

EFFECTS OF TEST EXPECTANCY, WORD FREQUENCY, AND WORD
CONCRETENESS ON ENCODING WORKLOAD

by

Frank Anthony Bosco Jr.

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Psychology

MONTANA STATE UNIVERSITY
Bozeman, MT

April 2005

©COPYRIGHT

by

Frank Anthony Bosco Jr.

2005

All Rights Reserved

APPROVAL

of a thesis submitted by

Frank Anthony Bosco Jr.

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

Dr. Richard A. Block

March 25, 2005

Approved for the Department of Psychology

Dr. Wesley C. Lynch

March 25, 2005

Approved for the College of Graduate Studies

Dr. Bruce R. McLeod

March 25, 2005

STATEMENT OF PERMISSION TO USE

In presenting this paper in partial fulfillment of the requirements for a master's degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library.

I have indicated my intention to copyright this thesis by including a copyright notice page, copying is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for permission for extended quotation from or reproduction of this thesis in whole or in parts may be granted only by the copyright holder.

Frank Anthony Bosco Jr.

April 18, 2005

ACKNOWLEDGEMENTS

Special thanks to Dr. Richard A. Block, and Dr. Keith A. Hutchison.

TABLE OF CONTENTS

1. INTRODUCTION	1
Two independent study strategies?	2
Evidence for dual-process models of memory.....	5
Frequency, concreteness, and relational-distinctive processing	6
Use of divided attention in memory research	9
Secondary task cost measurement issues.....	11
Temporal processing.....	12
The present research and predictions.....	14
2. EXPERIMENT 1	177
Method	177
Participants	177
Materials.....	177
Design	178
Procedure.....	178
Results	20
Temporal production analyses	20
Temporal production means analyses	20
Temporal production coefficient of variation analysis	21
Recognition test performance.....	23
Discussion	255
3. EXPERIMENT 2	288
Method	299
Participants	29
Materials.....	29
Design	30
Procedure.....	31
Results	32
Temporal production analyses	32
Aggregate temporal production means analysis.....	32
Aggregate temporal production coefficient of variation analysis	33
Intentional temporal production means.....	34
Intentional temporal production coefficient of variation	34
Recognition test performance.....	36
Recall test performance	37

TABLE OF CONTENTS - CONTINUED

Discussion	389
4. GENERAL DISCUSSION	41
REFERENCES	451

LIST OF TABLES

Table	Page
1. Mean temporal production as a function of memory condition and item type.....	21
2. Mean temporal production coefficient of variation as a function of memory condition and item type.....	21
3. Recognition test data as a function of memory condition and item type.....	24
4. Mean temporal production as a function of memory condition, intentionality, and item type.....	32
5. Mean temporal production coefficient of variation as a function of memory condition, intentionality, and item type	35
6. Recognition test data as a function of intentionality and item type.....	37
6. Mean number of words recalled as a function of intentionality and item type.....	38

LIST OF FIGURES

Figure	Page
1. Mean temporal production coefficient of variation as a function of memory condition	22
2. Mean temporal production coefficient of variation as a function of memory condition and word concreteness	23
3. Mean temporal production coefficient of variation for intentional items as a function of memory condition, word frequency, and word concreteness	36

ABSTRACT

The effort associated with relational and distinctive memory encoding was evaluated in two experiments. To this end, a secondary task was added during encoding, and the interference on the secondary task was used to infer encoding effort. Memory test expectancy was varied between-participants, with incidental, expect recognition, and expect recall conditions in Experiment 1, and recognition and recall conditions in Experiment 2. In both experiments, all participants encoded words varying in written frequency and rated concreteness. In Experiment 1, all participants received a recognition test regardless of test expectancy, and in Experiment 2, all participants received a test congruent with their expectancy. The results of Experiment 1 indicated that secondary task performance was sensitive to the addition of intentional encoding, as well as word concreteness. The results of Experiment 2 replicated those of Experiment 1 in that the secondary task appeared to be sensitive to differences in recognition and recall encoding operations. Further, the results demonstrated an interaction of frequency and concreteness on encoding effort when participants expected recall, whereas only concreteness affected encoding effort when participants expected recognition. The results are discussed in terms of Anderson and Bower's (1972, 1974) Human Associative Memory (HAM) model

INTRODUCTION

People can store information in memory in different formats (see Paivio & Madigan, 1970), using different strategies (e.g., Marschark & Hunt, 1989; Marschark, Richman, Yullie, & Hunt, 1987). For instance, students knowing whether an upcoming exam will be a multiple choice or an essay test may use this information to adjust their study strategies. Essay tests, aside from their reputation of being relatively difficult (Green, 1981; Zeidner, 1987), require different study strategies rather than simply a greater amount of studying. Differential strategies across test expectancy are also shown in note-taking (e.g., Rickards & Freidman, 1978). Indeed, difficult essays may elicit a study strategy involving the organization of the material in the form of an outline. Clearly, performance on essay tests should be heightened by the presence of a retrievable and intact outline, from which one could abstract details at the time of test.

Marchark and Hunt (1993) evaluated two differentiable types of processing during list learning; distinctive processing, characterized as processing differences among items in a list, and relational processing, characterized as processing similarities among list items. However, Marchark and Hunt qualified distinctive processing by adding, “This is not to say that distinctive processing requires...attention to differences, but rather that encoding of the item *qua* item will result in the encoding of differences” (p. 423). Study strategies for multiple-choice tests presumably require less relational processing and operate primarily on a discriminability basis (i.e., ability to choose correct from incorrect responses). Thus, performance on multiple-choice tests is most likely

augmented by thinking about each item (e.g., facts, events) distinctively during study. One question of interest to researchers regards the relative influence of distinctive and relational encoding strategies in the two tests. In the present article, the effort (Kahneman, 1973) required to carry out these presumably different encoding strategies is inferred from performance decrements associated with a concurrent secondary task.

Two Independent Study Strategies?

Dual-process models of recall and recognition memory such as the HAM model (e.g., Anderson & Bower, 1972, 1974) assume that both a search process and a recognition process are involved in recall retrieval, whereas only a recognition process is required for recognition retrieval. Anderson and Bower (1974) suggested that the initiation of the search process requires the use of associational pathways that were tagged during study. In essence, participants search a set of tagged pathways during recall retrieval, and submit possible responses to a recognition process. Thus, according to the HAM model, participants must tag pathways (i.e., encode relations among items) during recall encoding, but not during recognition encoding. Based on these assumptions, recall performance is a function of relational processing (i.e., pathway tagging), whereas recognition performance is primarily driven by distinctive processing (i.e., node tagging). Although several researchers (e.g., Tversky, 1973) support the idea that study for recognition tests involves primarily distinctive processing, there remains

confusion as to how distinctive processing plays a role in combination with relational processing during encoding for recall tests.

Undoubtedly, memory performance is a function of both encoding and retrieval processes (Loftus, 1971), and several memory theories posit the similarity of the two processes, including the encoding specificity principle (Tulving, 1983) and the transfer-appropriate processing theory (Morris, Bransford, & Franks, 1997). Baddeley, Lewis, Eldridge, and Thompson (1984) explicitly tested this assumption using a divided attention (DA) procedure. The dual-task approach, based on capacity theories of attention (e.g., Kahneman, 1973), assumes that the performance decrements associated with a concurrent secondary task (e.g., pursuit rotor tracking) reveals the workload properties of the primary task (e.g., memory processes). However, several researchers (e.g., Baddeley et al., 1984; Craik, Govoni, Naveh-Benjamin, & Anderson, 1996) further reported that concurrent secondary task execution interfered bidirectionally with encoding processes, evidenced by performance decrements for both memory and secondary tasks.

One can infer from the above results that memory performance is relatively more reliant on encoding processes than on retrieval process. One can easily agree that memory retrieval is difficult, or impossible, if the episodes were not encoded or not encoded optimally. These results have important implications for the study of memory. First, early generate-recognize theories (e.g., Kintsch, 1968) suggested that differences in memory performance are attributable to the characteristics of the retrieval environment,

but not to changes in encoding. It is clear now, however, that differences in the encoding processes have large effects on subsequent memory performance.

Although many studies have used DA to evaluate memory processes, their foci have seemed to converge on the similarities and differences between encoding and retrieval processes. To the extent that memory performance is relatively more dependent on encoding than retrieval, a dual-task methodology with the goal of assessing various effects on encoding workload could prove useful. What is needed is an evaluation of encoding workload across variables known to affect processing strategies (e.g., different test expectancies).

The goal of the present paper is to evaluate the effects of distinctive and relational processing on memory-encoding workload. Of particular interest are the effects of test expectancy (i.e., recall or recognition) on encoding effort, and accompanying interactions with lexical variables. From these manipulations, many clear predictions can be made. According to Anderson and Bower's (1974) HAM model, the workload associated with recognition test expectancy should be affected by varying distinctive information because of the relative dominance of distinctive processing in such conditions. In contrast, the workload associated with recall expectancy should vary as a function of the relational processing information because such conditions instigate primarily relational processing. Although several researchers (Craik, Govoni, Naveh-Benjamin, & Anderson, 1996) have examined such effects with respect to memory accuracy, relatively little evidence exists regarding the workload associated with these presumably different encoding strategies.

Evidence for Dual-Process Models of Memory

Evidence in support of dual-process theories includes a detrimental effect of recognition test expectancy on recall performance (Balota & Neely, 1980), a benefit of list organization on recall performance but not recognition performance (Bruce & Fagan, 1970; Kintsch, 1968), and a detrimental effect of rehearsal on recall performance but not recognition performance (Glenberg, Smith, & Green, 1977; Maisto, DeWaard, & Miller, 1977). In explaining the latter effect, it is assumed that rote rehearsal may prevent relational strategies, yet have benign or even enhancing effects on item-specific processing. Indeed, Schwartz and Humphreys (1974) reported that recognition performance benefited from oral rehearsal. Thus, these examples all suggest that relatively more relational encoding is engaged during recall tests and that recognition encoding is dependent primarily on item-specific processing.

Tversky (1973) indicated that recognition performance was improved with distinctive processing study instructions, and recall performance was improved with instructions emphasizing relations between items. These data align nicely to both the demands of the memory test phase described by Anderson and Bower (1974), as well as participant self-reports regarding encoding strategies. In retrospect, participants are more likely to report using a relational strategy when expecting a recall test, and they are more likely to report using distinctive processing strategies when expecting a recognition test (e.g., Eagle & Leiter, 1964).

Balota and Neely (1980) and Neely and Balota (1981) conducted a series of experiments to examine the relative amounts of node- and pathway-tagging occurring in recall and recognition. Although Neely and Balota criticized the usage of the theoretical construct pathway tagging, their so-called *cumulative node-tagging* essentially represents a relational strategy in which related items (not their pathways) are tagged. In the Neely and Balota account, the expectancy of recall essentially increases cumulative node tagging, as evidenced by a high frequency word recognition benefit for those participants expecting a recall test, in contrast to those expecting a recognition test. The important point is an interactive effect of word frequency and test expectancy, demonstrating clear quantifiable differences in encoding processes.

Frequency, Concreteness, and Relational-Distinctive Processing

Balota and Neely (1980) manipulated word frequency on the basis that high frequency words activate other memory nodes (i.e., have an inherently greater relational processing capacity) than low frequency words (Clark & Burchett, 1994; Glanzer & Bowles, 1976). Their results indicated that recall expectancy improved both recognition and recall performance, but to a greater extent for high frequency words than low frequency words. Thus, Balota and Neely provided evidence that encoding under recall expectancy is relatively more relational in nature. In accordance with these experiments, earlier researchers (e.g., Bousfield et al., 1958; Cofer, Bruce, & Reicher, 1966) have reported a greater tendency to cluster high frequency words than low frequency words

during recall. This result is consistent with the view that high frequency words have more related memory nodes. Thus, word frequency appears to be a sensitive marker of relational processing. More specifically, Balota and Neely's (1980) data suggest that the probability of relational processing with unrelated high frequency words is greater than that for unrelated low frequency words.

Paivio's (1971) dual-coding theory suggests that concrete items (e.g., "apple") are stored with both lexical and image-based codes, whereas abstract items (e.g., "enigma") are stored in only the former format. Thus, by dint of image-availability, concrete items benefit from more distinctive processing. According to Marschark and Hunt (1989), "concrete words have a distinctive advantage over abstract words" (p. 711).

Accordingly, recognition performance should benefit from increased distinctive processing in recognition tests. Concreteness as an item variable may prove useful in assessing the effects of distinctive processing. Although one experiment (Griffith & Johnston, 1973) has examined the effects of item concreteness on encoding workload in a paired-associate learning task, no research known to the author has examined these effects within recognition and free-recall tests. Furthermore, paired-associate learning has been considered by some to act as an operational midpoint between recognition and recall (Griffith & Johnston, 1973), and thus may not be an appropriate comparison for the present purposes.

Marschark (1985) and Marschark and Hunt (1989) suggested that concrete items may also benefit from relational processing. In essence, participants may use relational visual strategies during study in order to form an item-incorporating scene for later

retrieval. Along these lines, Paivio (1976) suggested that concreteness effects should be relatively more robust in free recall tests. Further, research has shown a greater tendency to cluster concrete than abstract words during recall (Frincke, 1968). Taken together, these results suggest that item concreteness, but not word frequency, should affect recognition encoding. Furthermore, due to the possibility that both high frequency words and concrete words can benefit from a relatively greater relational processing availability, one should expect both word frequency (presumably relationally-driven) and word concreteness (presumably both distinctive and relationally-driven) to affect recall encoding. More precise predictions will follow further discussion.

Another issue that deserves mention is possible interactions with distinctive and relational processing. In other words, what strategy will a participant use in recall encoding when both distinctive and relational processing is likely? Presumably, this will occur with concrete high frequency words. Marchark and Hunt (1989) suggested that “concreteness effects generally do not occur in memory for sentences comprising connected prose because distinctive processing of individual sentences usually is neglected in favor of relational processing among sentences” (p. 711). Thus, to the extent that relational processing strategies can be attributed to encoding connected prose, it appears that participants use less distinctive processing in such situations. The idea that differentially-engaged encoding strategies (i.e., in which relational processing can override distinctive processing), has been suggested by several researchers (e.g., Hunt & Einstein, 1981; Marchark & Hunt, 1989). Paivio and Madigan (1970) reported an interaction of frequency and imagery on free recall in which the typical high frequency

recall advantage is only present in concrete nouns. This could be due to combined, but not necessarily additive, effects of relational and distinctive processing. These distinctions have important implications for the present study.

Use of Divided Attention in Memory Research

Maisto, DeWaard, and Miller (1977) used oral repetition as a source of interference (i.e., DA) designed to prevent certain encoding strategies. In this study, encoding processes were inferred from memory data following interference. Johnston et al. (1970) and Griffith (1976) suggested a different use of DA; rather than experimentally altering the encoding process with another task, concurrent secondary task performance decrements could be used to infer the workload required to carry out the encoding processes themselves. This method has at least one major advantage. Most important, in order to infer encoding strategies from memory performance, the interference approach must assume that no interactions occur between encoding processes and retrieval processes. Should these interactions occur, inferences from memory performance cannot be attributed only to encoding. Indeed, evidence suggests that encoding and retrieval processes do not operate in essentially the same manner (e.g., Baddeley et al., 1984; Naveh-Benjamin et al., 2000).

Monitoring encoding processes with DA could prove useful for evaluating distinctive and relational processing of items across memory test expectancies. This use of DA has been used previously in memory studies as such to evaluate workload

associated with encoding across levels of word frequency (Naveh-Benjamin et al., 1998; Naveh-Benjamin & Guez, 2000), concreteness (Johnston, 1973), test expectancy (Craik et al., 1996), and item spacing (Johnston & Uhl, 1975); however, no study has systematically evaluated any of the above in a multifactor design. To test whether encoding workload is associated with different types of processing and test expectancies, such a design is necessary.

Naveh-Benjamin et al.'s (1998) results suggest that low frequency, concrete words are associated with larger recall encoding workload than high frequency, concrete words. Griffith and Johnston (1973) reported that item concreteness had no effect on encoding workload in a paired-associate task. However, the latter authors give no mention as to the frequency of the words. One possible explanation drawn from Griffith and Johnston's results is that the presence of relational processing attenuates any distinctive processing, as suggested by Marchark and Hunt (1989). Care must be taken in interpreting these results, and in drawing conclusions from inferred workload. If high frequency words benefit recall due to a larger network of related nodes, then more relational processing should occur in such conditions. Two possible explanations for Naveh-Benjamin et al.'s results are that participants (a) allocate more workload toward the relational processing of low frequency words due to an easier relational high frequency process, causing increased low frequency secondary task costs during encoding, or (b) when encoding low frequency concrete words, strategies involve a combination of distinctive and relational processing, and the effect is simply an artifact of

item selection (i.e., a special case for HC words). No studies have tested item frequency effects on encoding workload with abstract items.

To the extent that secondary task costs during encoding reflect processing types, a basic prediction can be made. If word frequency and concreteness have interactive effects on memory performance, these variables should have complimentary interactive effects on secondary task costs during encoding. This possible interaction on secondary task costs remains unexplored.

Secondary Task Cost Measurement Issues

The literature that has examined the nature of encoding processes via secondary task costs is not immune to problematic measurement issues. Indeed, Baddeley et al. (1984) noted the need for caution in interpreting such effects based on reaction time measures alone. The choice of secondary task becomes important due to the possibility that under DA conditions, controlled memory processes may operate in a qualitatively different manner. Perez-Mata, Read, and Diges (2002) presented lists of related-concrete or related-abstract words to participants expecting a free recall test. Their results indicated an interaction between attention level (i.e., full attention or DA) and word concreteness on veridical recall. Under DA conditions, word concreteness effects otherwise present under FA conditions were attenuated. The effects of concreteness are thus especially revealing, and a secondary task's tendency toward qualitative encoding changes could be tested with regard to item concreteness.

Unfortunately, the majority of DA and memory research involves the use of only concrete high frequency nouns (Baddeley et al., 1984; Craik et al., 1996). The relationship between secondary task costs and both normal memory performance and different processing strategies in these studies becomes difficult. Further, the use of a less obtrusive secondary task cost measure of workload is desirable. A secondary task bearing evidence of unidirectional interference (i.e., in which only the secondary task is affected) represents the ideal workload measure. There appears to be a lack of evidence demonstrating such evidence for an ideal task used in the extant DA and memory literature.

Temporal Processing

In light of complications regarding qualitative encoding changes as a result of DA (e.g., Perez-Mata et al., 2002), and in searching for the most ideal secondary task, the present experiments utilized 2-s temporal productions. In essence, temporal productions require participants to indicate the perceived completion of a temporal interval. In the present experiments, participants were asked to press a key when they thought a given stimulus had been displayed for two seconds. Interpolated was the task of memorization, which should interfere with timing according to the attentional-gate model of time estimation (Zakay & Block, 1996). According to this model, timing is disrupted by a concurrent task in manner consistent with capacity theories of attention (e.g., Kahneman, 1973). Recall encoding is presumed to be more difficult (i.e., require greater attentional

resources) than recognition encoding (e.g., Craik et al., 1996). Thus, as encoding workload increases, timing performance should exhibit increasing interference (i.e., longer productions, more variable productions, or both). According to the attentional-gate model, as the difficulty of any nontemporal task increases, the participant devotes less attention to the passing of time. Time estimate accuracy is thus sacrificed, to some extent, in maintaining the memorization task goals.

Brown (1997) reviewed the temporal processing interference literature and examined timing interference in three experiments. Brown noted that “only one nontemporal task was affected by concurrent timing, whereas timing was seriously disrupted by all three concurrent nontemporal tasks” (p. 1136). Brown also reviewed 18 studies that measured both temporal and nontemporal task performance and noted that the majority (10) of them caused interference in timing only. This evidence suggests that temporal tasks are relatively less resource-demanding than other concurrently performed tasks, and may be an ideal secondary task.

Several researchers (e.g., Block & Zakay, 1997; Zakay & Block, 1996; Zakay & Shub, 1998) noted that time estimates are a reliable measure of attentional demands, sensitive to many types of primary tasks. Brown (1997) noted that timing interference has occurred in the predicted direction with either the addition of or across varying difficulty levels of, physiologically-oriented tasks, motor tasks, and several cognitive tasks including the Stroop task and mental arithmetic. The extant DA and memory literature has not yet exhibited the use of a secondary task empirically demonstrated to be

reliable and widely-applicable. Furthermore, none of the previous secondary tasks have been associated with a tendency toward unidirectional interference.

The Present Research and Predictions

The present experiments were designed to address several issues. First, to what extent are temporal productions an ideal secondary task? To this end, Experiment 1 varied the difficulty associated with three memory expectancies. It is predicted by the attentional-gate model (Zakay & Block, 1996) that timing interference during encoding should increase from an incidental-expectancy condition (i.e., where no memory instruction is given) to a recognition-expectancy condition, and be greatest in a recall-expectancy condition. This same pattern of results has been found previously by Craik et al. (1996) with a different secondary task.

A second goal is to examine the effects of distinctive and relational processing on encoding workload. If more relational processing occurs in recall encoding, then high frequency words should attract more processing than low frequency words. If more distinctive processing occurs in recognition encoding, then concrete words should attract relatively more processing than abstract words. By examining these effects on workload across memory condition, one can (a) evaluate the extent to which secondary task costs are sensitive to these different encoding strategies, and (b) evaluate the extent to which recognition and recall tests differentially give rise to relational and distinctive processing. Along these lines, several predictions can be made.

First, in terms of Anderson and Bower's (1974) generate-recognize theory, it is predicted that recognition encoding workload will be affected by item concreteness but not item frequency. Although item concreteness has been attributed a small role in relational processing, this does not affect the present prediction. Of practical importance is the predicted effect of item concreteness, and lack of an effect of item frequency in recognition encoding. Thus, even if concreteness affects both relational and distinctive processing, word frequency should not affect recognition encoding.

An important issue is that of the predicted interference directions. It seems logical that more processing should lead to greater interference, which would lead to increased encoding workload for concrete items in recognition tests. Likewise, it seems logical that the abstract items in the recognition condition, for example, should elicit greater secondary task costs due to the increased encoding difficulty driven by a lack of distinctive details. Metcalfe and Kornell (2003) suggested that list learning evokes a tendency to study more readily the items that are perceived as easier to remember. Presumably, participants selectively encode different words once it seems impossible to remember an entire list. Words that seem more difficult to remember may receive less than optimal encoding; to some extent, they may be ignored. For the present purposes, increases in secondary task costs will be predicted to have a positive relationship with increases in processing effort. In other words, concrete items should lead to greater secondary task costs during recognition encoding due to an increased likelihood, of distinctive processing. This prediction is thus aligned with Metcalfe and Kornell's suggestion.

Recall predictions are less clear, although some definitive ones can be made. First, the predicted recognition pattern should not occur in recall. In line with Anderson and Bower (1974), and Balota and Neely (1980), recall encoding requires strategies beyond those used in recognition encoding due to the relationally-driven nature of the retrieval task. Specifically, the prediction is made that high frequency items should create greater secondary task costs during recall encoding because there is a greater chance that participants will relate other words to them. However, because the opposite pattern was found by Naveh-Benjamin et al. (1998) and Naveh-Benjamin and Guez (2000), the prediction is made that this effect is the result of an interaction involving differential processing of concrete items, in which distinctive processing plays a role. Thus, the high frequency increase in secondary task costs is predicted in recall only for abstract items, for which less distinctive processing is present, and relational processing in its absence can be observed.

Another prediction regards the unobtrusive nature of concurrently-performed temporal productions. If interference on primary tasks is typically minimal (cf. Brown, 1997), then the attenuation of the concreteness effect in recall observed by Perez-Mata et al. (2002) should not occur here. More specifically, typical memory performance results are expected in all conditions.

EXPERIMENT 1

MethodParticipants

A total of 136 undergraduates from Montana State University participated in the experiment for partial fulfillment of an introductory course requirement. This study was approved by the Human Subjects Committee of Montana State University. Participants were required to have normal or corrected to normal vision, and were required to be native English speakers.

Materials

From a set of 224 words, two sets of 112 words equated in terms of word length and number of syllables were created. Each of the sets contained four subsets of 28 stimuli varying in word frequency and word concreteness. The low frequency and high frequency words used had mean written frequency counts of 4.5 per million and 180 per million, respectively (Kucera & Francis, 1967). Abstract and concrete words exhibited mean concreteness ratings of 376.0 and 572.7, respectively (Coltheart, 1981). All stimuli were five or six-letter monosyllabic or disyllabic words. The two sets of 112 words were counterbalanced across subjects for use as study items and recognition test lure items. During the practice phase, six-letter strings comprised of asterisks (e.g., *****) were presented. All letter strings were presented in the same size 18 Times New Roman font.

All stimuli were presented in black on a white background. The experimental protocol was designed with E-Prime (Schneider, Eschman, & Zuccolotto, 2002).

Design

The independent variable *memory condition* was manipulated between subjects. One-third of the participants were unaware of a future memory test, one-third expected a recognition test, and one-third expected a recall test. All participants made 2-s temporal productions on all study items. The within-subjects variables were word frequency and word concreteness, resulting in a $3 \times 2 \times 2$ mixed-model design. For all participants, the dependent variables were temporal productions for each word presentation and later recognition memory performance.

Procedure

Participants first practiced the temporal production task with the practice stimuli. The temporal production required a spacebar response indicating the perceived completion of a 2-s duration which began at stimulus onset: Participants were instructed to press a spacebar once they felt the stimulus had been on the screen for 2 s. A 1500-ms intertrial interval separated each trial. The temporal productions were monitored by the program during practice, and participants completed the practice phase by reaching an accuracy criterion. Participants with a mean production below 1.8 s or above 2.2 s received another set of practice trials until the accuracy criterion was met. The mean production was calculated in running sets of 10 practice trials, and participants were

unaware of the transition between practice sets. The maximum number of practice trials was 100 (10 sets of 10), after which the participant advanced to next phase of the experiment regardless of accuracy.

During the experimental phase, words were presented individually in sets of 28. The temporal production task and intertrial interval used during experimental trials recapitulated those used during practice. Following the completion of each set of 28 trials, the participant encountered a rest break which was terminated by the participant at will with the spacebar. Each word presentation was drawn randomly without replacement from the 112-word pool.

The randomly-assigned between-subjects variable *memory condition* was varied by experimenter instruction. One third of the participants were simply told to perform the temporal production task on each stimulus (incidental condition). One third were told to expect a recognition test in which their task would be to indicate whether or not they had seen the word previously among lures (recognition condition). The final third of subjects were told to expect a recall task in which they would be later required to report any of the words they could recall on paper (recall condition).

Regardless of instruction, all participants received the recognition test after the presentation of all 112 words. The recognition test contained 224 words, half of which were old (i.e., previously presented), and half of which were new (i.e., not previously presented). Each word was drawn at random from the 224-word pool. Participants were told to press the “1” key to indicate that the item was old, and to press the “2” key to indicate it was a new (i.e., lure) item. Rest breaks were given after each set of 56

recognition trials and were terminated in the same manner as during the experimental trials; the intertrial interval remained at 1500 ms.

Results

Temporal Production Analyses

All analyses were conducted on both the mean (M) and the coefficient of variation (CV) of the production. The coefficient of variation is defined as the standard deviation divided by the mean, and is calculated with each participant's mean and standard deviation for each condition. This measure provides an index of mean-controlled variability. Both increases in production mean and coefficient of variation are considered indicative of increased workload (Brown, 1997). The outlier criterion for productions was set to 2.5 SDs from each cell mean. Production outliers within each cell were not removed, but replaced with each cell's upper or lower outlier criterion, depending upon whether the production was unusually long or short, respectively. Only one participant, with mean coefficient of variation greater than a criterion set by 2.5 SDs from the cell mean, was removed from the analysis.

Temporal Production Means Analysis

Table 1 shows mean temporal production for all conditions. ANOVA results revealed no significant effect of memory condition on temporal production mean, $F < 1$. Only one effect approached significance in the means analysis; incidental-condition participants

Table 1
Mean Temporal Production (in Milliseconds) as a Function of Memory Condition and Item Type

Item type	Memory Condition					
	Incidental (45)		Recognition (46)		Recall (44)	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
HC	2387	64.81	2327	67.02	2345	67.88
HA	2413	63.34	2321	63.88	2362	72.99
LC	2383	66.83	2333	65.64	2360	72.85
LA	2410	66.78	2338	67.04	2366	79.25
<i>M</i>	2398	66.19	2330	65.90	2358	73.24

Note. H = high frequency; L = low frequency; C = concrete; A = abstract. Numbers in parentheses represent the number of participants per group.

gave slightly longer mean productions to abstract ($M = 2412$, $SE = 10.72$) than concrete ($M = 2385$, $SE = 10.72$) items, $F(1,44) = 4.00$, $p = .05$, $d = 0.42$.

Temporal Production Coefficient of Variation Analysis

Table 2 shows mean temporal production coefficient of variation for all conditions. An omnibus ANOVA revealed a main effect of memory condition on production CV, $F(2, 132) = 3.81$, $p = .03$. Tests of simple effects indicated that recall

Table 2
Mean Temporal Production CV as a Function of Memory Condition and Item Type

Item type	Memory Condition					
	Incidental (45)		Recognition (46)		Recall (44)	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
HC	.142	.006	.163	.007	.181	.010
HA	.151	.008	.159	.007	.169	.008
LC	.142	.006	.162	.007	.177	.010
LA	.147	.007	.161	.006	.167	.009
<i>M</i>	.146	.007	.161	.007	.173	.009

Note. H = high frequency; L = low frequency; C = concrete; A = abstract. Numbers in parentheses represent the number of participants per group.

condition participants ($M = 0.17$, $SE = 0.01$) gave more variable productions than the incidental group ($M = 0.15$, $SE = 0.01$), $t(77) = 2.6$, $p = .01$, $d = 0.55$, and the recognition condition ($M = 0.16$, $SE = 0.01$) gave marginally more variable productions than the incidental condition, $t(77) = 1.8$, $p = .08$, $d = 0.38$. The recognition and recall conditions did not differ, $t < 1$. Figure 2 shows mean production coefficient of variation as a function of memory condition.

A significant effect of concreteness was found in the recall condition, $F(1,43) = 6.56$, $p = .01$, $d = 0.55$, in which concrete words ($M = 0.18$, $SE = 0.004$) received more variable productions than abstract words ($M = 0.17$, $SE = 0.004$). This effect was marginally significant in the opposite direction in the incidental condition, $F(1, 44) = 5.4$, $p = .06$, $d = 0.50$, in which abstract words exhibited more variable productions than

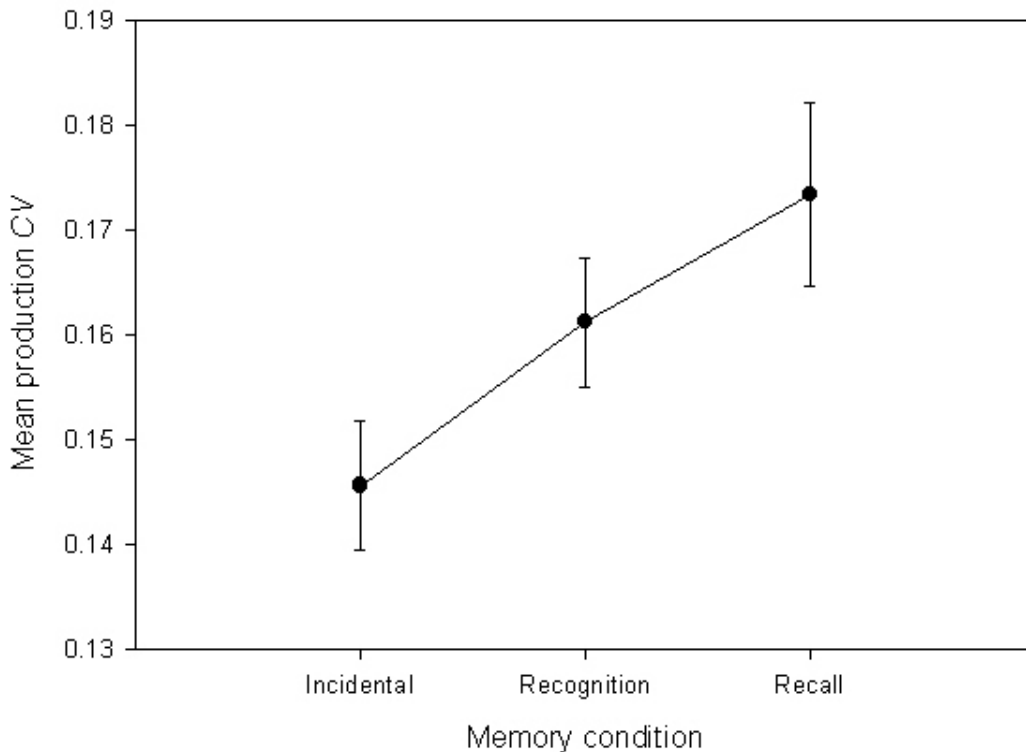


Figure 1. Mean temporal production CV as a function of Memory Condition. Each error bar is the standard error of the mean CV

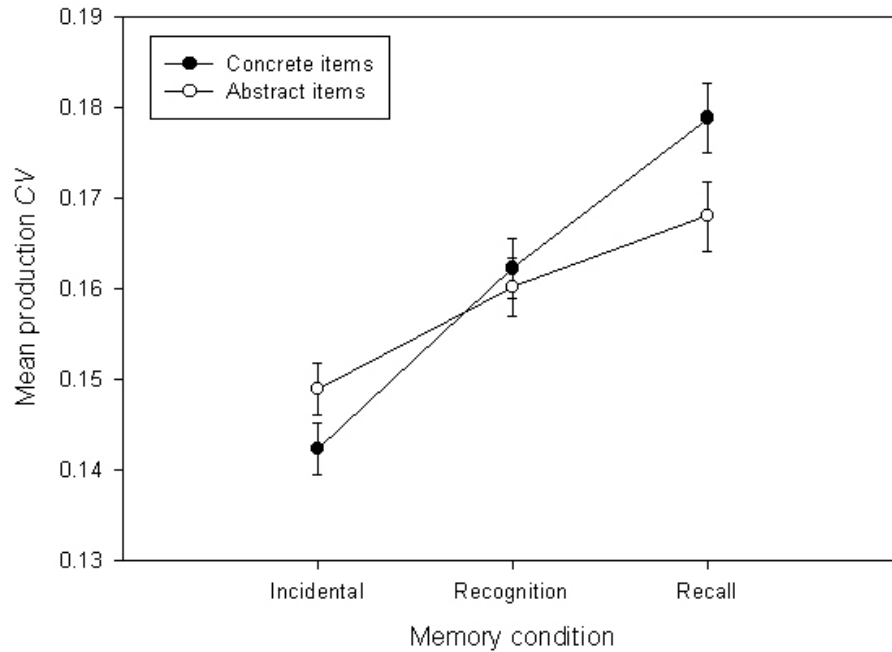


Figure 2. Mean production CV as a function of Memory Condition. Each error bar is the standard error of the mean CV.

concrete words. There was no effect of concreteness in the recognition condition, $F < 1$.

A Memory Condition \times Word Concreteness interaction on mean production coefficient of variation was also present, $F(2, 132) = 5.4$, $p < .01$, as shown in Figure 2. These results only partially support the present predictions; however, temporal production coefficient of variation does appear to be sensitive to varying levels of encoding difficulty.

Recognition Test Performance

Memory sensitivity was calculated as hits minus false alarms, shown in Table 3. ANOVA results revealed a significant main effect of memory condition on memory sensitivity, $F(2,132) = 15.68$, $p < .01$. Sensitivity increased from the incidental condition ($M = 0.26$, $SE = 0.01$) to the recognition condition ($M = 0.33$, $SE = 0.02$), and was

Table 3
Recognition Test Data as a Function of Memory Condition and Item Type

Item type	Memory condition								
	Incidental (45)			Recognition (46)			Recall (44)		
	<i>Hit</i>	<i>FA</i>	(<i>Hit-FA</i>)	<i>Hit</i>	<i>FA</i>	(<i>Hit-FA</i>)	<i>Hit</i>	<i>FA</i>	(<i>Hit-FA</i>)
HC	.573	.349	.223	.617	.344	.274	.679	.298	.381
HA	.581	.397	.185	.638	.390	.247	.647	.348	.300
LC	.566	.256	.310	.612	.239	.373	.696	.168	.529
LA	.531	.221	.310	.610	.198	.412	.641	.163	.478
<i>M</i>	.562	.306	.257	.619	.293	.327	.666	.244	.422

Note. H = high frequency; L = low frequency; C = concrete; A = abstract. Numbers in parentheses represent the number of participants per group. *Hit* = hit rate; *FA* = false alarm rate.

greatest in the recall condition ($M = 0.42$, $SE = 0.01$). This effect is consistent with the findings methodologically similar experiments in terms of test expectancy and type of test received (e.g., Balota & Neely, 1980).

A significant main effect of word frequency was revealed, $F(1, 132) = 165.73$, $p < .01$, $d = 1.57$, in which participants were better able to discriminate between old and new low frequency words than old and new high frequency words. The low frequency advantage in memory sensitivity is typical in the recognition memory literature (e.g., Gregg, 1976). Further analyses reveal that the low frequency advantage was present in the incidental condition, $F(1, 44) = 45.03$, $p < .01$, $d = 1.41$, in the recognition condition, $F(1, 45) = 64.99$, $p < .01$, $d = 1.68$, as well as in the recall condition, $F(1, 43) = 57.71$, $p < .01$, $d = 1.62$.

A main effect of word concreteness was also revealed $F(1, 132) = 7.30$, $p < .01$, $d = 0.33$; however, further analyses reveal that this effect was driven by the recall condition, in which concrete words were better remembered than abstract words $F(1, 43) = 12.39$, $p < .01$, $d = 0.75$. A Concreteness \times Memory Condition interaction was also

present $F(2, 132) = 4.64, p = .011$, demonstrating that those who expected the recall test had a greater high concreteness advantage than those who expected the recognition test, in which no such advantage was observed. Results also reveal a Frequency \times Concreteness interaction, $F(1, 132) = 6.42, p = .012$, in which concrete items led to better memory performance than abstract items for all memory conditions, but only for high frequency items.

Post hoc items analyses were conducted on the productions and recognition hit rate. Overall correlations were not significant; however, a significant correlation between the production coefficient of variation and later hit rate was revealed in the recognition condition, $r = -.167, p = .013$.

Discussion

The production analyses only partially confirm the predictions of attentional models of time estimation. Insofar as increased production coefficient of variation is indicative of an increase in workload, as the difficulty of test expectancy increased, the production coefficient of variation increased; however, production *Ms* did not show the same pattern. The marginal effects of concreteness on productions observed in the incidental condition suggest some degree of workload associated with lexical access. Participants not expecting a memory test exhibited marginally greater coefficients of variation, as well as means, for abstract compared to concrete words. This effect replicates a similar effect on temporal estimates found by Bosco (2002), and is assumed

to have occurred in the absence of a stimulus-specific response. In the incidental condition, participants were involved in making only temporal estimates, thus no memorization or stimulus-related response was involved. This effect highlights the importance of evaluating the possibility that lexical access can play a confounding role in DA and memory studies.

The presence of the concreteness effect in the incidental condition, although not statistically significant, suggests that workload is involved in processing abstract words in the absence of stimulus-related responses. Concreteness affects naming latencies and lexical decisions in the same direction (e.g., Groot, 1989). The concreteness effect in the recall condition suggests that either more distinctive or more relational processing had occurred, however; frequency did not affect recall condition coefficient of variation, so these data indicate that distinctive processing, but not relational processing, was used by participants in recall encoding.

These results stand in stark contrast to the current predictions. One explanation for unexpected effects in the recall condition is that another workload (i.e., that associated with the incidental condition) could have interfered with the workload associated with the item encoding. Another possibility for such an effect was the length of the test. Participants essentially prepared to recall as many words as they could from a 112-item list. This large number of instances is rather rare for word recall tests, and it could have qualitatively changed participant encoding strategies. Thus, the encoding workload inferred from the recall condition should be interpreted with caution.

The recognition test performance as a function of frequency, concreteness, and expectancy conforms to typical findings under FA conditions. Words of low frequency were recognized significantly better than those of high frequency in all conditions, and the high concreteness advantage in recognition performance was enhanced by the recall expectancy condition. Similarly, Paivio (1976) suggested that the effect of word concreteness is relatively more pronounced in recall in comparison to recognition memory performance.

The items analyses correlation between production coefficient of variation and recognition test hit rate is a speculative link. However, the link is strengthened by the fact that the correlation was restricted to a congruent condition in which a recognition test was expected and a recognition test was performed. No correlations existed between later recognition performance and productions for those in the incidental or recall conditions. According to previous research, encoding strategies appear to be a function of test expectancy (e.g., Tulving, 1983). Thus, insofar as encoding strategies relate to later memory performance, a correlation should be strongest in the recognition test expectancy condition. Of primary importance, this experiment demonstrated that temporal productions are a secondary task sensitive to encoding workload, and perhaps to qualitative differences in the encoding processes involved in recognition and recall.

EXPERIMENT 2

The primary purpose of Experiment 2 was to replicate and extend the findings of Experiment 1; to this end, several modifications were made. First, memory intentionality was manipulated within subjects. Second, in order to prevent the possibility that the secondary task costs associated with memory condition in Experiment 1 were attributable to confounded lexical access and encoding workloads (i.e., those evidenced by the effect of concreteness on production data in the incidental condition of Experiment 1), a lexical decision task was required prior to an intentionality instruction for each trial. Thus, only after a lexical decision for each item was the subject made aware of whether the item would be on a subsequent memory test. Analyses to follow represent productions made after the lexical decision (i.e., post lexical access), while the item and intentionality cue were both on the screen.

An ancillary purpose of Experiment 2 was to determine whether or not post-lexical access productions would be affected by word variables on incidental items. Such effects would be quite problematic; if incidental item presentations affect productions after a lexical decision, then the present measure of encoding workload becomes complicated.

One additional change to the procedure involved the number of stimuli recall-instructed participants attempted to remember. In Experiment 1, participants essentially expected to recall any of 112 words at the end of the presentation phase. It became

apparent that different encoding strategies could be present in this case, thus, participants expected to recall a mean of 10 items from each set of 20 presentations.

Method

Participants

A total of 149 undergraduates from Montana State University participated in the experiment for partial fulfillment of an introductory course requirement. Participation criteria recapitulated those of Experiment 1.

Materials

From a set of 256 words, two presentation sets of 128 words equated in terms of word length and number of syllables were created. Each of the sets contained four subsets of 32 stimuli varying in word frequency and concreteness. Mean frequency counts (Kucera & Francis, 1967) for low and high frequency words were 5.7 and 168.8 per million, respectively. Mean concreteness values (Coltheart, 1981) for abstract and concrete words were 342.1 and 557.2, respectively. Each participant received one of the two possible 128-item subs for study.

For all participants, one-half of the set was assigned as intentional (i.e., to-be-remembered) items, and one-half was assigned as incidental (i.e., not to-be-remembered) items. The assignment of stimulus set, as well as item intentionality, was randomized and counterbalanced across memory condition and subject. Intentionality was

manipulated with the presentation of a letter-cue. Above each test item appeared either the letter “I” (incidental) or the letter “R” (intentional). All letters appeared in 18 pt Courier New font.

The addition of the lexical decision task resulted in the addition of 32 nonwords held constant across all participants. Thus, each participant responded to 160 study-phase items, 128 words (64 intentional and 64 incidental) and 32 nonwords.

The memory test for those in the recognition condition included all 128 items, regardless of test expectancy condition, as well as 128 lures from the subset not received during study. This allowed the evaluation of the intent on subsequent memory performance. Nonwords were not tested in the recognition test, and all nonwords were designated incidental status during presentation.

In contrast to Experiment 1, all stimuli were presented in white on a black background. This was done to remedy the slightly fatiguing effects mentioned by some participants in Experiment 1.

Design

The independent variable memory condition was manipulated between subjects; half of the subjects participated in a recognition test, and half participated in a recall test. Unlike Experiment 1, test expectancy always matched the subsequent test. The independent variable intentionality was manipulated within subjects. One-half of the stimuli were items to appear on a future memory test, and one-half of the stimuli were items not to appear on a future memory test. The other within-subjects variables were

word frequency and word concreteness, resulting in a $2 \times 2 \times 2 \times 2$ mixed-model design. For all participants, the dependent variables were temporal productions for each word presentation and memory test performance.

Procedure

The procedure was very similar to that of Experiment 1, with a few necessary adjustments. Participants first practiced temporal productions using the same task from Experiment 1. The production task, intertrial interval, and rest-break protocol remained the same as those in Experiment 1. However, those in the recall condition recalled words on paper during rest breaks. These subjects were told to take a short break following each set recalled. Each presentation was drawn randomly from a 160-item pool. All participants studied 20-item lists between rest breaks.

Upon stimulus onset, participants were instructed to make a lexical decision with one of two keys, “Q” to indicate a word response and “P” to indicate a nonword response. Participants were given 1750 ms to make the lexical decision, after which the intentionality cue appeared regardless of the presence and accuracy of the response. Participants were instructed to begin making a 2-s temporal production upon intentionality cue onset. Participants were told that the “I” words would not be on the memory test, and that only their memory of the “R” words would be tested. The “I” and “R” intentionality cues were abbreviations for *ignore* and *remember*, respectively.

Results

Temporal Production Analyses

Production outliers were removed with the same method used in Experiment 1. Productions following only correct lexical decisions were analyzed. Incorrect lexical decisions, including non-responses, made up 5.3% of the responses. A total of 11 participants were removed from the analysis; the removal criterion was set to 2.5 *SDs* from each cell mean. Four participants were removed on the basis of high coefficient of variation, four were removed on the basis that their overall production mean was below 1 s, and three participants were removed because they made a large proportion of lexical decision errors.

Aggregate Production Means Analysis

Production means and standard errors for each condition are shown in Table 4.

Table 4
Mean Temporal Production (in Milliseconds) as a Function of Memory Condition, Intentionality, and Item Type

Item type	Memory Condition and Intentionality							
	Recognition				Recall			
	Intentional		Incidental		Intentional		Incidental	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
HC	2115	55.05	2092	58.87	1946	73.97	1900	72.05
HA	2151	57.79	2112	57.00	1942	70.56	1872	72.46
LC	2142	53.72	2109	53.34	1977	71.68	1897	70.40
LA	2174	54.16	2139	55.07	1985	72.08	1906	74.29
<i>M</i>	2146	55.18	2113	56.07	1962	72.07	1894	72.30

Note. H = high frequency; L = low frequency; C = concrete; A = abstract. Numbers in parentheses represent the number of participants per group. *Hit* = hit rate; *FA* = false alarm rate.

ANOVA results indicated a main effect in which production mean was greater for intentional items than for incidental items, $F(1, 136) = 22.49, p < .01, d = 0.57$.

Intentional items ($M = 2145, SE = 53.80$) were given longer productions than incidental items ($M = 2113, SE = 54.60$) in the recognition condition, $F(1, 67) = 6.46, p = .01, d = 0.44$. Likewise, intentional items ($M = 1962, SE = 70.90$) were given longer productions than incidental items ($M = 1894, SE = 71.20$) in the recall condition, $F(1, 69) = 16.40, p < .01, d = 0.68$.

A main effect of word frequency was also found, $F(1,136) = 7.54, p < .01, d = 0.33$, in which words of high frequency were given slightly longer productions than those of low frequency. Memory condition significantly affected productions in that participants the recognition condition produced longer durations than those in the recall condition $F(1,136) = 4.19, p = .04, d = 0.25$. No other main effects or interactions were significant.

Aggregate Production Coefficient of Variation Analysis

Production coefficients of variation for all conditions are shown in Table 5.

ANOVA results revealed that participants in the recall condition made more variable productions than those in the recognition condition, $F(1,136) = 16.28, p < .01, d = 0.48$.

A significant effect of intentionality on production *CV* was also found $F(1,136) = 15.15, p < .01, d = 0.47$. In contrast to the means analysis, intentional items were given less variable productions than the incidental items. A significant three-way interaction of Intentionality \times Frequency \times Concreteness was also found, $F(1, 136) = 4.64, p = .036$.

However, no effects of frequency, concreteness, or interactions occurred within the incidental items. Thus, analyses conducted on only intentional items are considered further.

Intentional Production Means

Contrary to predictions, ANOVA results revealed longer temporal productions for the recognition condition than for the recall condition, $F(1,136) = 4.19, p = .043, d = 0.25$. Low frequency words were given slightly longer productions than high frequency words, $F(1,136) = 4.02, p < .05, d = 0.24$; this effect did not interact with memory condition. Further analyses revealed the frequency effect was limited to the recall condition, $F(1,69) = 4.58, p = .04, d = 0.36$, in which low frequency words ($M = 1981, SE = 71.7$) were given longer productions than high frequency words ($M = 1944, SE = 71.2$). Concreteness affected productions in only the recognition condition, $F(1,67) = 4.86, p = .03, d = 0.48$, in which abstract items ($M = 2129, SE = 53.7$) were given longer productions than concrete items ($M = 1944, SE = 71.2$).

Intentional Production Coefficient of Variation

In support of the attentional-gate model (Zakay & Block, 1996), the present results reveal greater CVs for the intentional items ($M = .2405, SE = .01$) in the recall condition than for those in the recognition condition ($M = .1842, SE = .01$), $F(1, 136) = 18.13, p < .01, d = 0.51$. An effect of word concreteness was also found, $F(1,136) = 4.51,$

$p = .04$, $d = 0.25$, in which concrete items ($M = .2435$, $SE = .01$) items showed slightly greater CVs than abstract items ($M = .2375$, $SE = .01$).

As predicted, a Frequency \times Concreteness interaction was revealed, $F(1,136) = 4.50$, $p = .04$. However, further analyses revealed that this was limited to a main effect of concreteness in the recognition condition, $F(1,67) = 4.71$, $p = .03$, $d = 0.37$, and an interaction of Frequency \times Concreteness in only the recall condition $F(1,69) = 4.0$, $p < .05$. Within the recall condition, additional planned comparisons revealed a concreteness effect in low frequency words, $t(69) = 2.06$, $p = .04$, $d = 0.36$, but not in high frequency words $t(69) = .79$, $p = .43$, $d = 0.36$, and an effect of frequency in abstract words $t(69) = 2.04$, $p < .05$, $d = 0.27$, but not concrete words $t(69) = .80$, $p = .43$, $d = 0.36$. Figure 3 shows mean intentional item production coefficient of variation as a function of memory condition and item type.

Table 5
Mean Temporal Production CV as a Function of Memory Condition, Intentionality, and Item Type

Item type	Memory Condition and Intentionality							
	Recognition				Recall			
	Intentional		Incidental		Intentional		Incidental	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
HC	0.189	0.011	0.205	0.012	0.239	0.013	0.256	0.02
HA	0.183	0.008	0.196	0.010	0.247	0.013	0.257	0.01
LC	0.193	0.010	0.206	0.010	0.248	0.012	0.240	0.01
LA	0.173	0.008	0.197	0.010	0.228	0.012	0.259	0.01
<i>M</i>	0.184	0.009	0.201	0.010	0.241	0.012	0.253	0.01

Note. H = high frequency; L = low frequency; C = concrete; A = abstract. Numbers in parentheses represent the number of participants per group. *Hit* = hit rate; *FA* = false alarm rate.

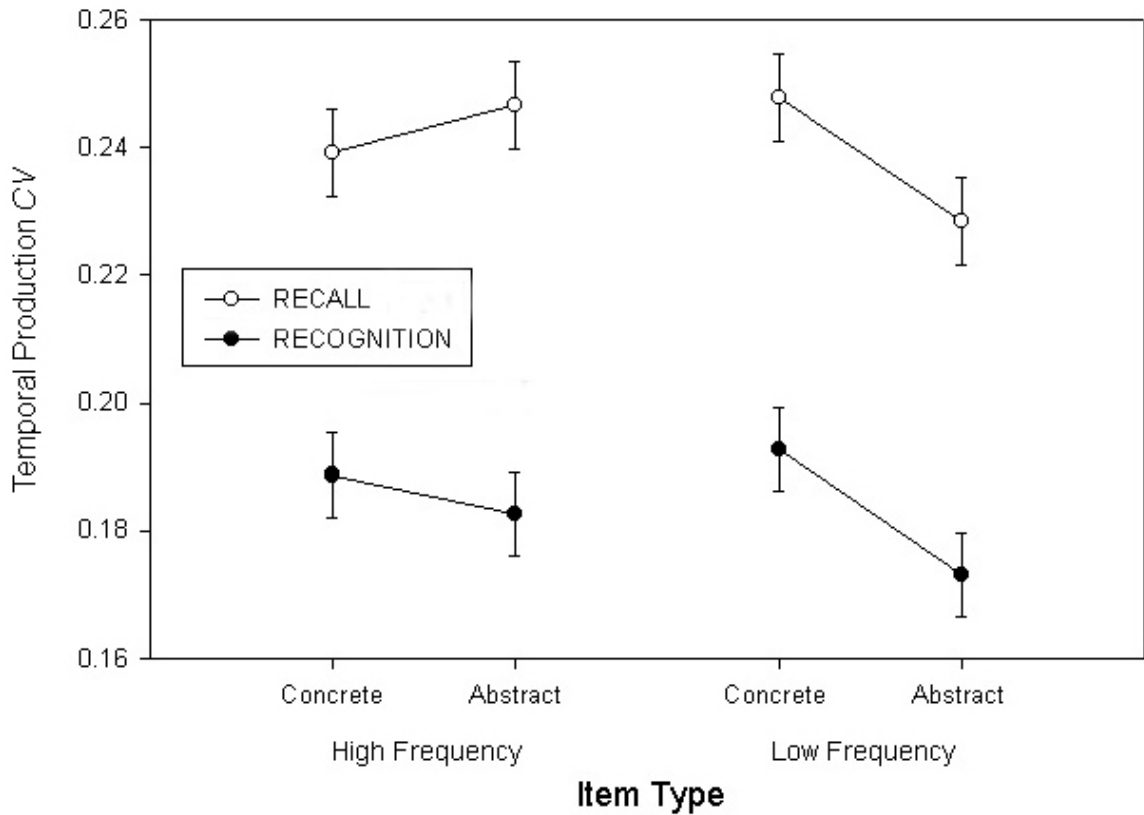


Figure 3. Mean temporal production *CV* for intentional items as a function of Memory Condition, word frequency, and word concreteness. Each error bar is the standard error of the mean.

Recognition Test Performance

Recognition test performance is shown in Table 6. As in experiment 1, intent to remember the words increased overall recognition performance $F(1, 76) = 4.88, p = .03, d = 0.36$. The typical low frequency advantage was present, $F(1, 76) = 373.53, p < .01, d = 3.11$. The effect of concreteness approached significance, $F(1, 76) = 3.28, p = .08, d = 0.29$; however, the Frequency \times Concreteness interaction was also significant $F(1, 76) = 8.12, p < .01$. Analyses conducted separately for intentional items show effects for both frequency, $F(1, 76) = 283.96, p < .01, d = 2.72$, and concreteness, $F(1, 76) = 7.70, p <$

Table 6
Recognition Test Data as a Function of Intentionality and Item Type

Item type	Intentionality					
	Incidental			Intentional		
	<i>Hit</i>	<i>FA</i>	<i>(Hit-FA)</i>	<i>Hit</i>	<i>FA</i>	<i>(Hit-FA)</i>
HC	.684	.314	.370	.728	.314	.414
HA	.709	.382	.327	.715	.382	.333
LC	.767	.196	.571	.790	.196	.594
LA	.767	.185	.581	.784	.185	.599
<i>M</i>	.731	.269	.462	.754	.269	.485

Note. H = high frequency; L = low frequency; C = concrete; A = abstract.
Hit = hit rate; *FA* = false alarm rate.

.01, $d = 0.45$. The Frequency \times Concreteness interaction was also significant, $F(1, 76) = 10.54$, $p < .01$, showing a relatively greater concreteness benefit in high frequency words. A separate analysis of only the incidental items revealed a frequency effect, $F(1, 76) = 32.23$, $p < .01$, $d = 0.91$, as well as a concreteness effect, $F(1, 76) = 8.56$, $p < .01$, $d = 0.47$, but no interaction.

Recall Test Performance

Recall test performance is shown in Table 7. Like the recognition data, intentionality had a significant effect on the mean number of words recalled $F(1,72) = 1196.99$, $p < .01$, $d = 5.73$. The aggregate ANOVA revealed significant main effects of both frequency, $F(1,72) = 5.26$, $p = .03$, $d = 0.38$, and concreteness, $F(1,72) = 74.03$, $p < .01$, $d = 1.42$. Further, three interactions were present: Intentionality \times Concreteness, $F(1,72) = 44.33$, $p < .01$, Frequency \times Concreteness, $F(1,72) = 15.82$, $p < .01$, and Intentionality \times Frequency \times Concreteness, $F(1,72) = 5.55$, $p = .021$.

The intentional items analysis showed that item concreteness affected recall $F(1,72) = 68.81, p < .05, d = 1.37$ yet word frequency did not ($p > .1$). However, an interaction of Frequency \times Concreteness was present, $F(1,72) = 11.81, p < .01$ in which for concrete items, high frequency words were better recalled than low frequency words. Within abstract words, trends indicated better recall for low frequency words, although this difference was not significant, $F < 1$.

The present data are typical of recall data; Paivio and Madigan (1970) reported that imagery and frequency are empirically differentiable, and they also reported an interaction comparable to the present results. Paivio reported that high frequency words were better recalled than low frequency words, but only for concrete items; the frequency effect was reversed for abstract items. Thus, due to the typical recall effects produced in Experiment 2, there is no reason to assume that the secondary task interacted with word variables as in the findings of Perez-Mata et al. (2002).

Table 7
Mean Number of Words Recalled as a Function of Intentionality and Item Type

Item type	Intentionality			
	Incidental		Intentional	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
HC	0.81	0.09	10.43	0.21
HA	0.38	0.09	7.96	0.21
LC	0.47	0.09	9.40	0.21
LA	0.38	0.09	8.26	0.21
<i>M</i>	0.51	0.09	9.01	0.21

Note. H = high frequency; L = low frequency; C = concrete; A = abstract.
Hit = hit rate; *FA* = false alarm rate.

Discussion

These results show some similarities and differences to the results of Experiment 1. An unexpected result was that not-to-be-remembered items resulted in more variable time estimates than to-be-remembered items. This result is not consistent with models of time estimation based on capacity theories of attention. Furthermore, mean productions showed the opposite pattern in that to-be-remembered items received longer productions. Although the mean production analysis follows attentional model predictions, the existence of opposite effects in the mean and coefficient of variation analyses is problematic. Brown (1997) noted that as concurrent workload increases, production mean, production variability, or both should increase.

The above contrast in effects suggests one of two possible explanations. The first possibility rests upon the change of memory condition from a between-subjects design in Experiment 1 to a within-subjects design in Experiment 2. It is widely held that rehearsal is one of the main strategies in list memorization (Belleza, Geiselman, & Aronovsky, 1975); thus, it is possible that the intervals following the lexical decision in the incidental trials in Experiment 2 were used for the rehearsal of previous intentional items. If this is the case, one would expect more variable productions in the incidental condition because of a larger range of possible stimuli and their corresponding memorization attempts. However, this does not fully explain the difference observed between production mean and coefficient of variation.

A second consideration regards the attentional-gate model (Zakay & Block, 1996). If instances occur in which production mean and coefficient of variation vary in opposite directions significantly, the model requires a modification to consider both these measures and their relation to workload. With few exceptions (e.g., Brown, 1997), the temporal processing interference literature tends to report only one of these measures.

Aside from the unexpected results, the main effects of interest were those that affected productions after the lexical decision across levels of intentionality. The results indicate that no lexical variables affected the incidental items following lexical decision. As previously mentioned, this could be the result of an increase in rehearsal variability. In contrast, several item effects were shown under intentional conditions. This contrast suggests that temporal productions are sensitive to at least some form of encoding process occurring in the absence of workload associated with lexical access.

Most important to the present discussion, the production coefficient of variation analyses suggest that (a) recognition encoding is driven primarily by distinctive encoding, as evidenced by an effect of concreteness, but not of frequency; (b) both relational and distinctive processing are involved in recall, as evidenced by the interaction of frequency and concreteness; and (c) the intuitive increase in encoding workload due to low frequency word encoding (e.g., Naveh-Benjamin et al., 1998, 2000) is limited to concrete words.

GENERAL DISCUSSION

Two experiments examined the effects of presumably different encoding strategies on secondary task performance decrements. In Experiment 1, secondary task costs calculated with coefficient of variation increased as a function of increased encoding difficulty (i.e., across test-expectancy). The results from Experiment 2 replicate these findings with an increase in coefficient of variation from the intentional recognition condition to the intentional recall condition. These data thus replicate similar effects found by other researchers (e.g., Baddeley et al., 1984; Craik et al., 1996).

Three relatively definitive results from both experiments show that: (a) The use of temporal production coefficient of variation is an ideal secondary task cost measure of encoding workload; (b) Recall encoding requires relatively greater workload than recognition encoding; and (c) Memory performance following DA (i.e., that created by concurrently performing temporal productions) appears to follow relatively normal patterns of memory performance found under conditions of full attention.

In terms of the effects of lexical variables and test expectancy on secondary task costs, relatively more discussion is given to the results of Experiment 2, in contrast to Experiment 1, because of its more appropriate recall test conditions. In Experiment 1, participants in the expect recall condition attempted to encode 112 items, which may have qualitatively changed encoding strategies. Further, the addition of a lexical decision task prior to the production in Experiment 2 afforded a measure of post-lexical access encoding strategies.

The prediction that recognition encoding should invoke distinctive processing was confirmed in Experiment 2, but not in Experiment 1. During recognition encoding, item distinctiveness should play a role, but relationships between items should not. Thus, it was predicted that concreteness, but not frequency, should have an effect on recognition encoding workload. Further, it was suggested that the same pattern should not exist in recall conditions, but instead that word frequency should affect recall encoding. The results of Experiment 1 do not support these predictions; however, they do not disconfirm them either. The presence of an effect of concreteness in recall can be explained in terms of an increase in relational processing (see Marchark & Hunt, 1989).

Accordingly, others (Hunt & Einstein, 1981) suggested that relational processing can override distinctive processing when both are available. Because less relational processing is attributed to low frequency words, distinctive processing appeared to influence workload, as evidenced by increased production coefficient of variation for concrete, low frequency items. These data fit nicely with Anderson and Bower's (1974) generate-recognize theory of memory. As predicted, concreteness increased recognition encoding workload. Further, both frequency and concreteness affected recall encoding, suggesting an involvement of both distinctive and relational processing.

The predictions concerning frequency, concreteness, as well as test-expectancy were confirmed by the results of Experiment 2. The production coefficient of variation results suggest that recognition encoding is affected only by concreteness. Workload was greater during recognition encoding for concrete, compared to abstract, words. Concreteness and frequency interacted during recall encoding. For low frequency items,

concreteness increased encoding workload, and for high frequency items, concreteness had a nonsignificant effect in the opposite direction. This is presumably due to varying levels of relational processing utilization across memory test expectancy. Thus, when participants encounter a low frequency item in a recall study phase, a distinctive strategy may seem more useful than a relational one. When a participant encounters a high frequency item, primarily relational processing is used, and concreteness does not have an effect on encoding.

The present data thus suggest that participants expecting a recognition test primarily focus on distinctive characteristics of the stimuli, but those expecting a recall test can engage either relational or distinctive processing depending on the characteristics of the items. In other words, the greater secondary task costs associated with concrete items in recognition encoding suggest that participants are engaging in distinctive processing. The interaction of frequency and concreteness in recall encoding suggests that participants are engaging in relatively more relational processing. As predicted, high frequency words created greater secondary task costs for only abstract items. However, it cannot be inferred from these results that recognition test study strategies involve only distinctive processing, or that those associated with recall tests are purely relational in nature.

In explaining the effects of concreteness on recall encoding from Experiment 1, it is suggested that the length of the study list affected how participants engage distinctive processing strategies. Thus, the production coefficient of variation for items in the recall condition from Experiment 1 could have been greater for concrete words than abstract

words because participants engaged in more distinctive processing. However, this assertion would support that recognition encoding workload would have been affected by concreteness.

An especially unusual finding in Experiment 2 was the opposite pattern found in the means and coefficient of variation analyses across item intentionality. Other than explaining this effect as an anomaly, an alternative explanation is that participants rehearsed previous intentional items during the incidental-item productions. If the latter explanation is true, one would expect that no lexical variables would have affected productions in this condition. Thus, although the unusual production analyses may appear to be a limitation to the present conclusions, the experimental conditions which led to this unusual finding could prove useful for studies of DA and memory. If it is the case that incidental production durations are used to rehearse previous intentional items, then participants are less likely to rehearse previous intentional items during intentional encoding. This could afford greater accuracy to the measure of encoding workload across lexical variables for intentional conditions.

Although previous research has examined the effort associated with encoding across lexical variables (e.g., Naveh-Benjamin et al., 1998) and across test expectancy (e.g., Craik et al., 1996), a multifactor design is required to evaluate differences in relational and distinctive processing, or node tagging and pathway tagging. The present experiments thus add to the extant literature an evaluation of encoding effort across lexical variables in different testing situations.

REFERENCES

- Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, 79, 97-123.
- Anderson, J. R., & Bower, G. H. (1974). A propositional theory of recognition memory. *Memory & Cognition*, 2, 406-412.
- Baddeley, A., Lewis, V., Eldridge, M., & Thomson, N. (1984). Attention and retrieval from long-term memory. *Journal of Experimental Psychology: General*, 113, 518-540.
- Balota, D. A., & Neely, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 576-587.
- Bousfield, W. A., Cohen, B. H., & Whitmarsh, G. A. (1958). Associative clustering in the recall of words of different taxonomic frequencies of occurrence. *Psychological Reports*, 4, 39-44.
- Block, R. A., & Zakay, D. (1997). Prospective and retrospective duration judgments: A meta-analytic review. *Psychonomic Bulletin & Review*, 4, 184-197.
- Brown, S. W. (1997). Attentional resources in timing: Interference effects in concurrent temporal and nontemporal working memory tasks. *Perception & Psychophysics*, 59, 1118-1140.
- Bruce, D., & Fagan, R. L. (1970). More on the recognition and free recall of organized lists. *Journal of Experimental Psychology*, 85, 53-154.
- Clark, S. E., & Burchett, R. E. R. (1994). Word frequency and list composition effects in associative recognition and recall. *Memory & Cognition*, 22, 55-62.

- Cofer, C. N., Bruce, D. R., & Reicher, G. M. (1966). Clustering in free recall as a function of certain methodological variations. *Journal of Experimental Psychology*, *71*, 858-866.
- Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology*, *33*, 497-505.
- Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, *125*, 159-180.
- Eagle, M., & Leiter, E. (1964). Recall and recognition in intentional and incidental learning. *Journal of Experimental Psychology*, *68*, 58-63.
- Frincke, G. (1968). Word characteristics, associative-relatedness, and the free-recall of nouns. *Journal of Verbal Learning & Verbal Behavior*, *7*, 366-372.
- Glanzer, M., & Bowles, N. (1976). Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 21-31.
- Gregg, V. (1976). Word frequency, recognition and recall. In J. Brown (Ed.), *Recall and recognition* (pp. 103-129). London: Wiley.
- Glenberg, A., Smith, S. M., & Green, C. (1977). Type I rehearsal: Maintenance and more. *Journal of Verbal Learning & Verbal Behavior*, *16*, 339-352.
- Griffith, D. (1976). The attentional demands of mnemonic control processes. *Memory & Cognition*, *4*, 103-108.
- Griffith, D., & Johnston, W. A. (1973). An information-processing analysis of visual imagery. *Journal of Experimental Psychology*, *100*, 141-146.

- Green, K. E. (1981). Test anxiety level and test format preference. *Psychological Reports, 48*, 537-538.
- Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning & Verbal Behavior, 20*, 497-514.
- Johnston, W. A., Greenberg, S. N., Fisher, R. P., & Martin, D. W. (1970). Divided attention: A vehicle for monitoring memory processes. *Journal of Experimental Psychology, 83*, 164-171.
- Johnston, W. A., & Uhl, C. N. (1976). The contributions of encoding effort and variability to the spacing effect on free recall. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 153-160.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.
- Kintsch, W. (1968). Recognition and free recall of organized lists. *Journal of Experimental Psychology, 78*, 481-487.
- Loftus, G. R. (1971). Comparison of recognition and recall in a continuous memory task. *Journal of Experimental Psychology, 91*, 220-226.
- Maisto, S. A., DeWaard, R. J., & Miller, M. E. (1977). Encoding processes for recall and recognition: The effect of instructions and auxiliary task performance. *Bulletin of the Psychonomic Society, 9*, 127-130.
- Marschark, M. (1985). Imagery and organization in the recall of prose. *Journal of Memory & Language, 24*, 734-745.
- Marschark, M., & Hunt, R. R. (1989). A reexamination of the role of imagery in learning and memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 710-720.

- Marschark, M., Richman, C. L., Yuille, J. C., & Hunt, R. R. (1987). The role of imagery in memory: On shared and distinctive information. *Psychological Bulletin, 102*, 28-41.
- Metcalfe, J., & Kornell, N. (2003). The dynamics of learning and allocation of study time to a region of proximal learning. *Journal of Experimental Psychology: General, 132*, 530-542.
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 519-533.
- Naveh-Benjamin, M., Craik, F. I. M., Gavrilesco, D., & Anderson, N. D. (2000). Asymmetry between encoding and retrieval processes: Evidence from divided attention and a calibration analysis. *Memory & Cognition, 28*, 965-976.
- Naveh-Benjamin, M., Craik, F. I. M., Guez, J., & Dori, H. (1998). Effects of divided attention on encoding and retrieval processes in human memory: Further support for an asymmetry. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1091-1104.
- Naveh-Benjamin, M., & Guez, J. (2000). Effects of divided attention on encoding and retrieval processes: Assessment of attentional costs and a componential analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1461-1482.
- Neely, J. H., & Balota, D. A. (1981). Test-expectancy and semantic-organization effects in recall and recognition. *Memory & Cognition, 9*, 283-300.
- Paivio, A., & Madigan, S. A. (1970). Noun imagery and frequency in paired-associate and free-recall learning. *Canadian Journal of Psychology, 24*, 353-361.
- Paivio, A. (1971). *Imagery and verbal processes*. Oxford: Rinehart & Winston.

- Paivio, A. (1976). Imagery in recall and recognition. In J. Brown (Ed.), *Recall and recognition* (pp. 103-129). London: Wiley.
- Pérez-Mata, M. N., Read, J. D., & Diges, M., (2002). Effects of divided attention and word concreteness on correct recall and false memory reports. *Memory, 10*, 161-177.
- Rickards, J.P., & Friedman, F. (1978). The encoding versus the external storage hypothesis in note taking. *Contemporary Educational Psychology, 3*, 136-143.
- Schwartz, R. M., & Humphreys, M. S. (1974). Recognition and recall as a function of instructional manipulations of organization. *Journal of Experimental Psychology, 102*, 517-519.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime Reference Guide*. Pittsburgh: Psychology Software Tools.
- Tulving, E. (1983). *Elements of episodic memory*. New York: Clarendon Press.
- Tversky, B. (1973). Encoding processes in recognition and recall. *Cognitive Psychology, 5*, 275-287.
- Zakay, D., & Block, R.A. (1996). The role of attention in time estimation processes. In M. A. Pastor, & J. Artieda (Eds.), *Time, internal clocks and movement* (pp. 143-164). Amsterdam: Elsevier.
- Zeidner, M. (1987). Essay versus multiple-choice type classroom exams: The student's perspective. *Journal of Educational Research, 80*, 352-358.