A. J. Corkill, personal communication, September, 2003). Subjects who did participate in the research, including the present study, appeared to do so reluctantly. Thus, it is appropriate to question their level of motivation.

Finally, there was great variability in scores on the dependent measures (immediate, 48-hour delayed, and 21-day delayed cued recall) for subjects across all conditions. Recall that the maximum score on the cued recall test (regardless of treatment condition) was 22 points. Scores on the immediate cued recall test had a range of 2 – 21, on the 48-hour delayed recall test scores ranged from 1 – 21.5, and the 21-day delayed recall test scores ranged from 0 – 21. This may be indicative of motivation variability on the part of study subjects. The wide variability may have also affected the potential emergence of the Generation Effect.

In summary, the current study had three apparent limitations: 1) limited power, 2) subject pool peculiarities, and 3) considerable variability on the dependent measure scores. Had there been greater power for the time-by-treatment interaction, more willing

and motivated participants, and lesser variability between subject performances on dependent measures, the Generation Effect might have appeared.

Future Research

The following section includes a discussion of several approaches that might be taken in future research. Each will be discussed in turn, followed by a summary of the potential research possibilities. First, the lack of power in the current study and how greater power might lead to different findings in future research is briefly discussed. Second, possible research combining the definition generation and sentence generation tasks is explained. Next, the potential value of a design change incorporating the scores on the Verbal Intelligence test as a blocking variable is discussed. Finally, the idea of comparing performance of subjects in a definition only group offered spaced presentations to performance of subjects in a sentence generation group offered single presentations is discussed.

The power of the current study was apparently limited by the sample size. Replication of this study with a larger sample might better effectuate a generation effect. A replication would also help to determine the extent to which the sentence generation and definition generation tasks lead to learning unfamiliar vocabulary. The task of generating a definition from experimenter provided context rich sentences was interpreted by the present researcher to be a more authentic generation task (according to the Generation Effect theory) than the task of generating meaningful sentences including target words. Recall that the accepted definition for the Generation Effect is that there is better recall for material that is self-generated as compared to material that is simply read (Slamecka & Graf, 1978; Graf, 1980). The definition generation subjects actually

generated the meaning of the word, whereas the subjects in the sentence generation condition generated meaningful sentences using the target words.

People often use methods similar to the definition generation task when they try to extract meanings of unfamiliar words from context. It is likely that people also use methods similar to the sentence generation task. That is, someone who knows the meaning of an unfamiliar word might attempt to help another learn it by using it in a sentence. Then the learner might check his/her understanding by offering his/her own sentence including the unfamiliar word. Because it seems that methods similar to both of the generation tasks in the present study are used by people who are attempting to learn new vocabulary, it would be interesting to investigate the results of having the same subjects use both generation tasks in an attempt to learn unfamiliar vocabulary.

This new condition would be a definition extraction/sentence generation condition. Subjects would be given a rich context sentence from which to extract the meaning of the target word. They would then use the target word in their own, personally relevant, meaningful sentence. This would result in a two-fold generation task. The relevant question, then, is would this two-fold approach lead to superior performance, or a stronger generation effect, over either of the single generation tasks.

In the current study, scores on the Verbal Intelligence test were used as a covariate. This allowed for the control of individual differences in verbal ability between subjects. An alternative approach would be to use verbal ability as a blocking variable. This would enable the researcher to determine which generation condition, if any, leads to better learning of subjects with particular levels of verbal ability. It is possible that people with lesser verbal ability would greatly benefit from one or the other of the generation tasks,

while people with better verbal ability would perform well regardless of assignment to generation or read-only tasks. People with higher levels of verbal ability are likely already efficient learners, thus an introduction of a new strategy (i.e., generation task) might not make a significant difference in their level of performance.

Perceived potential limitations of Dempster's (1987; 1989) study were previously discussed: insufficient time allocation for subjects in the experimental conditions and the spacing effect benefiting the read condition subjects more than the experimental condition subjects. The current study supported the notion of those perceived limitations. A further investigation including spaced presentations for control conditions versus single presentations for sentence generation conditions might determine whether Dempster's (1987, 1989) studies suffered from the fact that spaced rehearsal virtually always results in better retention. That is, it might determine if spaced, maintenance rehearsal is as beneficial as the act of generating personally relevant, meaningful sentences in the retention of to-be-learned vocabulary.

In this type of design, the spaced presentation group would see each of the unfamiliar target words and definitions and repeatedly copy each word and its definition. These subjects would be exposed to each target word and definition three times (spaced) at 20-seconds per presentation. The sentence generation group would be presented the target words and definitions one time and would be allocated 60-seconds to generate a meaningful sentence using the target word. The comparison of retention between the two groups would help determine whether Dempster's (1987, 1989) results were confounded by the power of the spacing effect. If the groups fail to differ in performance, then it would be necessary to address the fact that it might be just as beneficial to engage in rote

rehearsal over spaced presentations as to engage in single input elaborative rehearsal (sentence generation).

In summary, several approaches may be taken in future research: 1) the creation of a combined definition/sentence generation condition; 2) using scores of verbal ability as a blocking variable instead of a covariate; and 3) comparing performance resulting from single presentation of generate tasks and spaced presentations of read-only tasks.

APPENDIX A

OVERVIEW OF KEY EXPERIMENTS

Overview of Key Experiments

Study	Subjects	Materials	Conditions	Recall	Results	Contribution
Bobrow & Bower (1969)	University under-graduates	Exp. 1a: Noun pairs; $noun_1$-$noun_2$ pairs linked by sentence; $noun_1$-#-$noun_2$ triplets.	Exp. 1a: Generate: S's presented with pair, make up & said linking sentence. Read: S's presented linking sentence, read aloud & studied. Number: S's presented with N_1-#-N_2 triple, to retrieve verb associated with # & say sentence using that verb to link N_1 & N_2.	Exp. 1a: Cued recall (recall N_2 to N_1 cue) Cued recall (#'s 1-15 as cue, appropriate verb to link N_2 to N_1)	Exp. 1a: Means for generate, read, & number differed sig. (generate > number > read).	Exp. 1a: Evidence suggests that generate-read differences should not be attributed to active memory search or association.
		Exp.2b: 30 sentences containing an ambiguous word & two concrete nouns.	Exp.2b Continuation: S's read sentence aloud, made up & said sensible continuation. Repetition: S's read each sentence aloud 3 times.	Exp.2b Cued recall (recall N_2 to N_1 cue).	Exp.2b Recall was sig. better for continuous condition.	Exp.2b Generate & read conditions differ in comprehension of sentences & that comprehension aids retention.
Anderson, Goldberg, & Hidde (1971)	University under-graduates	Exp. 1.: 2 groups of 24 sentences: intact sentences & the same sentences with a blank for last word.	Exp. 1 3 study-test trials No blank: S's presented with complete sentences. Blank: S's presented with same sentences with blank for last word to be generated.	Exp. 1. Cued recall: cued by subject noun of a sentence to recall the last word of that sentence.	Exp. 1 Blank group performed sig. better on first test session.	Evidence that procedures that force meaningful processing facilitate learning, supporting Bobrow & Bower's (1969) comprehension hypothesis.
		Exp. 2. Same as Exp. 1.	Exp. 2. Same as Exp. 1, but with one presentation & test.	Exp. 2 Forward & backward cued recall (cue & target were reversed for backward test)	Exp. 2. Blank group performed sig. better on all measures.	

Study	Subjects	Materials	Conditions	Recall	Results	Contribution
Slamecka & Graf (1978)	University under-graduates	Exp. 1 & 2 100 word pairs blocked by 5 rules (associate, category, opposite, synonym, & rhyme).	Exp. 1 & 2: Read: S's presented with stimulus & response words to read & say aloud. Generate: S's presented with stimulus & first letter of response words. S's to read the stimulus word, generate the response word & say both aloud.	Exp. 1 & 2: Recognition test: sheet with 100 sets of 3 alternatives, one being the target response word, S's to indicate that word.	Exp. 1 & 2: Sig. better recognition for generate group.	Established the existence of the GE. They offer several theoretical interpretations for the effect (levels of processing, cue-target relationship enhancement, effort, semantic memory activation) & limitations of those, but commit to none.
		Exp. 3: 66 rhyme word pairs.	Ex. 3: Read or generate as within factor. 33 items presented as read & the remaining as generate. Group 1: informed of pending test. Group 2: not informed.	Exp. 3: Same procedure as Exp. 1 & 2. Two tests: response recognition & stimulus recognition.	Exp. 3: Generated items were sig. better recognized than read items. Large GE for responses, but not stimuli. Informed vs. not informed was inconsequential.	
		Exp. 4 60 word pairs: 20 opposite, 20 synonym, 20 rhyme. 10 of each as generate, 10 as read.	Exp. 4. Five alternating presentation-test trials. 2x3x5 Within-subject design (rules, input procedure, trials).	Exp. 4 Written free-recall of response words.	Exp. 4 Generated items were recalled sig. better. There was a sig. main effect for trials, & an interaction of condition (read, generate) & trials.	
		Exp. 5 66 rhyme word pairs	Exp. 5. Five alternating presentation-test trials. Read or generate as within factor. 33 items presented as read & the remaining as generate. Group 1: informed of pending test. To recall stimulus at test. Group 2: not informed of	Exp. 5 2 types of Cued recall: 1) stimulus cue, response target, & 2) response cue, stimulus target.	Exp. 5 Sig. main effects for read vs. generate, stimulus vs. response, & trials.	

Study	Subjects	Materials	Conditions	Recall	Results	Contribution
			pending test. To recall response at test.			
Graf (1980)	University under-graduates	Meaningful & anomalous sentences under read or generate formats. Read presentations offered anomalous sentences that were read aloud once. Generate presentations offered the words to be arranged in an anomalous sentence by the subject & read aloud once.	Exp. 1: Anomalous condition: S's received 8-read & 8-generate anomalous sentences. Meaningful condition: Same as above, but with meaningful sentences.	Exp. 1: Cued recall: 6-minute cued recall with verbs as cues & corresponding sentences (or parts of them) as targets.	Exp. 1: A GE was found for meaningful material only. Repeated Measures: Sig. main effects for materials & processing & an interaction of the two. Very low recall for anomalous condition.	Evidence that one consequence of generating is a greater degree of interword organization. The meaningfulness of input material is of critical nature in the GE, arguing that the interword organization is semantically based.
			Exp. 2 Same conditions as Exp. 1, but with three presentations.	Exp. 2 Same as Exp.1	Exp. 2 (ANOVA) Same as in Exp. 1. The anomalous condition benefited from 3 presentations, especially for generated items. Generate condition benefited from additional study trials.	
			Exp. 3 Same conditions as Exp. 2.	Exp. 3 Same test as Exp. 1 & 2, but given after each study trial.	Exp. 3 – GE for meaningful material only. Repeated Measures Sig. main effect for materials & trials; & sig. interaction for materials with trials & materials with processing condition.	
			Exp. 4 Group One: 12 S's received meaningful input & 12 received anomalous input (6 of each as read, 6 as generate), with one exposure to the list. Group Two: Same as in group 1, but with two	Exp. 4 Word-pair recognition test. Nouns of sentences arranged in 32 pairs: 16 from same sentence with 8 read & 8 generate, 16 from different	Exp. 4 Overall concurrence with Exp. 1, 2, & 3. Better performance for 2 exposures. ANOVA Sig. main effects for material & group; & an interaction for material with group & material with processing.	

Study	Subjects	Materials	Conditions	Recall	Results	Contribution
Graf (1980) continued			consecutive exposures to the list, without pause.	sentences, 8 read & 8 generate. S's to indicate if pair appeared in same or different sentence at input.		
			Exp. 5 Anomalous: 16 read & 16 generate format. Meaningful: 16 read & 16 generate format.	Exp. 5 Yes-No word recognition test. All words mixed with 128 distracters presented one at a time & S's were to indicate if they had seen the words at input	Exp. 5 There was a GE for both types of materials. Sig. main effect for processing condition, with no other sig. effects.	Exp. 5 Evidence that benefits of generating are not localized to the interword organization of input sentences, but show an increase in intraword organization as well. In generating a sentence, compared to reading one, a subject must pay close attention to individual words.
Ghatala (1981)	6th- & 7th- grade students	2 groups of 24 target sentences, intact sentences & the same sentences with a blank for last word	Read: S's presented with intact sentences to read aloud twice Generate: S's presented with same sentences with blank for last word to be generated & read aloud twice. Judgment: S's presented with intact sentences with the last word underlined. S's read the	Cued Recall: Forward test – subject noun as cue & last word of sentence as target response. Backward test – last word of sentence offered as cue &	ANOVA Sig. main effects for encoding condition & type of test. Generate & judgment conditions performed similarly & both outperformed read with significance. Test order did not affect the pattern of results.	Evidence that 6th- & 7th graders as well as college undergraduates exhibit the GE. The positive transfer explanation does not apply.

Study	Subjects	Materials	Conditions	Recall	Results	Contribution
Ghatala (1981) continued		(Anderson et al., 1971). Three additional filler sentences were used.	sentence & indicated if the underlined word properly fit the sentence.	subject noun of that sentence as target. Both tests were given to all S's.		Findings rule against cognitive operation involved in generating have a special mnemonic value. Generate advantages might be attributed to the comprehension hypothesis: generating requires the subject to meaningfully represent & integrate the words of a sentence.
McElroy & Slamecka (1982)	University under-graduates	Exp. 1 60 Paired associates: 30-word pairs & 30-nonword pairs.	Exp. 1 Within-subject factors- Read: S's to read aloud the stimulus & response Generate: S's to produce opposite for each stimulus word & use letter-transposition rule for nonword stimulus (all to be said aloud) Between-subject factors- Timed: 6-senconds per presentation Self-paced: Subject paced	Exp. 1 Recognition test: S's to indicate if item had appeared at input. Task allocation: to indicate if item had appeared as read or generate at input Confidence ratings for the above.	Exp. 1 GE for words on both measures, but not for nonwords.	Exp. 1 Evidence that the GE is semantically based.
		Exp. 2 30 nonword pairs	Exp. 2 Within-subject factors- Read: same as above	Exp. 2 Free-recall test- Five study-test trials.	Exp. 2 No GE for nonwords. Sig. effects for trials.	Exp. 2 Evidence that generation does

Study	Subjects	Materials	Conditions	Recall	Results	Contribution
McElroy & Slamecka (1982) continued			Generate: same as above Trials: 5 study-test trials	Timed at 4-minutes per test.		not always result in better retention (Exp. 1 & 2 show no GE for nonwords).
		Exp. 3 2 lists of 24 nonword pairs: 12 associated by rhyme rule & 12 by letter-transposition rule.	Exp. 3 Within-subject factors- For all S's: items presented twice, first in read & generate format, & then in read format. Rules: letter-transposition or rhyme. Tasks: Read or generated Trials: five study-test trials.	Exp. 3 Free-recall test- Five study-test trials. Timed at 3-minutes per test.	Exp. 3 No GE with rhyme or letter-transposition rules on any trial or across trials for nonwords.	Exp. 3 Evidence that the nonsemantic nature of material might lead to absence of GE with nonwords.
		Extension: Nonword-word pairs associated by rhyme rule Word-word pairs associated by opposites & rhyme rules	Extension: 8 subjects Same as above but with single test trial	Extension: Multi-trial free-recall tests	Extension: GE for both opposites & rhyme rule word-word pairs. No GE for word-nonword pairs.	Extension: Findings inconsistent with lexical activation view Exp 1-3 & extension: Evidence that benefits of generation are not an automatic consequence of the act of generating. Benefits are likely (at least partially) due to the involvement of semantic memory.

Study	Subjects	Materials	Conditions	Recall	Results	Contribution
Dempster (1987)	University under-graduates	Exp. 5 38 vocabulary terms and definitions.	Exp. 5 No-presentation baseline: S's not exposed to words before testing Control: three-spaced presentations. S's presented with word and definition 3 Context: three spaced presentations. S's presented with word and definition and three context sentences including the word.	Exp. 5 Sentence-cued recall for 19 items (given first) Word-cued-recall test with added instructions: if can't define word, try to write sentence using word. Both tests were timed at 10-minutes.	Exp. 5 Sentence-cued recall test was redundant across groups. The control group performed marginally better than 3-sentence context group.	Exp. 5 Failed to find evidence to support variable encoding. Exp. 1-5 Evidence to support that recall was not sig. affected by manipulations designed to affect the number of retrieval routes. Evidence to support spaced presentations in voc. learning.
Hirshman & Bjork (1988)	University under-graduates	14 first associate pairs & 14 third associate pairs	Exp. 1 Read group: S's received stimulus & response words (first & third associates) to copy Generate group: S's received stimulus & letter cues of response word (first & third associates). S's were to generate the response word & write both stimulus & response. Cued recall & free recall groups.	Exp. 1 Free recall S's given 5-min distracter task. Then to free recall & write response words. Cued recall S's given 20-min distracter task. Then to recall response words from stimulus word cues.	Exp. 1 a) Sig. main effects for encoding task & test type, & sig. interaction of those. GE larger in cued than free recall. GE sig. in cued, but not free recall. b) Sig. interaction for test type & associative strength. 3rd associates sig. better remembered with free recall, & 1st associate sig. better remembered with cued recall.	Exp. 1 Evidence that cued recall result in sig. larger GE. Cued recall more sensitive to relational factors than free recall. Item-specific explanation under lexical activation hypothesis not sufficient to explain GE.
			Exp. 2 Same as Exp. 1, except cued	Exp. 2 Same as Exp. 1	Exp. 2 a) Sig. interaction for test type &	Ex. 2 Analyses of

Study	Subjects	Materials	Conditions	Recall	Results	Contribution
Hirshman & Bjork (1988) continued			recall tested after 48 hours		encoding task. GE sig. larger for cued than free recall. GE sig. for cued, but not free recall. b) Sig. interaction for test type & associative strength. 3rd associates sig. better remembered for free recall & 1st associates sig. better remembered for cued recall. c) Sig. main effects for test type & encoding task. Sig. better performance for cued recall. Generated items sig. better recalled. Advantage of generate over read on 3rd associates for free recall.	results from Exp. 1 & 2 showed that GE in free recall was reliable.
			Exp. 2 extension Same as Exp. 1, except only informed free recall	Exp. 2 extension Free recall only	Exp. 2 extension a) Nonsig. advantage of generate over read on 1st & 3rd associates. b) Compared to combined-results of free-recall results from Exp. 1 & 2. Recall sig. better for in extension than in Exp. 1 & 2. Type of test did not interact with encoding task. For all 3 Exp., GE was sig. & 3rd associates sig. better recalled than 1st.	Exp. 2 extension combined with results from Exp. 1 & 2, Exp. 2 extension revealed that a small GE advantage shown in free recall with between-subject design. Evidence that one-factor item-specific theory, such as lexical activation not sufficient to explain GE.
			Exp. 3 Same as Exp. 1, except that all	Exp. 3 Same as Exp. 1	Exp. 3 a) Sig. main effects for test type	Exp. 3 Evidence to

Study	Subjects	Materials	Conditions	Recall	Results	Contribution
Hirshman & Bjork (1988) continued			S's were to only write the response word.		& encoding task. Sig. interaction of the two. b) Sig. GE in cued, but not free recall. c) Sig. interaction of test type & associative strength. 3rd associates sig. better free recalled. 1st associates sig. better remembered in cued recall. d) Sig. interaction of encoding task & associative strength. Sig. interaction of free recall & associative strength. 1st associates in the Read cond. sig. better free recalled than 1st associates in Generate cond. e) Numerical generation advantage for third associates.	support two factor (item specific, & stimulus-response relation) theory.
			Exp. 4 Same as Exp. 1, except that encoding task (read, generate) used as within-subject factor.	Exp. 4 Same as Exp. 1	Exp. 4 a) Sig. main effects for test type & encoding task. No interaction for the two. b) GE sig. for free & cued recall. Larger for cued recall.	Exp. 4 Evidence that within-subject designs effectuate larger GE than between-subject designs. Large, reliable GE in free recall with within-subject design.
Dempster (1989)	University under-graduates	Exp. 1 38 vocabulary terms & definitions.	Exp. 1: 3-spaced presentations Control: provided with words & definitions Context: provided with words & definitions & asked to covertly generate sentences using words.	Exp. 1 Sentence-cued recall for 19 items (given first) Word-cued-recall test with instructions: if can't define word, try	Exp. 1 No sig. difference between groups on either measure.	Exp. Evidence that providing contextual support fails to enhance vocabulary learning. S's

Study	Subjects	Materials	Conditions	Recall	Results	Contribution
Dempster (1989) continued				to write sentence using word. Both tests timed at 10-minutes.		covert generation of context sentences did not enhance learning.
		Exp. 2 Same as Exp. 1	Exp. 2 Same as Exp. 1, except two-spaced presentations	Exp. 2 Same as Exp. 1	Exp. 2 No sig. difference between groups on either measure.	Exp. 2 Same as Exp. 1
		Exp. 3 Same as Exp. 1	Exp. 3: one presentation Control: provided with words & definitions & asked to write them several times. Context: provided with words & definitions & asked to write one to three context sentences using the words.	Exp. 3 Same as Exp. 1	Exp. 3 Results favored control group. No sig. difference between groups on any measures	Exp. 3 Evidence that requiring students to generate & write context sentences may interfere with vocabulary learning.
McNamara & Healy (1995)	University under-graduates	Exp.2 30 word-nonword pairs	Exp.2 All S's exposed to list prior to pretest, & received 14 blocks of training. Read: presented with word to copy, & then nonword to copy. Process repeated after completion of list. Generate: presented with word to copy, to then generate nonword, then given nonword & to check accuracy & make corrections where necessary. Process repeated after completion of list.	Exp.2 Word-cued recall tests. Pretest: immediately after first exposure. Posttest: after training. Retention test: one-week after training. Mnemonic questionnaires: S's to report use of mnemonics for each item	Exp.2 GE for posttest & retention test. Sig. main effect for test reflecting learning form pre- to posttest & forgetting from posttest to retention test. Many S's in read condition reported trying to generate the nonword before being presented with it. A main effect for internally generating read S's approached significance. No difference in mnemonic scores between groups. For read condition there was a sig. correlation between mnemonic use, internal generation. S's who reported high mnemonic	Exp.2 Evidence that the GE can be extended to instances of learning new material. Evidence to support procedural account for the GE, where the critical factor leading to a generation advantage for learning new facts or skills is that cognitive

Study	Subjects	Materials	Conditions	Recall	Results	Contribution
McNamara & Healy (1995) continued					use & internal generating tended to score higher. Using mnemonic scores as a covariate revealed that generating enhanced learning beyond any effect of mnemonic use. Using training condition as a covariate, showed sig. better performance for S's who used mnemonics. Using mnemonic scores as a categorical variable showed that read S's with high mnemonic scores (who also tended to internally generate) scored comparably to generate S's. There were a sig. higher proportion of correct responses for items subject to semantic mnemonics, followed by nonsemantic mnemonics, & then no mnemonics.	procedures be developed at training & that these procedures be reinstated at test.

APPENDIX B

VERBAL INTELLIGENCE ASSESSMENT

Vocabulary Test – VB

This is a test of your knowledge of word meanings. Look at the sample below. One of the four numbered words has the same meaning or nearly the same meaning as the word at the top.

Indicate your answer by circling ON YOUR ANSWER SHEET the number in front of the word that you select.

Attempt
1-run
2-hate
3-try
4-stop

The answer to the item is number 3; therefore, a circle would be put around number 3 on the ANSWER SHEET.

Your score will be the number marked correctly. There is no penalty for guessing. Try to answer every question.

You will have 10 minutes for this test. When you have finished STOP. Do not go on until you are asked to do so.

MAKE NO MARKS ON THE QUESTION SHEET

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO.

1. rancor
   1-forbearance
   2-ridicule
   3-malice
   4-bravery

2. raucous
   1-empty
   2-quiet
   3-smooth
   4-harsh

3. gargoyle
   1-oil
   2-medicine
   3-carved waterspout
   4-ugly building

4. recrudesce
   1-purify
   2-renew activity
   3-lack refinement
   4-crush

5. specious
   1-plausible, but not genuine
   2-noteworthy
   3-class or variety
   4-roomy

6. bauble
   1-bubble
   2-showy plaything
   3-idle talk
   4-confusion

7. prolific
   1-scarce
   2-producing abundantly
   3-reckless
   4-speaking profanely

8. opulent
   1-party
   2-wealthy
   3-happy frame of mind
   4-semiprecious stone

9. coercion
   1-conspiracy
   2-strategy
   3-restraint
   4-attraction

10. hiatus
   1-animal
   2-calamit
   3-dread
   4-gap

11. germane
   1-microbe
   2-contagious
   3-relevant
   4-different

12. perfunctory
   1-fundamental
   2-formal
   3-superficial
   4-careful

13. diverge
   1-reveal
   2-chant
   3-distract the attention of
   4-differ or turn off from

14. evoke
   1-take away
   2-anger
   3-connect
   4-bring out

15. pertinent
   1-relevant
   2-lying next to
   3-necessary
   4-bold

16. holocaust
   1-entirety
   2-destruction
   3-saintly
   4-price

17. piquant
   1-mellow
   2-fish
   3-pungent
   4-cloth

18. firmament
   1-foundation
   2-heavens
   3-strong
   4-glue

19. bizarre
    1-market
    2-conventional
    3-odd
    4-imaginative

20. moral
    1-ethical
    2-esthetic
    3-mental state
    4-weak

21. implacable
    1-subdued
    2-relieved
    3-uncertain
    4-relentless

22. paradox
    1-ornamental box
    2-question
    3-infectious disease
    4-statement that says two opposite things

23. bigot
    1-foreigner
    2-cynic
    3-intolerant person
    4-insect

24. sumptuous
    1-luxurious
    2-sweet
    3-credulous
    4-cheap

25.tacit
    1-tactful
    2-elaborately developed
    3-unspoken but implied
    4-clever

26. harbinger
    1-forerunner
    2-well-tailored
    3-foruneteller
    4-port

27. panegyric
    1-medicine
    2-denunciation
    3-sports event
    4-laudation

28. cryptic
    1-grave
    2-escape
    3-hidden
    4-pretentious

29. descried
    1-described
    2-scolded
    3-saw
    4-denounced

30. querulous
    1-questioning
    2-complaining
    3-noisy
    4-agreeable

31. pecuniary
    1-involving money
    2-esthetic
    3-trifling
    4-unusual

32.carnage
    1-flower
    2-small eagle
    3-slaughter
    4-antique

33. subservient
    1-arrogant
    2-submissive
    3-undermining
    4-unnecessary

34. trepidation
    1-fear
    2-watering
    3-means of travel
    4-surgery

35. delineate
    1-limit
    2-straighten
    3-omit
    4-depict

36. preponderance
    1-statement
    2-dominance
    3-body of water
    4-thoughtfulness

APPENDIX C

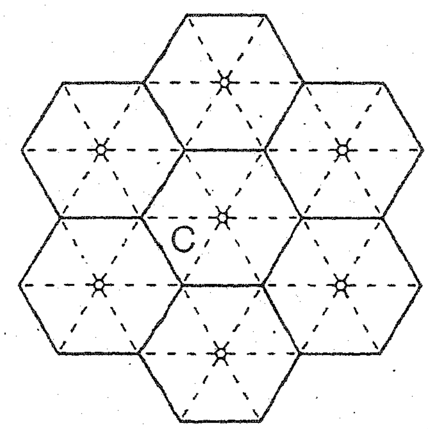
HEXAGON WORD MATCH PUZZLE

PUZZLE 3-S

Hexagon Match

Place the seven words into the hexagons so that each letter will match the letter in the adjacent hexagon. All the words will read in a clockwise direction. One letter has been entered to get you started.

CARTEL	PROFIT
CIRCLE	RECALL
EITHER	SEARCH
PREFER	



APPENDIX D

RECALL TESTS

Please attempt to define each of he words below. If you cannot define a word, try to write a sentence using the word.

	Circle Confidence level
1. Nadir	0 25 50 75 100%
2. Holm	0 25 50 75 100%
3. Burgee	0 25 50 75 100%
4. Odium	0 25 50 75 100%
5. Tarn	0 25 50 75 100%
6. Cognomen	0 25 50 75 100%
7. Cordite	0 25 50 75 100%
8. Nexus	0 25 50 75 100%
9. Éclat	0 25 50 75 100%
10. Cacophony	0 25 50 75 100%
11. Farceur	0 25 50 75 100%
12. Amulet	0 25 50 75 100%
13. Enigma	0 25 50 75 100%
14. Thaumaturgy	0 25 50 75 100%
15. Amanuensis	0 25 50 75 100%
16. Proselyte	0 25 50 75 100%
17. Foible	0 25 50 75 100%
18. Ratine	0 25 50 75 100%
19. Canard	0 25 50 75 100%
20. Travail	0 25 50 75 100%
21. Buskin	0 25 50 75 100%
22. Loggia	0 25 50 75 100%

Please attempt to define each of he words below. If you cannot define a word, try to write a sentence using the word.

	Circle Confidence level
1. Loggia	0 25 50 75 100%
2. Buskin	0 25 50 75 100%
3. Travail	0 25 50 75 100%
4. Canard	0 25 50 75 100%
5. Ratine	0 25 50 75 100%
6. Foible	0 25 50 75 100%
7. Proselyte	0 25 50 75 100%
8. Amanuensis	0 25 50 75 100%
9. Thaumaturgy	0 25 50 75 100%
10. Enigma	0 25 50 75 100%
11. Amulet	0 25 50 75 100%
12. Farceur	0 25 50 75 100%
13. Cacophony	0 25 50 75 100%
14. Éclat	0 25 50 75 100%
15. Nexus	0 25 50 75 100%
16. Cordite	0 25 50 75 100%
17. Cognomen	0 25 50 75 100%
18. Tarn	0 25 50 75 100%
19. Odium	0 25 50 75 100%
20. Burgee	0 25 50 75 100%
21. Holm	0 25 50 75 100%
22. Nadir	0 25 50 75 100%

Please attempt to define each of he words below. If you cannot define a word, try to write a sentence using the word.

	Circle Confidence level
1. Proselyte	0 25 50 75 100%
2. Éclat	0 25 50 75 100%
3. Farceur	0 25 50 75 100%
4. Canard	0 25 50 75 100%
5. Amanuensis	0 25 50 75 100%
6. Foible	0 25 50 75 100%
7. Loggia	0 25 50 75 100%
8. Ratine	0 25 50 75 100%
9. Thaumaturgy	0 25 50 75 100%
10. Enigma	0 25 50 75 100%
11. Tarn	0 25 50 75 100%
12. Travail	0 25 50 75 100%
13. Holm	0 25 50 75 100%
14. Buskin	0 25 50 75 100%
15. Nadir	0 25 50 75 100%
16. Cordite	0 25 50 75 100%
17. Cognomen	0 25 50 75 100%
18. Burgee	0 25 50 75 100%
19. Odium	0 25 50 75 100%
20. Amulet	0 25 50 75 100%
21. Cacophony	0 25 50 75 100%
22. Nexus	0 25 50 75 100%

REFERENCES

Anderson, A. C., Goldberg, S. R., & Hidde, J. L. (1971). Meaningful processing of sentences. *Journal of Educational Psychology*, 62, 395-399.

Bahrick, H. P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General*, 113, 1-29.

Bjork, R. (2001) How we learn versus how we think we learn: Implications for theory and practice. *Biomedical Science and Application Speaker Series*. University of Nevada, Las Vegas.

Bloom, K. F., & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *Journal of Educational Research*, 74, 245-248.

Bobrow, S. A. & Bower, G. H., (1969). Comprehension and recall of sentences. *Journal of Experimental Psychology*, 80, 455-461.

Brown, J. (1976). An analysis of recognition and recall and of problems in their comparison. In J. Brown (Ed.), *Recall and recognition*. London: Wiley.

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

Craik, F. I. M. (1979). Levels of processing: Overview and closing comments. In L. S. & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 447-461). Hillsdale, NJ: Erlbaum.

Crist, R. L. & Petrone, J. M. (1977). Learning concepts from context definitions. *Journal of Reading Behavior*, 9, 301-303.

Crutcher, R. J., & Healy, A. F. (1989). Cognitive operations and the generation effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 669-675.

Dempster, F. N. (1987). Effects of variable encoding and spaced presentations on vocabulary learning. *Journal of Educational Psychology*, 79, 162- 170.

Dempster, F. N. (1989). *Do sentence contexts facilitate vocabulary learning? Further evidence*. Unpublished manuscript.

Ekstrom, R. B., French, J. W., Harman, H. H., & Derman, D. (1976). *Kit of Factor Referenced Tests*. Princeton, NJ: Educational Testing Services.

Gardner, J., & Hampton, J. (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.

Glisky, E., & Rabinowitz, J. (1985). Enhancing the generation effect through repetition of operations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 193-205.

Ghatala, E. S. (1981). The effect of internal generation of information on memory performance. *American Journal of Psychology*, 94, 443-450.

Glover, J., Bruning, R., & Plake, B. (1982). Distinctiveness of encoding and recall of text materials. *Journal of Educational Psychology*, 74, 522-534.

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

Gipe, J. (1979). Investigating techniques for teaching word meanings. *Reading Research Quarterly*, 14, 424-464.

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

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.

Jenkins, J. R., Stein, M. L., & Wysocki, K. (1984). Learning vocabulary through reading. *American Educational Research Journal*, 21, 767-787.

Kane, J. H., & Anderson, R. C. (1978). Depth of processing and interference effects in the learning and remembering of sentences. *Journal of Educational Psychology*, 70, 616-625).

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M.M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625-632). New York: Academic Press.

Laying, W. G. (1994). *Epistemology and metacognition*. Unpublished master's thesis, University of Nevada, Las Vegas.

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

McNamara, D. S. & Healy, A. F. (1995). A generation advantage for multiplication skill training and nonword vocabulary acquisition. In A. Healy & L. Bourne, (Eds.), *Leaning and memory of knowledge and skills: Durability and specificity* (132-169). California: Sage Publications.

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly*, 20, 233-253.

Rea, C. P., & Modigliani, V. (1985). The effect of expanded versus massed practice on the retention of multiplication facts and spelling lists. *Human Learning*, 4, 11-18.

Reder, L. M., & Anderson, J. R., (1982). Effects of spacing and embellishment and memory for the main points of a text. *Memory & Cognition*, 10, 97-102.

Reynolds, J. H., & Glaser, R. (1964). Effects of repetition and spaced review upon retention of a complex learning task. *Journal of Educational Psychology*, 55, 297-308.

Rogers, T.B., Kuiper, N.N., & Kirker, W.S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, 35, 677-688.

Slamecka, N. J. & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology*, 4, 592-604.

Smith, S. M., & Rothkopf, E. Z. (1984). Contextual enrichment and distribution of practice in the classroom. *Cognition and Instruction*, 1, 341-358.

Tulving, E. (1962). Subjective organization in free recall of "unrelated" words. *Psychological Review*, 69, 344-354.

Tulving, E. (1966). Subjective organization and effects of repetition in multitrial free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 5, 193-197.

Tulving, E. (1976). Ecphoric processes in recall and recognition. In J. Brown (Ed.), *Recall and recognition*. London: Wiley.

VITA

Graduate College
University of Nevada, Las Vegas

Barbara Anne Badgett

Home Address:
1511 Dorothy Ave. #3
Las Vegas, NV 89119

Degree:
Bachelor of Science, Elementary Education, 2000

Publications:
Putney, L, G., Perkins, P., & Badgett, B. (2003, April). *Evaluation of Cultural Competency through Shared Professional Expertise.* Round table session presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Thesis Title: Vocabulary Acquisition and the Generation Effect

Thesis Examination Committee:
Chairperson, Dr. Alice J. Corkill, Ph.D.
Committee Member, Dr. Carolanne M. Kardash, Ph.D.
Committee Member, Dr. Gregory Schraw, Ph. D
Graduate Faculty Representative, Dr. Mark A. Guadagnoli, Ph. D.