The Generation Effect: A Detailed Analysis of the Role of Semantic Processing

Danielle S. McNamara

Institute of Cognitive Science
University of Colorado
Boulder, CO 80309

Technical Report 92-02

The Generation Effect: A Detailed Analysis of the Role of Semantic Processing

Danielle S. McNamara

University of Colorado, Boulder

Contents

I. Introduction

In 1978, Slamecka and Graf systematically examined a phenomenon that is often considered a simple and obvious characteristic of human learning. They were studying the notion that there is an advantage to learning by doing, that active involvement during the learning process is more beneficial than passive perception. Simply, their goal was to determine empirically whether a self-generated word is better remembered than one that is externally presented. Prior to Slamecka and Graf's report, other researchers had also been concerned with what is now called the *generation effect* (e.g., Anderson, Goldberg, & Hidde, 1971; Bobrow & Bower, 1969; Davies, Milne, & Glennie, 1973; Dosher & Russo, 1976; Erdelyi, Buschke, & Finkelstein, 1977; Gardiner, Craik & Bleasdale, 1973; Griffith, 1976; Jacoby, 1978; Johnson, Taylor, & Raye, 1977; Kane & Anderson, 1978; Pelton, 1969; Russo & Wisher, 1976; Schwartz, 1971; Schwartz & Walsh, 1974). Slamecka and Graf's (1978) study differed in that it focused solely on the generation effect, and the findings clearly demonstrated the robustness and generality of this phenomenon. Their methodology also overcame several methodological problems that weakened the validity of many previous studies. Foremost, Slamecka and Graf were careful to equate all aspects of the overt behavior of *reading* and *generating*. One way that they did this was to impose word generation constraints such that generated items would be identical to those presented in the read condition. This was accomplished by imposing generation rules which related the stimulus word to the response. The subject was provided with the stimulus word, the first letter of the response, and a rule relating the two. For example, with the rule of *opposite*, the stimulus *hot*, and the letter *c*, the word *cold* would be generated. The use of this paradigm eliminated confounding effects from idiosyncratic item selection habits that could give an unfair advantage to the generate condition. Additionally, because the stimuli are identical in both conditions, the possibility of an advantage for the read condition is also eliminated. This

methodology has since been used in the majority of the experiments investigating the generation effect.

Since Slamecka and Graf's (1978) report of their findings there has been a plethora of studies conducted to explore the breadth and limits of the generation effect. The extent of this literature, along with the great number of different theoretical accounts, renders a complete and unbiased understanding of this phenomenon difficult at best. One of the most debated issues that has emerged since Slamecka and Graf's report is the role of semantic activation or conceptual processing for the generation effect. The purpose of this paper is to review and critically analyze the findings of the studies which have focused on this issue, and to discuss the conclusions that can be drawn from this accumulated effort.

Slamecka and Graf (1978)

Slamecka and Graf (1978) reported the results of five experiments showing that the generation effect held across a variety of training and stimulus conditions. Because this report was used as a springboard for many of the generation effect studies to follow, and also because it remains one of the best studies on the generation effect, it is essential to describe their experiments and results in detail. First a description of the five experiments and their results is provided, followed by a summary of the results and the hypotheses generated by Slamecka and Graf as plausible explanations of this phenomenon.

In Slamecka and Graf's first experiment the subjects read or generated target words (between-subjects) following five rules of relatedness to cue words (i.e., associate, category, opposite, synonym, and rhyme; within-subjects). The presentation of the word cards was either timed (i.e., every 4 s) or self-paced. Of most importance, the results of this experiment, in terms of both the forced choice recognition test and confidence ratings, showed a main effect of condition (read vs. generate). This finding is what Slamecka and Graf then dubbed a generation effect. In addition, main effects of rule

type were found for the recognition measure and confidence ratings, though the main effect of presentation rate was not significant, nor were any interactions. The lack of an interaction of generate versus read with rules indicates that the generation effect is constant or invariant across the kinds of encoding operations used; this finding is consistent throughout the subsequent experiments using different rules.

Their second experiment addressed whether the generation effect would persist using a within-subjects design and without intentional learning instructions. Thus, generate versus read and rule type were within-subjects variables and type of instructions (informed vs. uninformed) was a between-subjects variable. They used a timed condition of 4 seconds for all subjects (this presentation rate was used for all three subsequent experiments), and again tested using recognition as the retention measure, with the added allocation task of indicating whether the item had been read or generated. The results of the first experiment were confirmed: read and generate were again significantly different using a within-subjects design; rule type was significant as a main effect, and no interactions were significant. In addition, there was no effect of instructions; that is, it made no difference whether or not the subject was aware that memory for the word-pairs would be tested. Also, performance was significantly above chance on the allocation task, though there was no difference between the read and generate items.

Experiment 3 examined the possibility that the act of generation leads to more attention being allotted to all aspects of the situation. If this were the case, then the stimulus word would also receive more attention in the generate condition as it constrains the selection of the response word. In contrast, the read condition might entail only superficial processing of the display. If this hypothesis were correct, then the generate condition would result in superior memory for the stimulus words as well as the associated response. Because the stimuli are never generated, the *generation effect* would

then be a misnomer. In this experiment, read versus generate was a within-subjects factor, and stimulus versus response recognition was a between-subjects factor. Informed versus uninformed was also included as a between-subjects variable, and cued recognition (with the pair-word as the cue) was used as the retention measure to equate the read and generate conditions in word accessibility, as backward access from recallable responses would thus be avoided. The only rule used was rhymes. The results showed a generation effect for responses and none for stimuli, both for the recognition measure and for confidence ratings.

The goal of the fourth experiment was to determine whether the generation effect would hold with free-recall testing. For this experiment, read versus generate and rules (synonym, opposite, and rhyme) were within-subjects variables and five study/test trials were included. There was, again, a main effect of generate versus read, a main effect of trials, and an interaction between the two factors (reflective of a greater difference between the two conditions in the first trials than the last). Again, the interaction of generate versus read with rules was not significant.

The purpose of their fifth and last experiment was to replicate the findings of the third experiment which examined the recognition of the stimulus words, this time using cued-recall as the retention measure. The decision to use cued- rather than free-recall was predicated on the necessity to equate the two conditions on functional accessibility of the untested member, because with cued-recall it would always be supplied. In this experiment, in contrast to the third, all subjects were informed that they would be tested and there were five study/test trials. Again, there were main effects of generate versus read, trials, and stimulus versus response. No interactions were significant, including the one of interest, that between generate versus read and stimulus versus response, which would have indicated that the generation effect happens solely for the response words and not for the stimulus words. However, the reported means are in a trend which would indicate a lack of a

generation effect with the stimulus words (.53 for generate and .51 for read). They did not, however, report separate analyses which may have supported this conclusion.

In summary, Slamecka and Graf found a generation effect within a wide range of circumstances: (1) cued and uncued recognition, (2) free and cued recall, (3) confidence ratings, (4) single and multitrial learning, (4) five different encoding rules, (5) both between- and within-subjects experimental conditions, (6) paced or unpaced presentation rates, and (7) both intentional and incidental testing instructions. In short, Slamecka and Graf clearly established the robustness of this phenomenon across a variety of experimental manipulations.

Slamecka and Graf addressed four classes of theoretical explanations of the generation effect. The first class of hypotheses, similar to approaches such as strength theory, frequency theory, or the law of exercise, relates performance differences to the consequences of different frequencies of overt or covert events. Simply, the generation effect would be a result of more practice occurring in the generate condition. This theory was rejected by Slamecka and Graf on the basis that the procedures used in Slamecka and Graf's experiments always equated the generate and read conditions on the amount of overt responding to each card at input, and also that the use of pacing permitted no more, and perhaps even less, covert rehearsal for the generate condition than for the read condition.

A second theory addressed by Slamecka and Graf focuses on the relation between the stimulus and response; specifically that generation would result in a distinctive encoding of this relationship, whereas reading the word-pairs which may not effectively demand that the relationship between them be processed. This theory could be neither confirmed nor disconfirmed by their findings in this study. However, the findings of a lack of a generation advantage for the stimulus words would imply that this is not the case.

A third hypothesis is that generation is actually an act of recall, and that an initial recall task benefits subsequent memory tests on the same material. However, a recall task generally involves retrieving information that originates from an external origin -- typically the experimenter, whereas the relevant information in the generate task used by Slamecka and Graf was self-produced. Slamecka and Graf argue that these two types of paradigms (i.e., recalling and generating) are not comparable due to the source of the target information. Nevertheless, assuming that generating is an act of recall, it could be further argued that this proposal approximates a restatement of the very findings that it seeks to explain, that is, a generated item is better recalled than an item that is read -- because it was generated (i.e., recalled).

The fourth explanation of the generation effect addressed by Slamecka and Graf is related to theories attributing memory differences to levels or types of processing (Craik & Lockhart, 1972; Hyde & Jenkins, 1969). According to these theories, deeper or more elaborate processing of an item at input will result in superior retention of this item and superior performance at testing. Deeper processing is defined as semantic in nature, whereas shallow processing is generally acoustic, visual, or otherwise superficial in nature. Slamecka and Graf offer three arguments against adopting this principle as an explanation of the generation effect. Their first argument is that this theory would predict a generation advantage for the stimulus words, as well as for the responses. They argue that according to levels-of-processing theories the stimulus would receive at least as much processing as the response because without adequately analyzing the stimulus the response could not be generated. It follows that if the stimulus is encoded at least at the same depth as the response, it should be as well remembered as the response. This prediction, however, was contradicted by Slamecka and Graf's results both in Experiments 3 and 5. Although, in Experiment 5, they did not find an interaction between learning condition and recall item, they did find a main effect of recall item

(response vs. stimulus). This indicates that overall, the responses were better remembered than the stimuli. Slamecka and Graf argue that this finding directly contradicts a levels-of-processing explanation. Secondly, they argue that a levels-of-processing view would predict not only a main effect of rules, which they consistently found, but also an interaction between learning condition and rules. According to Slamecka and Graf, the rhyme rule should produce only a shallow level of processing compared to the other rules, resulting in an attenuation or absence of the generation effect in its case, which is not what they found. Thirdly, they address the speculation that it is the act of generation itself, regardless of the encoding rule, that entails a deeper level of processing. They argue, however, that this hypothesis would be an extension of the levels-of-processing view and not an inherent prediction on its part.

Slamecka and Graf thus set the stage for the slough of studies to follow, the primary goal being to determine which of these hypotheses, and others to follow, would hold true under an even wider set of conditions. The hypothesis that has received the most attention has been that the generation effect is dependent on semantic activation occurring during the act of generation (e.g., McElroy & Slamecka, 1982). This hypothesis is most related to the aforementioned prediction that the act of generation entails a deeper level of processing than the read condition. This paper will describe the studies that have addressed the role of semantic processing and focus on the specific theoretical explanations of the generation effect and how these theories stand given the accumulation of findings.

Determining the role of semantic processing in generating can be approached from a number of different theoretical viewpoints and by way of many different methodological paradigms. The most fundamental issue is to determine whether semantic processing is a necessary and/or sufficient condition for the generation effect to occur. If it is a necessary condition, then the generation effect cannot occur for tasks which are void of semantic processing. If it is both necessary and sufficient, then the generation effect cannot occur for

tasks which are void of semantic processing, and any semantic information that is generated should show a retention advantage, regardless of any other factors. Of course, the case in which semantic processing is found to be only a sufficient condition is much less interesting. This means that the generation effect may or may not occur, independent of whether or not the generated information was semantically processed. Not much information is gleaned from this type of finding, though it would imply the existence of another, and perhaps related, construct necessary for the explanation of the generation effect.

There have been several methodologies applied to the problem of varying the presence or absence of semantic processing in generating. The underlying goal has been to vary orthogonally the degree of conceptual or semantic processing, as opposed to data-driven processing, along with the read and generate tasks. This has been attempted both by means of the task itself, and by means of what information is tested. The prediction from the viewpoint that semantic processing is a necessary condition of the generation effect, is that an effect will only occur for conceptual information and not for data-driven information. An obvious method of varying the degree of conceptual processing within the read and generate tasks is to manipulate the meaningfulness of the generated item, that is, to determine whether or not a generation effect can occur when the generated information is void of meaning.

II. Varying the Meaningfulness of the Generated Item

Graf (1980) conducted five experiments that examined the effects of reading versus generating either meaningful or anomalous sentences. He proposed a semantic framework which was based on the combination of two previous theories. The first theory was the comprehension interpretation (Bobrow & Bower, 1969) of the generation effect, which argues simply that the generated material is comprehended better than the read material. Furthermore, viewed from an organizational perspective (Mandler, 1979), generating a sentence results in an increase in the integration or interword

organization of the sentence. Graf predicted that the generation effect would be influenced by variables that affect the interword organization of sentences, in this case, whether or not the sentences were meaningful. Graf's subjects were presented either a complete sentence or a list of four words to be used to generate the sentence. The subjects were told to read the sentence aloud when presented intact, and to create the sentence using the provided grammar when the words were presented in list form. The type of sentence (meaningful or anomalous) was a between-subjects factor. The first three experiments all used a cued-recall test, with the verb of the sentences serving as the cue for the remainder of the sentence. These three experiments differed solely in terms of the number of study and test trials. In summary, for all three experiments the results showed a generation effect when the sentences were meaningful and the lack of a generation effect when the sentences were anomalous. The fourth experiment confirmed the same pattern of results using a word-pair recognition test as the retention measure.

Graf interpreted the results of these first four experiments as evidence that generating, in contrast to reading, results in an increase in the interword organization of sentences when the sentences are meaningful, but not when they are anomalous. That is, the generation of meaningful sentences strengthens the relationship between the words in the sentences, thus improving the subsequent recall of the sentence. These four experiments tested only the interword organization of sentences, that is, memory for two or more words within a sentence. Graf's fifth experiment tested whether the generation effect in this paradigm also extended to *intraword* organization (Mandler, 1979), that is, memory for individual words, apart from other words in the sentence. Thus, subjects in this last experiment were tested for the recognition of the individual words in the sentence. In contrast to the first four experiments, a generation effect was found for words presented in both meaningful and anomalous sentences. This was interpreted as evidence that generating a sentence results in the allocation of more attention to

each individual word "in order to place it correctly into the specified sentence frame."

The experimental paradigm employed by Graf (1980) is very different from that of Slamecka and Graf (1978) and other researchers studying the generation effect (but also see Gollub & Healy, 1987, and Griffith, 1976, for a similar paradigm). Foremost, previous researchers had required subjects to generate only one word rather than whole sentences. Thus, it is questionable how well Graf's findings and conclusions can generalize to the traditional results on the generation effect. Graf argues, however, that for the case in which subjects read or generate word-pairs (e.g., Slamecka & Graf, 1978), the act of generating forces subjects to bring to mind not only the word-pairs but also the relationship between the two items in the pair which "ensures that the pair is encoded into an organized memory trace." (p. 326). Graf's explanation of his data based on semantic organization of the memory traces is quite appealing. However, while Graf recognized that there was a difference between the paradigm used in this study compared with previous ones in the amount of what was generated (i.e., a sentence vs. a word), he did not note the distinctly qualitative difference between what was generated. It could be argued that what the subjects in Graf's experiment were doing was generating the organization of the sentence, not the words in the sentence. That is, while in previous studies words had to be generated based on cue words and relationship rules, in Graf's (1980) experiments the words were always provided in complete form, and the organization of these words is what was generated.

Although not mentioned by Graf (1980), nor elsewhere in the literature, Graf's results can also be understood in terms of the degree of automaticity and the processing demands of the task. Reading meaningful sentences is clearly a familiar and well practiced task. Furthermore, the process of organizing words to produce sentences is also a highly practiced task -- we do it everyday both when we speak and when we write. Thus, in the case in which the

subjects are shown meaningful sentences, both reading and generating (i.e., reorganizing and then reading) the sentences would be familiar and practiced tasks. In contrast, it may require more cognitive processing to reorganize words into sentences when the sentences are anomalous than when they are meaningful. It is well documented that when a task is more familiar and practiced, performance improves on secondary tasks. Thus, in the case in which subjects produce anomalous sentences, if this is not a familiar task it will require more cognitive processing. This additional processing leaves fewer cognitive resources to be devoted to processes which might aid in the recall of the words in the sentences. This explanation, however, leaves open the question of why, in the last experiment in which the subjects recalled only individual words, a generation effect was found regardless of whether the sentences were meaningful or anomalous. One possible argument, using the same framework of differing processing demands, is that while reading words and reorganizing them into anomalous sentences is not a familiar task, reading the words themselves is a highly familiar and practiced task. Thus, as explained above, when tested on intra-sentence associations, poor performance would be expected for the anomalous sentences. In contrast, when tested for the retention of the individual words, superior performance would be expected, and a generation effect would be expected to occur[1].

Realizing the possible effects of processing demands offers another theoretical framework with which Graf's (1980) findings can be understood. It explains why the generation effect should not be expected to occur when processing demands due to the unfamiliarity of the task possibly preclude other cognitive processes from occurring. Nevertheless, this framework does not explain what

---

1 On the other hand, it could be argued that the processing of the organization of the sentence is occurring at the same time, and using the same pool of resources, as the processing of the individual words. However, one important factor to consider concerns the point at which the organization of the sentence (i.e., the integration of the words in the sentence) is assumed to occur while reading a sentence. If it is assumed to occur only at the end of the sentence or clause (e.g., Kintsch & van Dijk, 1978), then this argument is not valid.

additional cognitive processes may be causing the generation effect, that is, it does not explain why the generation effect occurs when it does.

Nonwords as a Manipulation of Meaningfulness

Graf (1980) varied the degree of conceptual processing within the read and generate tasks by manipulating the meaningfulness of generated sentences. However, meaningless sentences still retain a meaningful element, namely the words. Assuming that nonwords lack semantic properties, if semantic processing is necessary for the generation effect to occur, it should never occur for nonwords. This is what McElroy and Slamecka (1982) proposed.

As a refinement to the hypotheses emphasizing the necessity of semantic activation, McElroy and Slamecka (1982) proposed what they called the *lexical activation hypothesis* as an explanation for the generation effect. According to this hypothesis the generation effect is dependent upon activation of a lexicon in semantic memory. Accordingly, the generation task would activate more of the attributes associated in semantic memory in comparison to the read task. Consequently, there would be a greater probability of accessing traces of the generated responses as these would be richer in the amount of information encoded at input. This theory was proposed to account specifically for findings regarding nonword data. In McElroy and Slamecka's first experiment subjects were presented both word pairs and nonword pairs. The word pairs were opposites whereby in the generate condition the subject would be presented *hot-c* on an index card and be expected to say aloud *hot-cold.* The nonword pairs were related by a formal letter-transposition rule requiring that the first three letters of the stimulus be put in reverse order following an initial consonant. For example, in the generate condition, given *dand-s*, the subject would respond *dand-snad.* Task (generate or read) and item type (word or nonword) were within-subjects factors. An additional between subjects factor was the presentation rate of the items: in one condition as soon as completing an item the subjects went on to the next one; in the other,

the subject studied each item for 6 seconds. Retention was tested with a recognition test. In summary, McElroy and Slamecka found a generation effect for the words, and no generation effect for the nonwords. To establish that the absence of a generation effect with nonwords was not due to the testing of recognition memory, their second experiment employed free recall testing. This experiment used only the nonwords, and gave five study/test trials with a 30 second filled interval between each trial. This procedure also failed to yield a generation effect. They concluded that that the generation effect would not occur when the material is not in semantic memory.

In order to interpret further the results of this first and second experiment, the nonword pair task could be understood in terms of processing demands. The process of reversing the letters in a nonword (or word) in order to produce another nonword, would be a very process consuming task. In other words, whereas reading letter strings (words or nonwords) is an automatic and highly practiced process, manipulating the letters in these words is not. The more automatic a process the more resources will be available for other processes or activities. Thus, in the case in which the subject is presented words, there should be additional resources available for other processes. In the generate condition this might result in more links or a greater activation of links between the stimulus and response. In the read condition, whether or not these additional resources were used to enhance the memorability of the word pairs may be a consequence of numerous factors, for example, the instructions given the subject or the motivation of the subject to learn the word pairs. In contrast, in the case in which the subject's task is to produce nonwords by transposing letters (i.e., in the generate condition), the processing demands may be too great to allow additional resources for the production or strengthening of links or traces between the stimulus and response. If this explanation were correct, the timed presentation conditions, allowing more time for the processing of the items, would result in superior retention of the nonword items compared to the self paced condition

for which no extra time is allotted. This is what McElroy and Slamecka found.[2]

Although there was a generation effect trend in the self-paced condition, in the timed condition the effect was reversed -- that is, the read condition yielded better performance than the generate condition. This result neither supports nor disconfirms the hypothesis that these results are due to processing demands of the task -- the generation task, compared to the read task, with nonwords still made more processing demands. The purpose of McElroy and Slamecka's third experiment was to establish that the letter-transposition rule was not the source of the generation effect with nonwords. In this experiment both a letter-transposition and a rhyme rule were used with nonwords. Also, all items were presented twice in succession, the second time in the read condition, to ensure that all items were seen (that is, regardless of performance in the generate condition). This experiment employed recall as the test and like the second experiment included five study/test trials. Again, no generation effect was found either with the rhymes or the letter transpositions, although it should be noted that in the rhyme condition the trend was in the direction of a generation advantage.

McElroy and Slamecka concluded that it was the nonsemantic nature of the material that led to the absence of the generation effect with nonwords, and furthermore that only one manipulation, the lexical status of the stimulus materials, determines whether or not the generation effect will occur. Based on these findings, they argued against the hypothesis that the generation effect was due to greater effort expended during generation. They also argued against the hypothesis that the generation advantage was due to relational

---

2 McElroy & Slameka (1982) found a main effect of presentation rate, that is, an advantage for the timed condition for both words and nonwords. It is interesting to note that Slamecka and Graf's (1978) first experiment found no effect of presentation rate for words. However, their timed condition was 4 seconds while McElroy and Slameka's was 6 seconds. The extra 2 seconds may be sufficient to allow extra processing resulting in a difference between the two conditions.

information due to the lack of the generation effect with rhyming nonwords.

In summary, several problems regarding the conclusions of McElroy and Slamecka remain. First, the nonwords clearly differ from the words used in several ways apart from their lexical status, one of the ways being the processing demands of reading or producing them. Second, in the second and third experiments only nonwords were used, thus they did not test for the critical interaction of item type (i.e., word vs. nonword) and learning condition (i.e., read vs. generate). Finally, the second and third experiments included five study/test trials which results in the paradigm in which both the read and generate items are recalled every other trial. Hence, read items are not solely "read"; read items are being "generated" as many times as they are read, in effect, half as many times as the generate items. The use of this methodology may wash out effects that may have emerged if the read items had been solely copied or read.

Nairne, Pusen, and Widner (1985) furthered the investigation of the role of the lexical status of the generated information. In each of the three experiments, subjects were asked to generate or read nonwords while the instructions given subjects regarding the status of the nonwords was the principle variable manipulated. They employed the methodology in which the first item and the first letter of the second item was provided and the subject was to read or generate the rhyming pair. The goals of Experiment 1 were first to replicate McElroy and Slamecka's (1982) findings, and second to determine whether knowledge about the lexical status of the items could account for the failure to find a generation effect with nonwords. One group of subjects was told that all items were experimenter-created nonwords; the other that the nonwords were rare English words. Although Nairne et al. reported that "no generation effect was obtained for either group", they did report a significant interaction between trials (there were five testing trials) and the read/generate variable. Indeed, the graphical presention of

the results indicated a greater improvement over the five trials for the generated items than for the read items. Moreover, contrary to Nairne et al.'s hypothesis that a generation effect would only occur when the subject were told that the nonwords were rare words, the improvement over trials for the generated items is particularly evident for the subjects who were explicitly told that the nonwords were indeed nonwords. This significant result, along with the graphical presentation of the means, should have indicated a generation advantage over the five trials. It can be concluded from the results of this first experiment that whether the subject was instructed that the nonword was a word or a nonword made no difference. In addition, the group of subjects told that the nonword was a word showed the least evidence of a generation advantage. Therefore, as Nairne et al. predicted, a **failure** to find the generation effect with nonwords (i.e., McElroy & Slamecka, 1982) cannot be ascribed to knowledge about the lexical status of the to-be-remembered items. Nairne et al. also concluded that the results of the first experiment constituted a replication of McElroy and Slamecka's (1982) findings of a lack of a generation effect with nonwords. This conclusion is, however, only valid if you look at the first study/test trial, in which the read condition displays superior performance than the generate condition for the "nonword" group and approximately equivalent performance for the "words" group. If, on the other hand, the results are considered in terms of the effect of generating on learning, that is, over the five study/test trials, then the results indicate that generating, in contrast to reading, may have a positive effect on learning, even for nonwords. Moreover, it also should be considered that the read items were recalled (or "generated") after each of the five trials. If there had been only two memory tests -- a pretest after the first trial and a posttest after the fifth trial, a significant generation advantage may have emerged.

The goal of Nairne et al.'s (1985) second experiment was to determine whether the assignment of meaning to the nonword via a provided definition would make it more memorable when generated than when read. In contrast to the first experiment, input condition

(read vs. generate) was a between-subjects factor in this second experiment and whether or not the nonword was defined was a within-subjects variable. There were two testing trials; first however, all subjects learned to criterion the meanings of half of the nonwords (all nonwords were presented during this phase). There, again, was not a main effect of input condition. The results showed significantly superior recall for the defined nonwords than for the undefined nonwords, a main effect of trials with superior performance on the second trial, and, again, a significant interaction between trials and input (read/generate) condition. The finding of a significant interaction between trials and input condition is again difficult to interpret. In the case of the undefined words, there appears to be a negligible difference between the read and generate conditions, with a very slight generate advantage. In the case of the defined nonwords, the read condition shows a clear advantage over the generate condition after the first trial. However, while the read condition shows no learning, and even slightly worse performance, from the first to the second trial, the generate condition showed much greater learning after the second study/test trial. Here again, when the results are considered in terms of learning, the failure to find a generation effect becomes debatable. Although this second experiment was subsequently replicated using a completely within-subjects design yielding no hint of a generation effect, there was only one recall trial, and thus not reflective of learning in the sense of the reported second experiment.

Nairne et al.'s (1985) third experiment examined the role of word frequency for the generation effect with the reasoning that if the effect depends upon the number of associates to the to-be-remembered item, then a generation effect would only be expected for those items with a high number of associates. The generate task differed from the first two experiments in that the subjects were asked to switch the first two letters of an item if underlined and write down the resulting word or nonword. If no letters were underlined (i.e., read items), the subject was to simply copy down the presented item. There were four types of to-be-remembered items:

nonwords, low-frequency words, medium-frequency words, high-frequency words. Subject's were given a recognition test for the items following a 1-minute distractor task. As Nairne et al. predicted, a generation effect was found for only the medium- and high-frequency words, and not for the nonwords and low-frequency words. Nairne et al. concluded that, in dispute of the lexical activation hypothesis, the representation of the generated item in the mental lexicon is not sufficient for producing a generation advantage. Rather, not only is the activation of the item's location important, but also the activation of associated concepts, ". . . thereby creating a network of potential retrieval routes for the activated item."

A final interpretation of Nairne et al.'s (1985) findings is not straightforward. The results of the first two studies seem to indicate that for single-trial learning, no generation effect occurs for nonwords, lending some support for the lexical activation hypothesis. However, when subjects are given multiple trials and are learning the associated pairs, this learning process, even with nonwords, may be benefited by the generate task. The results of the third study seem to indicate that, for single-trial exposure, the generation effect may be dependent upon the degree to which the to-be-remembered item is associatively linked to other words or concepts. Nevertheless, there are two important criticisms of the paradigm employed in this third experiment. First, similar to Graf's (1980) study, the subject's task in the generate condition was to generate the organization of the response term, not to complete an item and thus generate the response term. This alone renders questionable the generalizabilty of these findings to conclusions concerning the generation effect. Secondly, the use of this paradigm results in a major confound -- the difference in the stimulus items presented in the read and generate tasks. In the traditional read task the subject is provided with a stimulus term which is linked by a rule to a presented response term and the subject copies the response term. In the generate task the subject sees the stimulus term and a fragment of the response term and the subject provides the complete response term. What is critical to the traditional read/generate paradigm is that the stimulus and response and the overt behavior of writing the response

be **identical** for both tasks. In contrast, Nairne et al. (1985) did not equate the stimulus terms for the read and generate tasks. As explained above, the generate items were presented with the first two letters underlined and the subject was to switch the letters to produce the response. Thus, the stimulus and response in the generate task were different. In contrast, the read items were presented intact and were to be simply copied; thus, the stimulus and response were identical in this case. This paradigm, in effect, renders the read and generate task stimuli incomparable.

Payne, Neely, and Burns (1986) point out that both McElroy and Slamecka (1982) and Nairne et al. (1985) potentially confounded the lexicality of the stimulus and response items, and that the failure to obtain a generation effect with nonwords could be due to the lexicality of the response item (i.e., the nonword), the stimulus, or both. They tested this possibility by orthogonally varying the lexicality of the stimulus and response items. A rhyme rule was used to generate both word and nonword responses and the stimulus and response always differed only by the first letter. In their first experiment stimulus type (word vs. nonword) and response type (word vs. nonword) were between-subject factors, and task (read vs. generate) was a within-subjects variable. In the second and third experiments all of the above factors were within-subjects while in the third experiment subjects were required to say aloud whether or not each response term was a lexical item. For all three experiments there was a generation effect only when a word was the response item (i.e., word-word, nonword-word) and not when the response item was a nonword (word-nonword, nonword-nonword). These results are consistent with McElroy and Slamecka's (1982) lexical activation hypothesis and replicate the absence of a generation effect with nonwords in the paradigm of single-trial learning.

Nairne and Widner (1987) proposed that a crucial factor which had led to the failure to find a generation effect with nonwords was an incompatibility between the initial item generated and the item tested, and that *test appropriateness* was necessary in order to test validly a

generation effect with nonwords. The idea of test appropriateness (see also Rabinowitz & Craik, 1986) is similar to the concepts of encoding specificity (Tulving, 1983) or transfer-appropriate processing (Morris, Bransford, & Franks, 1977). In previous studies, the generate task had consistently consisted of generating the remaining fragment of the provided letters in the response term. Thus, what is actually generated is not the complete response term, but the missing fragment, whereas it is the memory for the complete response term that is tested. Because with nonwords the complete response term is meaningless, the functional product of generation may be merely the fragment generated. Their first experiment was designed to examine memory for both the generated fragment of the response and the complete response term. The generate task consisted of switching two underlined letters in the response term and writing down the resulting word or nonword and underlining the letters that had been switched; and the read task simply involved copying down the presented item along with the underlined letters. This methodology is similar to the paradigm used by Nairne et al. (1985, Experiment 3). The retention test consisted of a recognition test of the underlined letters in the response terms. This test required the subject to choose between two response terms that were identical apart for which letters were underlined. Following the recognition test for the underlined letters, the subjects were given a free-recall test for the response items. As predicted, Nairne and Widner found a generation effect for both nonwords and words for the recognition of the originally underlined letters, and a generation effect only for words for the recall of the response terms. This finding, they concluded, shows that when memory for the generated elements (i.e., the underlined letters in this case) is tested, a generation effect will occur regardless of the lexical nature of the response term.

Although the findings of Nairne and Widner's (1987) first experiment are appealing, it is again essential to point out that the paradigm used in this experiment, as in Nairne et al. (1985, Experiment 3), fails to equate the stimulus terms for the read and generate tasks. As discussed above, it is critical to equate the read and generate tasks in terms of stimuli, responses, and overt behavior, to examine

adequately the effects of generation. Not only do the differing read and generate stimuli confound the results of this experiment, the findings can be explained more simply in terms of this confound. In effect, the generate task provides two exposures to the underlined letters with different contexts for the two exposures, whereas the read task provides two exposures to the underlined letters, but with identical contexts. Thus, according to the encoding variability hypothesis (Madigan, 1969; Martin, 1972), one would predict superior recognition for the underlined letters in the generate condition due to the increased number of contextual cues (i.e., for the underlined letters). Thus, the concept of encoding variability can account for the generate advantage for the underlined letters for both words and nonwords. However, there was a generation advantage for the **recall** of words but not for nonwords. What can account for this finding? This result can be ascribed to the unfamiliarity of the nonwords in comparison to the words. The percent recall for both the copied (read) and generated nonwords was very low in comparison to the recall performance for words. It is possible that repeated exposures to the nonword could eliminate possible floor effects reducing effect sizes. In fact, Neal & Healy (1987) used the same paradigm as Nairne and Widner and found a generation effect for the recall of the nonwords after ten exposures to the nonword. Thus, when performance is raised above floor levels, the generation effect can emerge even for the recall of the nonwords.

The purpose of Nairne and Widner's (1987) second experiment was to examine the effects of generating at both the encoding and testing phase of the experiment. Unfortunately the confound of differing stimulus terms for the read and generate task was not eliminated in Experiment 2 of this study, and the encoding phase was identical to that used in Experiment 1. At testing, subjects first either read or generated (switch the letters if underlined) the item and then indicated whether the item was old or new. Of most interest is the finding of a generation effect with nonwords in the generate-generate condition. They interpreted this as evidence that when attention is directed to the relevant surface features (presumably through the

process of generating prior to recognition), then the generation effect will occur regardless of what is actually generated.

Unfortunately, the incomparability of the read and generate task in the paradigm used in by Nairne and Widner (1987) again clouds the findings. Nevertheless, the idea that the failure to obtain a generation effect with nonwords is due to a lack of test appropriateness remains a viable hypothesis.

Johns and Swanson (1988) pointed out a potential confound for the previous studies that had examined the generation effect with nonwords that is directly related to the concept of test appropriateness proposed by Nairne and Widner (1987). They noted that one characteristic of the generate condition for nonwords is that the subjects never see the generated item in its entirety, apart for in handwritten form. Thus, the items in the generate condition are never seen in the same typewritten form as in the test. In the case of nonwords, which the subject is unlikely to have ever encountered prior to the experiment in typewritten form, this would give an advantage to the read condition due to test appropriateness. Therefore, in the four experiments in this study that examined the generation effect with nonwords, items presented in the generate condition were presented in their entirety after an interval of 4 seconds.

McElroy and Slamecka (1982) had briefly alluded to an unreported investigation which had failed to find a generation effect with nonword-word pairs. The purpose of Johns and Swanson's (1988) first experiment was to establish that a generation effect could occur for words regardless of whether the cue was a word or a nonword. In this experiment the generate task consisted of reversing the letters in a word (e.g., live) or nonword (e.g., egap) to produce a word (i.e., evil and page, respectively). The read items (i.e., both the cue and target) were presented intact. A generation effect occurred, regardless of the cue, for both the free recall measure and a subsequent recognition test.

Johns and Swanson's (1988) subsequent experiments in this study established that a generation effect could occur for nonwords

in the case of a nonword-word pair and a nonword-nonword pair. However, Experiment 2 failed to find a generation effect in the case of a word-nonword pair. The generate task in Experiments 2, 3, and 4, consisted of completing a word fragment so that it would rhyme with the stimulus word or nonword. For these experiments, memory for the responses was measured by a forced-choice recognition test with two alternatives. The results of these three experiments established that given this paradigm, a generation effect could occur for nonwords, regardless of whether the nonword was "wordlike" or not (Experiment 3), but that the effect was dependent upon feedback being given in the generate condition (Experiment 4). However, you may recall that McElroy and Slamecka's (1982) third experiment also provided feedback in the generate condition; that is, all items were presented twice in succession, the second time in the read condition, to ensure that all items were seen. Thus, it is not clear why there would be a difference in findings between Johns and Swanson's experiments and McElroy and Slamecka's third experiment. One difference is that McElroy and Slamecka employed free recall as the retention measure, whereas Johns and Swanson used a recognition test. Another possibility is that McElroy and Slamecka tested fewer subjects (N=16) and perhaps simply lacked the statistical power to detect a difference between the read and generate groups. This former supposition would be supported by the previously noted, though nonsignificant, trend in the rhyme condition in the direction of a generation advantage.

Two recent studies have confirmed Nairne and Widner's (1987) and Johns and Swanson's (1988) finding that it is indeed possible to obtain a generation effect with nonwords (Greenwald & Johnson, 1989; McNamara & Healy, 1991). Greenwald and Johnson (1989) were actually focusing on memory for cues in the generation task, but in their third experiment examined memory for word-pair cues and a nonsense word that resulted from the combination of the cues. Specifically, in a within-subjects design, in the generate condition subjects were given a word-pair, such as *doctor-night*, and asked to respond with the first letter of the first word combined with all but

the first letter of the second word (i.e., *dight*). Half of the word-pair cues were related (e.g., day-night) and the other half were not (e.g., doctor-night). Each word-pair cue was presented for 5 seconds either with (read task) or without (generate task) the nonsense syllable response. The response was then presented alone for 3 seconds for both tasks, thus providing the feedback, and hence test appropriateness, shown previously to be critical to obtain a generation effect with nonwords. As Greenwald and Johnson (1989) predicted, a generation effect was found in free recall for the nonword responses that were cued by related word-pairs and not for those cued by unrelated pairs. However, 75 percent of the recalled responses were ones for which there was full or partial recall of the word-pair cue, and for 96 percent of these, the nonword response was spontaneously written next to the word-pair cue. This can be interpreted as evidence that the generation effect for the nonwords was mediated by the recall of the related word-pair cues. Although this finding may be expected, because the items were associated during the task and hence would be recalled together; it also could be argued that the nonword itself was not recalled, but rather that the word-pair cue was recalled and the nonword was reconstructed during the retention test from the word-pair. Therefore, because of the nature of the task, it could be argued that the study by Greenwald and Johnson (1989) is not a good example of a generation effect with nonwords.

McNamara and Healy (1991, Experiment 2) examined the long-term memorial consequences of generating versus reading nonwords. Their study differed from previous studies of the generation effect with nonwords in three ways: (a) The read and generate task was treated as a between-subjects variable, (b) subjects were given 14 blocks of training on the nonwords, and (c) both immediate and long-term retention was measured. Subjects were presented English word-nonword pairs under the guise that nonwords had the same meanings in a foreign language as the corresponding English words. Subjects were trained over two consecutive days for approximately one hour each day. Generate subjects wrote down the English word

and then attempted to produce the corresponding nonword. They were then shown the correct nonword, which they copied if their generated answer was incorrect. Read subjects simply copied the English word and nonword consecutively. Subjects recall of the nonwords was tested on the first day before training (after 10 minutes exposure to the word-nonword pairs), on the second day after completing training, and again one week following the training sessions. They found that those subjects who had generated the nonwords during training had significantly better immediate and long-term retention of the nonwords than did those who had simply copied the pairs. In addition, it was found that within the read condition, those subjects who had spontaneously utilized mnemonic strategies to learn and retrieve the nonwords were indistinguishable from the generate subjects at posttest and retention test. McNamara and Healy (1991) understood these findings in terms of theories of proceduralization. They hypothesized that subjects in a generate condition are more likely to develop procedures during training than are subjects in a read condition, and that these procedures, developed during training, are subsequently reinstated at test.

It could be reasonably argued that McNamara and Healy's (1991) findings do not provide evidence that lexical or semantic activation is unnecessary for the generation effect on the grounds that subjects who recalled the nonwords had transformed the nonwords into meaningful items via mnemonic codes. However, the use of mnemonics only resulted in differences among the subjects in the read condition; subjects in the generate condition did not differ as a function of whether or not mnemonics were employed. It could be further argued that perhaps researchers such as Johns and Swanson (1988) obtained a generation effect with nonwords because their subjects also had developed some kinds of mnemonic procedures to recall the nonwords. However, given that the relevant task used by Johns and Swanson was to link two nonwords via a rhyming rule, the use of semantic mnemonic procedures would seem unlikely. Since they had no measure of covert strategies, there is no way to answer this question without further evidence. Nevertheless,

the bulk of the evidence indicates that as long as the subject has the opportunity to see the nonword in the same form as when it is tested, then a generation effect with nonwords can be obtained. McNamara and Healy (1991) showed that this is the case even with a between-subjects design and with a retention interval of one week. Thus, the previously supposed lack of a generation effect with nonwords no longer provides clear evidence of the necessity of lexical or semantic activation for a generation effect to occur.

Further Manipulations of Meaningfulness

The lexical activation hypothesis predicts that the involvement of semantic memory is necessary for a generation effect to occur. The most evident stimuli with which to test this hypothesis have been nonwords; nonwords are clearly void of semantic properties. Researchers have also examined the possibility of a generation effect with other types of information of varying degrees of meaningfulness, for example: numbers, letter bigrams, and low frequency words.

Gardiner and Rowley (1984) showed, in a within-subjects design, that a generation effect occurs in both recognizing and recalling which answers to simple multiplication problems were encountered during either a single trial or a multitrial exposure. Crutcher and Healy (1989), also using a within-subjects design, replicated Gardiner and Rowley's finding of a generation effect for the recall and recognition of the answers to simple multiplication problems. In addition, they showed that this generation advantage survived both a two-day and a one-week delay interval.

The finding of a generation effect for numbers discounts only a strong version of the lexical activation hypothesis which would suppose that only words and perhaps their associated concepts would make up the mental lexicon. A weaker version would simply maintain that numbers as well as words are represented in the mental lexicon. It is important to note that McElroy and Slamecka never directly suggested that numbers would not be included as such, and indeed, numbers are read aloud and can be written as words.

Gardiner and Hampton (1985, Experiment 1) had subjects select two specific letters from a string of five letters and reverse them to produce either a meaningful or a meaningless letter bigram (e.g., ET or EC, respectively). In both the generate and read tasks, the subject called out each letter from the string, said "equals", then the letter bigram (e.g., *G T N E I* equals *E T*). In the generate condition, the bigram was not provided. A generation effect occurred only for the meaningful letter bigrams. Gardiner and Hampton (1985, Experiment 2) then used the same paradigm with numbers, such that subjects were given a string of five numbers and were to read or generate two specific numbers. They found that when subjects generated or read the numbers in a unitized form (e.g., twenty-eight) versus in a nonunitized format (e.g., two, eight), a generation effect occurred only for the unitized numbers. They then applied the same paradigm to a series of words (Experiment 3) whereby the subject was shown a set of five words which constituted a noun phrase (e.g., a cake made of cheese, a cake made of tomato), and was asked to reverse the two constituent nouns to produce either a familiar noun compound (e.g., cheese cake) or an unfamiliar one (e.g., tomato cake). A generation effect was observed only for the familiar noun compounds. They concluded that generation effect should depend neither on lexicality nor on the number of associative links in memory (Nairne et al. 1985), but on an integrated conceptual representation in semantic memory.

Gardiner, Gregg, and Hampton (1988) attempted to replicate Nairne et al.'s (1985, Experiment 3) finding of the absence of a generation effect with low frequency words. However, in a series of four experiments they found, to the contrary, a consistent generation effect with low frequency words both with recognition and recall measures, with both single trial and multitrial learning, and regardless of the retention interval (i.e., 30 s, 20 min, 80 min, and 24 hr). However, they employed the same paradigm as Nairne et al. (1985), that is, the generate subjects reversed the letters in a word, and the read subjects simply copied the word. As explained above in the context of the Nairne et al. study, this methodology is intrinsically

flawed in that the generate and read conditions differ in terms of the stimuli presented to the subjects. On that basis alone, both the Nairne et al. (1985) and the Gardiner et al. (1988) studies are difficult to interpret.

Nairne and Widner (1988) suggested that the difference between the two studies' findings was due to differences in familiarity of the low frequency words. They pointed out that the relationship between the frequency of a word and the familiarity of that word is not perfect, and thus low frequency words that were more familiar to the subjects (e.g., ladybug or dinosaur) may produce generation effects. As in the previous experiments conducted by Nairne and his colleagues, the two experiments reported in this study also employed the letter switching task in the generate condition and the simple copying task in the read condition (methodological weaknesses of this task are discussed above). All variables for both experiments were manipulated within subjects. In their first experiment subjects read or generated nonwords, unfamiliar low frequency words, and familiar low frequency words. Subjects also made a lexical decision for all of the items presented. In the first experiment subjects were given an incidental recognition test after a 10 minute distractor interval. Their second experiment was similar but involved a five-day interval before the recognition test. As they predicted, for both experiments there was a generation advantage only for the low frequency words that were more familiar, and thus contrary to Nairne et al.'s (1985) finding, low frequency words do produce generation effects, as long as they are rated as familiar to the subject. Because low frequency words that were nevertheless recognized as words produced no generation effect, they concluded that lexical representation of the generated item is neither a necessary nor a sufficient condition for a generation effect to occur.

In summary, the findings regarding the role of semantic processing in the generation effect with research employing stimuli such as numbers, letter bigrams, and low frequency words remain inconclusive. The generation effect was found for the recall of the answers to previously solved simple multiplication problems (Crutcher

& Healy, 1989; Gardiner & Rowley, 1984), showing that the generated item does not have to be a "word" for the phenomenon to occur. There is also evidence contrary to Nairne et al.'s (1985) findings that the generation effect can occur for low frequency words (Gardiner et al., 1988), though this result may be dependent on that word being highly familiar to the subject (Nairne & Widner, 1988). It has also been shown that if the generated item is not a conceptual unit, the generation effect does not occur (Gardiner & Hampton, 1985). Thus, a generation effect has not been found for either unfamiliar, low frequency words (Nairne & Widner, 1988) or for nonunitized paired items (Gardiner & Hampton, 1985). However, it should be noted that neither of these studies equated study and test conditions. For example, it is likely that unfamiliar, low frequency words have never been seen in the same printed form as presented at test. It is possible that if encoding and testing conditions are made equivalent, a generation effect would occur for these stimuli, as it has with nonwords.

III. Data-driven versus Conceptually Driven Processing

The concept of implicit and explicit memory processes and tasks has also been applied to the generation effect in an effort to show that generation involves primarily conceptual processing. Implicit memory is assumed to be tapped by memory tests that do not require deliberate or conscious recollection of an event or experience, such as lexical decision or word-fragment completion tasks. In contrast, explicit memory requires conscious recollection and is generally tested with measures such as recall and recognition tests. Some have argued that differential performance on these two types of tests reflects the use of two distinct memory systems (e.g., Cohen, 1984; Squire, 1982; Tulving, 1985). A contrasting view is that differences in performance are the result of the different processes required to complete the task (e.g., Jacoby, 1983; Roediger & Blaxton, 1987; Weldon & Roediger, 1987). More specifically, performance observed in implicit memory tests would be mediated by data-driven or surface processing, and in explicit memory, by conceptually driven or semantic processing. This processing, according to their

view, would occur in a unitary memory system, and operate on a continuum depending on the task, rather than as a dichotomy.

As we have seen, the generation effect has been generally studied with the use of explicit memory tests such as recall and recognition. In contrast, when tests of implicit memory are employed, read advantages, or "negative generation effects", have been found (Jacoby, 1983). Jacoby (1983) had subjects verbally read words both out of context and in context, and generate words from the appropriate context. He found that reading the words enhanced performance on a later perceptual identification task, a test of implicit memory; whereas generating the words resulted in superior performance on a recognition test and inferior performance on the perceptual identification task. This finding was interpreted as evidence that the generate task enhances processes leading to superior retention of explicit, and thus conceptual, information about an item, and to inferior retention of surface information.

The findings of the studies described above imply that the advantage of generating is due to processes occurring in explicit and thus conceptual memory. Additionally, Slamecka and Fevreiski (1983) showed that unsuccessful attempts to generate words still produced generation effects. They interpreted this finding as evidence that surface information about the generated item is not necessary, and in contrast semantic information is the necessary information for the generation effect to occur.

However, contrasting evidence has indicated that given the right conditions the generation effect can occur in the absence of semantic information and with implicit tests of memory. As discussed above, Nairne and Widner (1987) found a generation effect when the subject's task was to recognize a surface feature of nonwords. This shows that a generation effect can occur in the absence of semantic information, and moreover with only surface information. Additionally, Glisky and Rabinowitz (1985) found a generation effect in recognition following a word-fragment completion task. Likewise, Slameka and Graf (1978) found a

consistently reliable generation effect with only a rhyming rule relating the stimulus and response items. These findings show that a generation effect can occur in the absence of strong semantic cues. Moreover, Gardiner (1988) showed that performance on a word-fragment completion test is enhanced when subjects previously generate rather than read the words -- provided the same word fragments are used at test as at study. In Jacoby's (1983) study, the items presented in the generate condition were never seen; they were generated verbally. Thus, perceptual identification of the generated words would be expected to be worse than in the other conditions, as it was the only condition in which the word to be identified had never been seen by the subjects.

Gardiner (1989) proposed that the enhanced priming effects found by Gardiner (1988) may have been the result of conscious retrieval strategies adopted by his subjects. The use of strategies would have been possible as subjects were explicitly informed at test that some fragments were those words from the study list, but that, regardless, they were to complete as many fragments as possible. If they recognized the fragments, they may have tried to recollect the words that they had generated. Gardiner (1989) modified the previous procedure such that subjects could not be aware that a memory test was occurring. The subjects first read aloud the generate and read clues and targets. An example of a clue and target in the generate condition is *It swings in a clock -E-D-L-M*. They were told that this initial task was not the primary focus, but rather that the experimenters were most interested in the subsequent task -- a series of tests of general knowledge which were comprised of questions from *Trivial Pursuit*. The latter phase of the experiment was followed by a test including the original word fragments mixed in with additional definition and anagram tasks. With this procedure, none of Gardiner's subjects reported being aware that some of the fragments in the test had occurred in the study list; nevertheless, the original finding by Gardiner (1988) of a generation effect in unconscious memory was replicated.

Gardiner's (1988, 1989) findings can be interpreted as evidence that under the right conditions, that is, with test appropriateness, generation effects could be reflective of data-driven processing, rather than conceptual processing. However, Gardiner used a different implicit measure (i.e., word-fragment completion) than did Jacoby (1983; i.e., word identification). Schwartz (1989) found varying performance on different indirect tests. Whereas, like Jacoby (1983), Schwartz found that word indentification performance was lower for words generated rather than read, she also found that word completion performance for generated words was equivalent to that for words that were read.[3] She concluded that the two indirect measures tapped varying degrees of conceptual information, with word completion tasks requiring the greatest amount of conceptual processing. Thus, perhaps a better method of investigating this issue would be to replicate Jacoby's (1983) methodology and use of the word identification task as an implicit measure of retention, but provide the necessary feedback in the generate condition, giving the opportunity for the generated words to be seen at least once before testing.

In summary, the findings of read advantages in data-driven implicit tests by Jacoby (1983) are contrary to Gardiner's (1988, 1989) findings of generation advantages using the same types of implicit tests of memory. However, as Gardiner (1988, 1989) pointed out, the difference in the results of these studies can be simply explained in terms of the encoding specificity principle or the principle of test appropriateness. When Gardiner matched the fragments used at study and at test, the generation effect on an implicit measure was found. However, because of the different implicit measures of retention used in Jacoby's (1983) and Gardiner's (1988, 1989) studies, more empirical evidence is necessary to resolve this issue.

---

3 Schwartz (1989), like Jacoby (1983), did not ensure test appropriateness. That is, items presented in the generate condition were never *seen* by the subjects. Given these conditions, a generation effect would not be expected -- especially for a data-driven or implicit retention test.

IV. Conclusions

We must conclude from this conglomeration of research that semantic activation of the generated item is not sufficient, nor is it necessary for a generation effect to occur. There have been two major lines of research directed at determining the extent of the role of semantic processing in the generation effect.

One effort involved the manipulation of the meaningfulness of the generated item. Initial findings indicated that when the items were not meaningful, such as nonwords (e.g., McElroy & Slamecka, 1982), anomalous sentences (Graf, 1980), or meaningless bigrams (Gardiner & Hampton, 1985), the generation effect did not occur. However, further exploration of the issue has indicated that as long as the subject is tested on the same items as presented at test, the generation effect occurs regardless of the nature of the generated item. This principle has been known as the principle of test appropriateness (Nairne & Widner, 1987; Rabinowitz & Craik, 1986), the encoding specificity principle (Tulving, 1983), or the concept of transfer-appropriate processing (Morris et al., 1977).

Another line of research was directed at the distinction between implicit and explicit memory, and the supposition that the generation effect was solely the enhancement of explicit memory and thus involved only conceptual processing. Initial findings supported this view (e.g., Jacoby, 1983). However, the principle of encoding specificity overruled again; when Gardiner (1988, 1989) equated the conditions of study and test in an implicit paradigm, the generation effect reappeared.

Where do these findings leave us? The robustness of this phenomenon is clearly not in doubt. It has been found for recognition measures (e.g., Crutcher & Healy, 1989; Gardiner & Rowley, 1984; Glisky & Rabinowitz, 1985; Graf, 1980; Jacoby, 1978; Slamecka & Graf; 1978), recall measures (e.g., Crutcher & Healy, 1989; Donaldson & Bass, 1980; Gardiner & Authurs, 1982; Gardiner &

Hampton, 1985; Gardiner & Rowley, 1984; Ghatala, 1983; Graf, 1980; Jacoby, 1978; McFarland, Frey, & Rhodes, 1980; McNamara & Healy, 1991; Slamecka & Graf, 1978), and confidence ratings (McElroy & Slamecka, 1982; Slamecka & Graf, 1978). It has been found both with single trial and multitrial learning (e.g., Gardiner et al., 1988; Gardiner & Hampton, 1985; Gardiner & Rowley, 1984; Graf, 1980; McNamara & Healy, 1991), and has been found to sustain retention intervals of up to a week (Crutcher & Healy, 1989; McNamara & Healy, 1991). What remains lacking is a parsimonious theory that is capable of accounting for what drives the generation effect.

Some researchers have hypothesized that the generation effect is due to the enhancement of the relationship between the cue and the target (e.g., Donaldson & Bass, 1980; Rabinowitz & Craik, 1986). Although this hypothesis is similar to theories emphasizing the importance of semantic activation or processing, it differs in that the critical factor is the nature of the cue-target relationship, not the nature of the target itself. Donaldson and Bass (1980) suggested that the act of generating resulted in a superior encoding of the cue-target relationship and that the underlying factor for the generation effect was that the subject perform a check on each generated target to ensure that the response adequately meets its semantic relation to the stimulus. They found that a read task which also required the evaluation of the goodness of the relationship resulted in a memorial advantage for the target items similar to that found for a generate task. This finding is similar to McNamara and Healy's (1991) finding that the retention of the read subjects who developed mnemonics when learning word-nonword pairs resembled that of the subjects who had generated the nonword. Some researchers have argued against the importance of the relationship between the cue and the target on the basis that no memorial advantage is found for the cues in the generate task (Slamecka & Graf, 1978). Other researchers, however, have found generation effects for cues (Greenwald & Johnson, 1989). On the other hand, Glisky and Rabinowitz (1985) found a generation effect when single words were generated from word fragments; this result cannot be easily explained in terms of

relational processing. These contradictory findings leave this issue unresolved, though it seems clear that relational factors are not sufficient to explain the generation effect.

Some studies have not obtained a generation effect with a between-subjects design (Begg & Snider, 1987; Hirshman & Bjork, 1988; McDaniel, Waddill, & Einstein, 1988; Nairne, Riegler, & Serra, 1991; Schmidt & Cherry, 1989; Slamecka & Katsaiti, 1987). Slamecka and Katsaiti (1987) proposed that in a mixed-list design (i.e., read vs. generate as a within-subject variable), subjects displace their rehearsal selectively from read items to generated items. In a between-list design, such selective rehearsal cannot take place. Watkins and Sechler (1988) argued that this explanation is not possible due to their finding of a generation effect under conditions of incidental memorization during which selective rehearsal would not have been possible given the paradigm that they employed. Begg and Snider (1987) found that the generation effect is absent for cued recall when unrelated cues are used at encoding and retrieval, and for recognition when read versus generate was a between-subjects variable. They concluded that there are inhibitory consequences of generating such that read items are only cursorily processed. However, Begg and Snider's (1987) findings are at odds with Slamecka and Graf's (1978) finding of a generation effect with a between-subjects design and recognition tests.

Because of the findings of a lack of a generation effect in between-subjects designs, and also due to the inconsistency of the findings regarding memory for the generated cues, some authors have suggested that the generation effect is due to more than one factor (e.g., Hirshman & Bjork, 1988; McDaniel et al., 1988) Hirshman and Bjork (1988) proposed that generating is superior to reading because it activates features of the response term in memory (i.e., lexical or semantic activation) and strengthens the stimulus-response relation in memory (i.e., enhanced cue-target relationship). Though Hirshman and Bjork did obtain a generation effect with a cued-recall test using a between-subjects design, they did not find a

generation effect on a free-recall test. They assume that free-recall measures are more sensitive to the activation of the response term and inhibited by the stimulus-response relation, and that cued-recall measures are facilitated by response activation, but more sensitive to the stimulus-response relation. McDaniel et al. (1988) proposed still a third factor -- that in some contexts, generation will interfer with whole-list processing. More specifically, good free recall depends on whole-list relational processing and the task of generating enhances stimulus-response relational processing thus interfering with whole-list processing. This hypothesis was based on the finding that a generation effect does occur in a between-subjects design for free recall when the words are structured by categories whereby the cue word is the name of the category. McDaniel et al.'s (1988) conclusions have been more recently supported by Nairne et al. (1991) who compared memory for item and order information in both read and generate tasks. They found that a generate task requiring word-fragment completion impaired the retention of serial order information, though it enhanced the memory for the individual items.

Hirshman and Bjork's (1988) and McDaniel et al.'s (1988) interpretations of the generation effect are dependent upon the lack of a generation effect in a between-subjects design with free recall used as the retention measure. There exist, however, two studies that found generation effects under those conditions (Gollub & Healy, 1987; McFarland et al., 1980, Experiment 2). Gollub and Healy had subjects generate sentences using target words; yoked subjects evaluated target word usage in the sentences generated by the first group. Free recall performance for the target words was vastly superior for the generate subjects than for the read subjects. McFarland et al. had subjects either decide if a given word fit a specified context or generate a word to fit the context. They found a generation effect using a between-subjects design with free recall as the dependent measure, though the effect was slightly reduced from the results using a within-subjects design. These findings weaken

the arguments and conclusions made by Hirshman and Bjork (1988) and McDaniel et al. (1988).

The theories described above predict that the critical factors for the generation effect are the nature of the generated target, the stimulus, or the relation between the two. These theories emphasize the nature of the task and the items or information within the task, rather than the cognitive processes used to complete the task. Another class of theories have emphasized the process of generation itself. One recurrent hypothesis of this type has been that the generation effect is due to increased amount of effort (e.g., Griffith, 1976; McFarland et al., 1980) or arousal (Jacoby, 1978). Indeed, many researchers have found what is called an *effort effect*, wherein the more "effort" that is expended during encoding, the better the subsequent recall of the encoded items (e.g., Eysenck & Eysenck, 1979; Griffith, 1976; Jacoby, Craik, & Begg, 1979; Kolers, 1973; 1975; Tyler, Hertel, McCallum, & Ellis, 1979; see also Mitchell & Hunt, 1989, for a review of the literature). Effort, in these studies is generally operationalized in terms of the difficulty of the task (e.g., Jacoby, 1978) or the amount cognitive processing resources required (e.g., Griffith, 1976). According to *effort hypotheses* of the generation effect, it is the increased effort associated with generating a response that results in superior performance on retention tasks. There are several problems associated with any hypothesis that the amount of effort used for a task has memorial consequences. The first and foremost concerns the difficulty of defining or operationalizing the construct of "effort" consistently. Once defined, a subsequent problem is the difficulty of isolating the amount of effort used for a task and keeping constant all other variables that can affect later recall. One of the most important confounding variable is the amount of time spent on a task. It is difficult, and perhaps impossible, to specify tasks that require equal amounts of time, and yet varying degrees of effort. A common solution to that is to require subjects to spend equal amounts of time on all tasks. That unfortunately leaves open the question of what cognitive processing is actually occurring during the excess time for the easier tasks. That is, while the subject

may spend the same amount of time completing both a hard and an easy task, there is no guarantee that after the easy task is completed the extra time is actually allotted to processing the information in the task (they may be thinking about what to have for dinner). However, despite the methodological and theoretical problems associated with the concept of effort, it remains intuitively appealing as a partial explanation of why the generation effect occurs. Though generation effects have been found for tasks which required virtually no effort at all (e.g., Glisky & Rabinowitz, 1985), in most cases the task of generating is simply more difficult (and thus more effortful) than the task of reading. Nevertheless, effort has not been isolated as the primary factor leading to the generation effect.

Another hypothesis, which like those that emphasize the amount of effort involved focuses on cognitive processes, is that the generation effect is dependent upon cognitive procedures being developed during the generate task (e.g., Crutcher & Healy, 1989; Glisky & Rabinowitz, 1985; Kolers & Roediger, 1984; McNamara & Healy, 1991), and moreover, that these same procedures must be reinstated at test. This theory was proposed on a general cognitive level by Kolers and Roediger (1984). Glisky and Rabinowitz (1985) found that the memorial benefits associated with generation were enhanced by the repetition of the crucial operations. Furthermore, Crutcher and Healy (1989) showed that the crucial factor in obtaining the generation effect for the answers to simple multiplication problems was that the subject perform the necessary mental operations to derive the answers. More recently, McNamara and Healy (1991) showed that when subject performed simple and difficult multiplication problems, a generation effect only occurred for the difficult problems. They proposed that in the case of the simple problems, subjects already had well entrenched procedures for the retrieval of the answers to the problems, and thus procedures would not be developed during training. On the other hand, procedures for solving the difficult problems, in contrast to the simple problems, could be developed during training. Thus, whether or not the subject could develop procedures during training was

determined to be the critical factor determining whether or not a generation effect would occur. As explained above in the discussion of the nonword studies, McNamara and Healy's second experiment showed that subjects who had generated the nonwords during training had significantly better long-term retention of the nonwords than did those who had simply copied the pairs. More important to this issue, they found that subjects in the read condition who had spontaneously utilized mnemonic strategies to learn and retrieve the nonwords were indistinguishable from the generate subjects at posttest and retention test. This showed that an important factor leading to the generation advantage is the development of procedures, in this case the development of mnemonic strategies linking the word-nonword pairs.

In conclusion, after over a decade of generation experiments several critical factors have emerged. The first and foremost seems to be that any conclusion regarding the cause of the generation effect must be based on studies that ensure that the tested items are identical to the encoded items. When test appropriateness is assured, the generation effect may occur regardless of the nature of the generated item. There is further evidence that another critical factor may be that cognitive procedures or operations be developed during the generation task. Moreover, it has been evidenced by the finding of similar memorial advantages for read subjects who developed mnemonic strategies (McNamara & Healy, 1991), that the development of these procedures may not be exclusive to the task of generation, they may also be mediated by individual differences in motivation. In this case, whether or not the information is generated may not be the critical factor leading to superior retention, but rather that cognitive operations be performed on the information, and that these same cognitive operations be reinstated at test.

Acknowledgements

I gratefully acknowledge Lyle Bourne, Charles Judd, Richard Olson, and Nancy Butler Songer for their thoughtful comments on earlier drafts of this document. I would like to especially thank Alice Healy for all of her help in the preparation of this document.

References

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

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

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

Cohen, N.J. (1984). Preserved learning capacity in amnesia: Evidence for multiple memory systems. In L. Squire & N. Butters (Eds.), *The neuropsychology of memory* (pp. 83-103). New York: Guilford Press.

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

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

Davies, G.M., Milne, J.E., & Glennie, B.J. (1973). On the significance of "double encoding" for the superior recall of pictures to names. *Quarterly Journal of Experimental Psychology, 25,* 413-423.

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

Dosher, B.A., & Russo, J.E. (1976). Memory for internally generated stimuli. *Journal of Experimental Psychology: Human Learning & Memory, 2,* 633-640.

Erdelyi, M.H., Buschke, H., & Finkelstein, S. (1977). Hypermnesia for Socratic stimuli: The growth of recall for an internally generated memory list abstracted from a series of riddles. *Memory & Cognition, 5,* 283-286.

Eysenck, M.W., & Eysenck, M.C. (1979). Processing depth, elaboration of encoding, memory store, and expended processing capacity. *Journal of Experimental Psychology: Human Learning & Memory, 5*, 472-484.

Gardiner, J.M. (1988). Generation and priming effects in word-fragment completion. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 14*, 495-501.

Gardiner, J.M. (1989). A generation effect in memory without awareness. *British Journal of Psychology, 80*, 163-168.

Gardiner, J.M., & Authurs, F.S. (1982). Encoding context and the generation effect in multitrial free-recall learning. *Canadian Journal of Psychology, 36*, 527-531.

Gardiner, J.M., Craik, F.I.M., & Bleasdale, F.A. (1973). Retrieval difficulty and subsequent recall. *Memory & Cognition, 1*, 213-216.

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

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

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

Ghatala, E. (1983). When does internal generation facilitate memory for sentences. *American Journal of Psychology, 96*, 78-83.

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

Gollub, D., & Healy, A.F. (1987). Word recall as a function of sentence generation and sentence context. *Bulletin of the Psychonomic Society, 25*, 359-360.

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

Greenwald, A.G., & Johnson, M.M.S. (1989). The generation effect extended: Memory enhancement for generation cues. *Memory & Cognition, 17,* 673-681.

Griffith, D. (1976). The attentional demands of mnemonic control processes. *Memory & Cognition, 4,* 103-108.

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

Hyde, T.S., & Jenkins, J.J. (1969). Differential effects of incidental tasks on the organization of recall of a list of highly associated words. *Journal of Experimental Psychology, 82,* 472-481.

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

Jacoby, L.L. (1983). Remembering the data: Analyzing Interactive processes in reading. *Journal of Verbal Learning & Verbal Behavior, 22,* 485-508.

Jacoby, L.L., Craik, F.I.M. & Begg, I. (1979). Effects of decision difficulty on recognition and recall. *Journal of Verbal Learning & Verbal Behavior, 18,* 586-600.

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

Johnson, M.K., Taylor, T.H., & Raye, C.L. (1977). Fact and fantasy: The effects of internally generated events on the apparent frequency of externally generated events. *Memory & Cognition, 5,* 116-122.

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,* 626-635.

Kintsch, W., & van Dijk, T.A. (1978). Towards a model of text comprehension and production. *Psychological Review, 85,* 363-394.

Kolers, P.A. (1973). Remembering operations. *Memory & Cognition, 1,* 347-355.

Kolers, P.A. (1975). Memorial consequences of automatized encoding. *Journal of Experimental Psychology: Human Learning & Memory, 1,* 689-701.

Kolers, P.A., & Roediger, H.L. (1984). Procedures of mind. *Journal of Verbal Learning & Verbal Behavior, 23,* 425-449.

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

Mandler, G. (1979). Organization & repetition: Organizational principles with special reference to rote learning. In L.G. Nilsson (Ed.), *Perspectives on Memory Research.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Martin, E. (1972). Stimulus encoding in learning and transfer. In A.W. Melton and E. Martin (Eds.), *Coding Processes in Human Memory*, Washington, D.C., Winston.

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

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

McFarland, C.E. Jr., Frey, T.J., & Rhodes, D.D. (1980). Retrieval of internally versus externally generated words in episodic memory. *Journal of Verbal Learning & Verbal Behavior, 19,* 210-225.

McNamara, D.S., & Healy, A.F. (1991, November). *A generation advantage for multiplication skill and nonword vocabulary acquisition.* Paper presented at the meeting of the Psychonomic Society, San Francisco, CA.

Mitchell, D.B., & Hunt, R.R. (1989). How much "effort" should be devoted to memory? *Memory & Cognition, 17,* 337-348.

Morris, C.D., Bransford, J.D., & Franks, J.J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behavior, 16,* 519-533.

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

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

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

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

Neal, M.M., & Healy, A.F. (1987, May). *The generation effect with nonwords: Implications for representation in memory.* Paper presented at the Rocky Mountain Psychological Association Convention. Albuquerque, NM.

Payne, D.G., Neely, J.H., & Burns, D.J. (1986). The generation effect: Further tests of the lexical activation hypothesis. *Memory & Cognition, 14,* 246-237.

Pelton, L.H. (1969). Mediational construction vs. mediational perception in paired-associate learning. *Psychonomic Science, 17,* 220-221.

Rabinowitz, J.C., & Craik, F.I.M (1986). Specific enhancement effects associated with word generation. *Journal of Memory & Language, 25*, 226-237.

Roediger, H.L., & Blaxton, T.A. (1987). Effects of varying modality, surface features, and retention interval on priming in word-fragment completion. *Memory & Cognition, 15*, 379-388.

Russo, J.E., & Wisher, R.A. (1976). Reprocessing as a recognition cue. *Memory & Cognition, 4*, 683-689.

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

Schwartz, B.L. (1989). Effects of generation on indirect measures of memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15*, 1119-1128.

Schwartz, M. (1971). Subject-generated versus experimenter-supplied mediators in paired-associate learning. *Journal of Experimental Psychology, 87*, 389-395.

Schwartz, M., & Walsh, M.F. (1974). Identical subject-generated and experimenter-supplied mediators in paired associate learning. *Journal of Experimental Psychology, 103*, 878-884.

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

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

Slamecka, N.J., & Katsaiti, L.T. (1987). The generation effect as an artifact of selective displaced rehearsal. *Journal of Memory & Language, 26*, 589-607.

Squire, L.R. (1982). The neuropsychology of human memory. *Annual Review of Neuroscience, 5*, 241-273.

Tulving, E. (1983). *Elements of Episodic Memory*. New York: Oxford University Press.

Tulving, E. (1985). How many memory systems are there? *American Psychologist, 40,* 385-398.

Tyler, S.W., Hertel, P.T., McCallum, M.C., & Ellis, H.C. (1979). Cognitive effort and memory. *Journal of Experimental Psychology: Human Learning & Memory, 5,* 607-617.

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

Weldon, W.A., & Roediger, H.L. (1987). Altering retrieval demands reverses the picture superiority effect. *Memory & Cognition, 15,* 269-280.