

Positive and negative generation effects, hypermnesia, and total recall time

NEIL W. MULLIGAN and MARQUINN D. DUKE
Southern Methodist University, Dallas, Texas

Self-generated information is typically remembered better than perceived information (the generation effect). Experimental design produces an important limiting condition for this effect: Generation enhances recall in within-subjects designs, but typically not in between-subjects designs. However, Mulligan (2001) found that the generation effect emerged over repeated recall tests in a between-subjects design, calling into question the generality of this limiting condition. Two experiments further delineated the emergent generation effect. Experiment 1 demonstrated that this effect does not require multiple discrete recall tests but may emerge on a single recall test of long duration. Experiment 2 demonstrated that the negative generation effect (a reversal of the typical generation effect produced under certain conditions) is abolished by multiple recall tests. In both experiments, the generate condition produced greater hypermnesia (increased recall over tests) than did the read condition.

Self-generated information is typically remembered better than information that is merely perceived, a result known as the *generation effect* (Slamecka & Graf, 1978). Generation effects are quite common and have been observed with a variety of generation conditions, materials, and memory tests. For example, generating words from antonyms, semantic associates, rhymes, word fragments, definitions, or second-language translations can all enhance memory, as compared with reading or hearing the same words. Generation enhances memory not only for words, but also for meaningful abbreviations, word compounds, sentences, text, numbers, and pictures. Generation effects have been observed on a number of memory tests, such as recognition memory, free and cued recall, and comprehension tests (see Greene, 1992, and Mulligan, 2001, for reviews).

Although the generation effect has substantial generality, generation does not always enhance memory. From a theoretical perspective, the most important limiting condition for the generation effect is experimental design. When encoding conditions vary randomly within a list, generation produces substantially better recall than does reading. However, when encoding conditions are varied between subjects (or in pure lists, within subjects), generation typically does not enhance recall (e.g., Begg & Snider, 1987; Grosofsky, Payne, & Campbell, 1994; Hirshman & Bjork, 1988; Kinoshita, 1989; Slamecka & Katsaiti, 1987). In fact, under some conditions, between-subjects manipulations of generation produce a negative generation effect, in which reading words produces better recall than

does generating them (see Steffens & Erdfelder, 1998, for a review).

The most successful account of the generation effect is the multifactor account (e.g., deWinstanley & Bjork, 1997; Hirshman & Bjork, 1988; Hunt & McDaniel, 1993; McDaniel, Waddill, & Einstein, 1988; Mulligan, 2001; Steffens & Erdfelder, 1998), which is based on the distinction between item-specific and relational information (Hunt & McDaniel, 1993). According to this view, generation enhances encoding of item-specific features of the target item, those characteristics that differentiate the item from other items in the list and increase item distinctiveness. Consistent with this view, if generation is implemented in either a between- or a within-subjects design, it enhances recognition memory, a test particularly sensitive to item-specific encoding (Begg, Snider, Foley, & Goddard, 1989; Hunt & McDaniel, 1993; Kinoshita, 1989). If the target word is generated from a cue word, the multifactor view proposes that generation also enhances the processing of the cue–target relation, which is particularly important in contributing to generation effects in cued recall (e.g., Hirshman & Bjork, 1988; Steffens & Erdfelder, 1998).

Intertarget (or list-wide) relational processing is a second type of relational information (besides cue–target information). This refers to relationships between target items of different study trials (rather than between a cue and the target item within a study trial). For present purposes, intertarget relations are the most relevant type of relational information and will often be referred to simply as *relational information*.

Free recall of targets relies heavily on this type of relational encoding (in addition to item-specific encoding; e.g., Hunt & McDaniel, 1993; Steffens & Erdfelder, 1998). According to the multifactor view, the act of generation focuses encoding resources on the target item and the cue–target relation, drawing resources away from processing

This research was partially supported by Grant 1-R03-MH61324-01 from NIMH. We gratefully acknowledge the assistance of Melissa Dela-Garza and Chirine Boukarroum. Address correspondence to N. W. Mulligan, Department of Psychology, Southern Methodist University, Dallas, TX 75275-0442 (e-mail: mulligan@mail.smu.edu).

associations between target items (e.g., Burns, 1992; Hunt & McDaniel, 1993; Steffens & Erdfelder, 1998). In pure lists, generation disrupts intertarget processing over the entire list and detracts from the usual generation advantage in recall. This causes generation effects to disappear or even reverse in free recall for between-subjects designs (Grososky et al., 1994; Hirshman & Bjork, 1988; Slamecka & Katsaiti, 1987; Steffens & Erdfelder, 1998), even though the generation advantage persists in recognition, a test with little reliance on relational processing (e.g., Begg et al., 1989; Kinoshita, 1989). When generate and read items are intermixed, disruptions of intertarget processing produced by generation affect both generate and read items. Consequently, read and generate items are expected to produce equivalent relational encoding in a mixed list, which in turn permits the superior item-specific encoding in the generate condition to produce a recall advantage (Hunt & McDaniel, 1993).¹

The fact that between-subjects designs produce null and even negative generation effects has played a critical role in the development of the multifactor account, as well as other accounts of the generation effect (e.g., Begg & Snider, 1987; Begg, Vinski, Frankovich, & Holgate, 1991; Grososky et al., 1994; Schmidt & Cherry, 1989; Slamecka & Katsaiti, 1987; see Steffens & Erdfelder, 1998, for a review). It is of interest, then, that the recent results of Mulligan (2001) call into question the extent to which experimental design is actually a limiting condition for the generation effect. Mulligan (2001, Experiment 1), using a between-subjects design, presented study pairs in either a read (e.g., *hot-cold*) or a generate (e.g., *hot-c*____) condition. In the latter condition, subjects generated the target word from its antonym and its initial letter. Subsequently, the subjects were presented with a series of five free recall tests for the target words. Consistent with earlier studies, no generation effect was observed on the initial (or the second) test. However, a significant generation effect emerged on Test 3 and persisted on Tests 4 and 5. This emergent generation effect has also been found when words are generated from semantic or category associates, although not all generation manipulations (e.g., generating from word fragments) produce this pattern (Mulligan, in press). One last aspect of Mulligan's (2001) results deserves mention: The generate (but not the read) condition exhibited increased recall over tests, a phenomenon known as *hypermnnesia* (Erdelyi, 1996; Payne, 1987).²

On the basis of previous research (Mulligan, 2001), we know that a between-subjects manipulation of generation producing no effect on an initial recall test may produce significant generation effects on subsequent tests (the emergent generation effect). Because of the theoretical importance of the effects of experiment design on the generation effect, it is important to further delineate this result. In Experiment 1, we examined whether the emergent generation effect requires multiple tests or whether a single recall test of extended duration would produce the same result. In Experiment 2, we examined whether repeated testing can eliminate the negative generation effect. Re-

peated recall tests convert a null generation effect into a positive generation effect. To the extent that generation produces greater hypermnnesia than does reading (Mulligan, 2001), repeated tests are expected to eliminate the negative generation effect.

EXPERIMENT 1

The results of Mulligan (2001) demonstrated that multiple recall tests alleviate the limiting condition for the generation effect produced by a between-subjects manipulation. In the context of hypermnnesia research, it is natural to wonder whether the emergent generation effect is due to the use of multiple recall tests per se or whether the effect is due to the increased total retrieval time afforded by multiple tests. This type of question is typically assessed by comparing recall performance across several, discrete recall tests with performance on a single (long) recall test of equal total duration (e.g., Roediger & Thorpe, 1978; Turtle & Yuille, 1994). On the one hand, there are reasons to expect that the two testing methods might induce differences in retrieval practice and retrieval strategies (see Mulligan, 2002b). Multiple tests require that the study list be recalled anew several times, whereas in the single-test condition, each word need be recalled only once. Consequently, the multiple-test condition affords more retrieval practice, which may render items more easily recalled on later tests (Bjork, 1988; Dempster, 1996; Morris & Fritz, 2000). In addition, there is evidence that multiple recall tests produce changes in search strategies, as compared with a single, extended recall attempt (e.g., McDaniel, Moore, & Whiteman, 1998; Mulligan, 2002b). The single and multiple tests may differ in terms of the emergent generation effect, as well.

On the other hand, extant data suggest similarity between multiple tests and a single, time-matched test. Consider the results of Roediger and Thorpe (1978). In this study, subjects were presented with a list of words and then recalled the words, using either a series of three successive 7-min tests or a single 21-min recall test. Recall was assessed by examining the total number of different items recalled (cumulative recall) at equivalent points in time (e.g., the end of Test 1 for the multiple-recall group vs. the 7-min mark for the single-test subjects, the end of Test 2 for the multiple-recall group vs. the 14-min mark for the single-test subjects, etc.). Specifically, for cumulative recall, an item was scored as correct on its first recall, and subsequent recalls of the same item were ignored. Roediger and Thorpe found no difference in cumulative recall between the groups, a result replicated several times with varying materials and conditions (e.g., Payne, 1986; Turtle & Yuille, 1994). This finding implies that repeated-recall tests produce hypermnnesia because they provide additional retrieval time. The consequent implication is that multiple-recall tests are not functionally different from a single-recall test of equivalent total time (this is a core assumption of an important view of hypermnnesia, the cumulative recall level hypothesis; Payne, 1987; Roediger &

Challis, 1989). Relating these results to the emergent generation effect implies that a single, long recall test would produce the same emergent generation effect as that reported by Mulligan (2001).

Experiment 1 used a between-subjects manipulation of generation and was modeled on Mulligan (2001, Experiment 1). The study lists consisted of word pairs in which the second item was an antonym of the first item and was presented either intact to be read (e.g., *hot-cold*) or as a one-letter word stem to be generated (e.g., *hot-c_____*). Antonym generation was used because it is a standard generation manipulation (e.g., Masson & MacLeod, 1992; Slamecka & Graf, 1978). There were two testing conditions. One group received four consecutive free recall tests, in which they attempted to recall the second word of each pair. Each of the tests lasted 5 min. The second group was given a single, 20-min recall test. The multiple test group was a replication of Mulligan (2001, Experiment 1), and the same results were expected—specifically, no generation effect on the initial test, but significant generation effects on later tests. On the basis of the hypermnnesia literature (Payne, 1986; Roediger & Thorpe, 1978; Turtle & Yuille, 1994), it seemed likely that the single-recall group would exhibit cumulative recall similar to that of the multiple-recall group. This implies that over the first several minutes (e.g., the first 5 min, equivalent to the end of the Recall Test 1 in the multiple-test group), no significant difference between read and generate would be detected (i.e., a null effect over that time range) but that a significant generation effect would be observed at a later time in the recall period.

Method

Subjects. One hundred twenty-eight undergraduates at Southern Methodist University participated in exchange for extra credit in psychology courses.

Design and Materials. Encoding condition (read vs. generate) and test condition (single test vs. multiple tests) were manipulated between subjects. Within the multiple-test condition, recall test (1–4) was manipulated within subjects.

The critical stimuli were 44 cue–target antonym pairs used in Mulligan (2001, Experiment 1). The target words were between four and six letters in length and varied in Kučera–Francis (1967) frequency from 1 to 1,070, with a mean of 192. The 44 critical word pairs were randomly ordered. Five additional antonym pairs were developed, one of which served on a practice trial and the remaining four of which were used as primacy (two) and recency (two) buffers in the study lists. Two versions of the study lists were created. In the read condition, the study stimulus consisted of both words presented in an intact form (e.g., *hot-cold*). In the generate condition, the study stimulus consisted of the first word and the first letter of the second word followed by an underscore (e.g., *hot-c_____*). The study stimuli were copied to slides and were presented via a slide projector.

Procedure. The subjects were tested in groups of 3 to 8. The subjects in the read condition were instructed to read both words (silently) and to write the second word of each pair on their answer sheet.³ They were further instructed to try to remember the second word (the one written down) for a later (unspecified) memory test. In the generate condition, the subjects were told that, on each trial, they would see a word followed by the first letter of a second word. They were instructed to read the first word and to try to think of an

antonym for the word that started with the letter presented. They were to write their answer on the sheet and to try to remember the written word for a later memory test. A practice trial preceded the study list (in both the read and the generate conditions) to ensure that the subjects understood the task. Each study pair was presented for 8 sec.

Following the study task, the subjects in the multiple-test condition were given the first free recall memory test. The subjects were presented with a test sheet and were asked to recall the target words from the study phase. It was made clear that the target words were those words written down on the answer sheet during the study phase. The test lasted 5 min. The subjects were encouraged to use the entire time to remember as many words as possible. At the end of the first test, the test sheets were collected, and the subjects were then presented with a second free recall test. The subjects were given a new test sheet and again were asked to recall as many target words from the study list as possible. The second test also lasted 5 min. The subjects were encouraged to continue trying to remember words for the entire 5-min period. This procedure was repeated for Tests 3 and 4.

In the single-test condition, the study phase was followed by a single-recall test of 20 min. As in the multiple-test condition, it was made clear that the to-be-recalled words were those written down on the answer sheet during the study phase. The subjects were asked to use the entire time to try to recall the words. They were informed that people continue to remember words over the entire time period. Every 5 min, the subjects were asked to draw a line on their test paper. Any items recalled subsequently were to be written below the line, enabling computation of the number of items recalled in each 5-min epoch. In addition, at the 5-min intervals, the subjects were encouraged to continue trying to recall words.

Results and Discussion

At study, target words were correctly produced on 96% and 100% of the trials in the generate and in the read conditions, respectively. Although performance was quite high in the generate group, as well as in the read group, analyses were performed on the test data both conditionalized on correct performance at study (i.e., excluding items that were not generated correctly) and unconditionalized. Not surprisingly, the two sets of analyses led to the same conclusions. Only the unconditionalized analyses are reported for this and the subsequent experiment.

Multiple-test condition. The multiple-test condition was identical to Mulligan (2001, Experiment 1), with the exception that four, rather than five, recall tests were administered. Consequently, it is important to first analyze the multiple-test condition in isolation to determine whether it replicated the results of the prior experiment. The recall data are presented in Table 1. Net recall (the proportion of studied words recalled on a given test) was analyzed with a 2×4 analysis of variance (ANOVA), using encoding condition as a between-subjects factor and recall test (1–4)

Table 1
Experiment 1, Multiple-Recall Condition: Mean Net Recall as a Function of Encoding Condition and Recall Test

Encoding Condition	Recall Test			
	1	2	3	4
Generate	.21	.24	.26	.29
Read	.19	.20	.20	.21

Table 2
Experiment 1, Multiple-Recall Condition: Mean Number of
Item Gains and Losses as a Function of Encoding Condition
and Recall Test

Result	Encoding Condition	Between Tests			Total
		1-2	2-3	3-4	
Gains	generate	3.13	2.38	2.69	8.20
	read	1.91	1.59	1.44	4.94
Losses	generate	2.00	1.72	1.56	5.28
	read	1.69	1.25	0.97	3.91

as a within-subjects factor (an alpha level of .05 was used for this and subsequent tests). The main effects of encoding condition [$F(1,62) = 4.98, MS_e = 0.0308$] and of test [$F(3,186) = 15.34, MS_e = 0.0016$] were significant, as was their interaction [$F(3,186) = 4.56, MS_e = 0.0016$]. Planned follow-up tests revealed a significant effect of test in the generate group [$F(4,93) = 18.21, MS_e = 0.0016$], but not in the read group ($F = 1.67, p > .15$), indicating that hypermnesia occurred only under generation instructions. In addition, planned contrasts within tests indicated no generation effect for Test 1 ($|t| < 1$), a marginally significant generation effect for Test 2 [$t(62) = 1.98, p < .06$], and significant generation effects for Tests 3 and 4 [$ts(62) = 2.34$ and 2.89 , respectively].

These results are consistent with Mulligan (2001, Experiment 1) in several ways. First, hypermnesia occurred for the generate condition, but not for the read condition. As Mulligan (2001) noted, the absence of hypermnesia in the read condition is consistent with prior research, which has shown that verbal materials typically do not produce hypermnesic recall unless the materials are presented multiple times or are encoded under instructions encouraging elaboration or the formation of mental images (e.g., Belmore, 1981; Latour & McKelvie, 1994; Payne, 1986, 1987). These conditions do not appear to apply to the read condition (see Mulligan, 2001). Second, generation produced no effect on the initial recall test, as was expected on the basis of prior between-subjects experiments that utilized a single recall test (e.g., Grosofsky et al., 1994; Hirshman & Bjork, 1988; Kinoshita, 1989; McDaniel et al., 1988; Slamecka & Katsaiti, 1987). Third, a significant generation effect emerged on the third and fourth tests.

Consistency between the present and prior experiments can also be assessed by examining gains and losses between tests. When multiple recall tests are used, some items are recalled on later tests that had not been recalled earlier (item gains), whereas other items that were successfully recalled earlier may not be recalled on later tests (item losses). Although not the primary focus of the present experiment, an analysis of item gains and losses was conducted because they have been used to index item-specific and relational encoding. Specifically, conditions fostering item-specific encoding increase item gains, whereas conditions fostering relational encoding reduce item losses (e.g., Burns & Gold, 1999; Klein, Loftus, Kihlstrom, & Aseron, 1989; McDaniel et al., 1998). These recall components have thus been used to assess theories

framed around the item-specific-relational distinction (e.g., Burns & Gold, 1999; Hunt & McDaniel, 1993; McDaniel et al., 1998; Mulligan, 2000; Olofsson, 1997). Mulligan (2001, Experiment 1) found both more gains and more losses in the generate condition than in the read condition, consistent with the multifactor view that generation simultaneously enhances item-specific encoding and disrupts relational encoding in between-subjects designs. The analysis of gains and losses is presented here to facilitate comparison with Mulligan (2001).

Item gains for test i were computed as the number of words recalled on test i , but not on test $i - 1$. Item losses for test i were computed as the number of words recalled on test $i - 1$, but not on test i . The number of gains and losses (Table 2) were analyzed with separate 2×3 ANOVAs, using encoding condition as a between-subjects factor and recall test (2-4) as a within-subjects factor. For gains, the effect of encoding condition was significant [$F(1,62) = 12.58, MS_e = 4.47$], indicating more gains in the generate condition than in the read condition. The effect of test approached significance [$F(2,124) = 2.47, p < .10$], and the interaction was not significant ($F < 1$). The analysis of losses produced two significant effects: (1) the main effect of encoding condition [$F(1,62) = 5.19, MS_e = 1.94$], indicating more losses in the generate condition than in the read condition, and (2) the main effect of test [$F(2,124) = 4.70, MS_e = 1.16$], indicating greater losses on the earlier tests. This latter result is consistent with earlier studies (e.g., Mulligan, 2000; Payne & Wenger, 1992; Roediger & Payne, 1985). The interaction was not significant ($F_s < 1$). As a set, the results are consistent with Mulligan (2001, Experiment 1) in showing both more gains and more losses in the generate condition than in the read condition.

Comparison of single- and multiple-test conditions. Cumulative recall (Table 3) is used to compare the single- and the multiple-test conditions. In the single-test condition, cumulative recall is simply the proportion of old items recalled by the end of each 5-min interval. In the multiple-test condition, cumulative recall is the proportion of old items recalled at least once across tests, treating each test as a 5-min interval.

The results of the single-test condition were similar to those of the multiple-test condition. Cumulative recall was analyzed with a $2 \times 2 \times 4$ ANOVA, using encoding condition and test condition (single test vs. multiple tests) as between-subjects factors and recall time (5, 10, 15, or 20 min) as a within-subjects factor. The analysis produced

Table 3
Experiment 1, Mean Cumulative Recall as a Function of Test
Condition, Encoding Condition, and Recall Time

Test Condition	Encoding Condition	Recall Time (min)			
		5	10	15	20
Multiple tests	generate	.21	.28	.32	.36
	read	.19	.24	.26	.28
Single test	generate	.23	.29	.35	.39
	read	.21	.25	.28	.32

three significant results: a main effect of encoding condition [$F(1,124) = 8.20, MS_e = 0.0400$], a main effect of recall time [$F(3,372) = 326.12, MS_e = 0.0012$], and a significant retrieval time \times encoding condition interaction [$F(3,372) = 13.96, MS_e = 0.0012$]. No other effects were significant ($F_s < 2, p_s > .10$). These results indicate that cumulative recall was greater in the generate condition than in the read condition, that cumulative recall increased with recall time (not surprisingly), and that this increase was greater in the generate condition than in the read condition. To further explore the interaction, separate 2 (encoding condition) \times 2 (test condition) ANOVAs were carried out for each of the recall times. The analysis indicated no generation effect at 5 min ($p > .15$) but significant generation effects at all three of the later recall times [$F_s(1,124) > 6.5$]. All other effects in this set of analyses were nonsignificant ($p_s > .14$).

An additional analysis verified that the single-test condition produced the same pattern of results when considered in isolation. A 2 (encoding condition) \times 4 (recall time) ANOVA of cumulative recall in the single-test condition revealed that encoding condition [$F(1,62) = 4.00, MS_e = 0.0427$] and recall time [$F(3,186) = 193.92, MS_e = 0.0011$] were significant, as was their interaction [$F(3,186) = 7.49, MS_e = 0.0011$]. Comparisons within tests indicated no generation effect at 5 min ($t = 1.08$), a marginally significant effect at 10 min [$t(62) = 1.74, p < .10$], and significant generation effects at 15 and 20 min [$t_s(62) = 2.32$ and 2.39 , respectively].

Several aspects of the cumulative recall analysis should be noted. Most important, cumulative recall reveals the same emergent generation effect as that found with net recall for the multiple-test condition. Specifically, a between-subjects manipulation of generation initially produced the expected null effect, followed by a significant generation effect at later times during the long recall test (or on later tests in the multiple-test condition). Second, test condition did not produce a main effect or significantly interact with any other variable. This indicates that the multiple- and single-test conditions produced comparable levels and similar patterns of cumulative recall (Roediger & Challis, 1989). This indicates that multiple-recall tests are not required to produce the emergent generation effect; a single, long recall test produces the same phenomenon.

This result underscores a point made some time ago by Roediger and Thorpe (1978), who noted that researchers typically measure recall for a fixed, usually brief period. However, recall typically does not reach asymptotic levels in the first few minutes, and conditions may continue to diverge over longer recall periods. Consequently, the length of the recall test may be an important variable in determining whether experimental conditions produce equivalent levels of recall. This seems to be the case in the present situation. Prior between-subjects studies of generation have used recall tests of 5 min or less (e.g., Burns, 1990, 1992; Grosofsky et al., 1994; Hirshman & Bjork, 1988; Kinoshita, 1989; Slamecka & Katsaiti, 1987). Consistent with the results of these prior studies, the present experi-

ment indicated that, at that recall time, the generate and the read conditions did not significantly differ. However, at longer recall times (e.g., 15 or 20 min), a significant difference was found.

EXPERIMENTS 2A AND 2B

Although generation typically enhances memory, it may, under some circumstances, harm recall (the negative generation effect; e.g., Burns, 1992; Schmidt & Cherry, 1989). In a review of the literature, Steffens and Erdfelder (1998) accounted for the negative generation effect in terms of the multifactor account, arguing that negative generation effects occur when the generation task conflicts with the intertarget relations. For example, Steffens and Erdfelder presented a series of study pairs in which the cue and the target words rhymed. In the generation condition, the target word was generated from the rhyming cue; in the read condition, both words were presented intact. Importantly, target items from sequential study pairs were members of the same category. Under these conditions, the intertarget relation and the cue-target relation do not match. The former relation is categorical-semantic, and the latter is a rhyme relationship. Steffens and Erdfelder argued that, under these conditions, the generation task is especially disruptive of intertarget relational processing, whereas the reading condition is relatively conducive to intertarget relational processing. Steffens and Erdfelder argued that, under these conditions, the generation condition produces worse recall on a test with heavy reliance on intertarget relational information. In contrast, when the generation task (and the cue-target relation) matches the intertarget relations (i.e., both are semantic or categorical), the generation task enhances intertarget relational encoding and produces a consequent positive generation effect in recall (e.g., McDaniel et al., 1988; Steffens & Erdfelder, 1998; see note 1).

In Experiment 2A, we examined whether the negative generation effect is moderated by multiple retrieval attempts. The encoding conditions were modeled on Steffens and Erdfelder (1998, Experiment 1), because these conditions produce a negative generation effect on an initial recall test. The study session was followed by a series of five recall tests. The first test was expected to replicate Steffens and Erdfelder in producing a negative generation effect. The question of most interest was whether the negative generation effect would be eliminated by multiple retrieval attempts. Because generation leads to greater hypermnesia than does reading (the present Experiment 1; Mulligan, 2001), the negative generation effect was expected to be diminished or eliminated on later recall tests. As will be discussed in the General Discussion section, recent theorizing by McDaniel et al. (1998) has pointed to the same outcome as this empirically derived prediction.

The prediction of worse relational encoding in the generate condition can be examined more directly by examining the amount of category clustering at recall. Category clustering (typically assessed with the adjusted ratio of

clustering, or ARC, score; Roenker, Thompson, & Brown, 1971) is a standard method for assessing relational encoding (e.g., Hunt & Einstein, 1981; Hunt & McDaniel, 1993). The multifactor account predicts less category clustering in the generate condition than in the read condition, a result observed in earlier studies (Burns, 1990, 1992). In addition, when the list items are made up of members of common categories, retrieval becomes increasingly organized around the categorical structure of the list over recall trials (Mulligan, 2001, 2002b). Thus, the amount of clustering was expected to increase across tests in the present experiment.

Finally, it should be noted that the multifactor account proposes that generation generally enhances item-specific encoding, regardless of its effect on relational encoding. To determine whether generation had this effect under the present conditions, a secondary experiment (Experiment 2B) was conducted in which the study session was followed by a single test of recognition for the target items. Recognition is especially sensitive to item-specific information and relatively insensitive to relational encoding (e.g., Begg et al., 1989; Hunt & McDaniel, 1993). Consequently, a positive generation effect is expected in recognition.

Method

Subjects. The subjects were undergraduates at Southern Methodist University, who participated in exchange for extra credit in psychology courses. Seventy-two participated in Experiment 2A, and 32 participated in Experiment 2B.

Design and Materials. In Experiment 2A, encoding condition (read vs. generate) was manipulated between subjects, and recall test (1–5) was manipulated within subjects. In Experiment 2B, the recall tests were replaced by a single recognition test.

The critical target words consisted of 14 examples from each of six categories. The examples were three to seven letters in length and of high to moderate category frequency (Battig & Montague, 1969). Two master sets of 42 critical items each were created by randomly dividing the examples from each category into two sets. From each master set, two study lists were created, a read list and a generate list. Each critical word was paired with a rhyming word that bore no obvious semantic relationship to the critical item. In the read lists, both the cue and the target words were presented intact in lower case (e.g., *booth-tooth*). In the generate lists, the target item was presented in fragment form, in which from one to three letters (depending on the length of the word) were replaced by underscores (e.g., *booth-to_ _*). The word pairs were organized so that all the target items from the same category appeared in sequence. Four additional examples were selected from a seventh category to serve as buffer items. These examples were also paired with rhyming cues. Two of the buffer pairs were presented at the beginning and two at the end of each study list and were not scored for recall. The study stimuli were copied to slides and were presented via slide projector. The recognition test of Experiment 2B consisted of all of the critical items, half of which were presented as target items at study (the old items) and half were not (the new items). Because old and new items came from the same categories, category membership could not be used as a basis for the old–new decision. The recognition list was presented in two columns on a two-page test.

Procedure. The subjects were tested in groups of 2 to 6. The study phase of the experiment was the same as that in Experiment 1, with changes to the generation instructions to reflect the rhyming relationship of the cue and the target words. In Experiment 2A, the

study task was followed by a series of recall tests. This procedure was identical to the multiple-recall condition of Experiment 1, with the exception that five tests were administered. In Experiment 2B, the study task was followed by a single recognition memory test. The subjects were given the recognition test sheet and were asked to circle words that were read or generated during the study session. The recognition test lasted 5 min.

Results

At study, target words were correctly generated on 99% of the trials in both Experiments 2A and 2B. Of the read words, 100% were correctly copied in both experiments.

Recall results of Experiment 2A. Net recall (Table 4) was analyzed with a 2×5 ANOVA, using encoding condition and recall test as factors. The analysis showed a significant main effect of test [$F(4,280) = 23.42$, $MS_e = 0.0029$], revealing hypermnesia, and a test \times encoding condition interaction [$F(4,280) = 6.01$, $MS_e = 0.0029$], indicating greater hypermnesia in the generate condition than in the read condition. The effect of encoding condition was not significant ($F = 1.11$). Planned follow-up analyses revealed significant hypermnesia in both the generate [$F(4,140) = 16.08$, $MS_e = 0.0047$] and the read [$F(4,140) = 8.68$, $MS_e = 0.0011$] conditions. In addition, planned contrasts within tests indicated significantly *worse* performance in the generate condition than in the read group on Test 1 [$t(70) = -2.26$], but no significant differences on Tests 2–5 [$t(70) = -1.44, -1.07, -0.58$, and 0.12 , respectively, $ps > .15$]. Thus, a negative generation effect was observed on the first test, but not on subsequent tests.

Cumulative recall is also reported because this figure is important to theoretical accounts of hypermnesia (as described in Experiment 1; e.g., Payne, 1986; Roediger & Thorpe, 1978) and because it cannot be derived from the other reported results. Cumulative recall was analyzed with a 2 (encoding condition) \times 5 (recall test) ANOVA. The analysis yielded results consistent with net recall, including a significant main effect of test [$F(4,280) = 151.36$, $MS_e = 0.0016$] and a significant test \times encoding condition interaction [$F(4,280) = 15.82$, $MS_e = 0.0016$]; the main effect of encoding was not significant, $F < 1$. These results indicated that cumulative recall increased over tests and that the increase was greater in the generate condition than in the read condition.

Table 4
Experiment 2A: Recall Performance as a Function of Encoding Condition and Recall Test

Performance	Encoding Condition	Recall Test				
		1	2	3	4	5
Net recall	generate	.34	.35	.39	.42	.46
	read	.41	.41	.43	.44	.45
Cumulative recall	generate	.34	.41	.46	.49	.53
	read	.41	.45	.48	.50	.51
ARC scores	generate	.31	.31	.42	.49	.51
	read	.52	.56	.52	.64	.61

Note—Net recall and cumulative recall are equivalent on Test 1. ARC, adjusted ratio of clustering.

Table 5
Experiment 2A: Mean Number of Item Gains and Losses as a
Function of Encoding Condition and Recall Test

Result	Encoding Condition	Between Tests				Total
		1-2	2-3	3-4	4-5	
Gains	generate	2.83	3.28	2.64	2.67	11.42
	read	1.75	1.83	1.22	1.19	6.00
Losses	generate	2.31	1.78	1.42	1.28	6.78
	read	2.11	0.92	0.81	0.89	4.73

Category clustering was assessed with the ARC score, which measures the number of pairs of category members recalled in adjacent positions in the recall protocol, corrected for the number of adjacent category pairs expected from a random ordering of the recalled items (see Roenker et al., 1971, for details). A value of 0 indicates chance clustering, and a value of +1.00 indicates perfect clustering. ARC scores were computed for each subject on each test (Table 4). To determine whether generation disrupted intertarget relational encoding, the analysis was restricted to Test 1. This is because clustering on the initial test provides the purest assessment of differences in intertarget relational processing arising during encoding; clustering on later tests presumably reflects influences of study-list processing, as well as reprocessing that takes place during the earlier tests. On Test 1, ARC scores were significantly lower in the generate condition than in the read condition [$t(70) = 2.11$], indicating that generation disrupted intertarget processing at encoding. This replicates the results of Burns (1990, Experiment 1C; 1992, Experiment 3) in experiments in which similar materials and encoding instructions were used. In fact, a significant effect of encoding on clustering persisted on Test 2 [$t(70) = 2.38$], but not on later tests [$|t|s < 1.5, ps > .10$]. In addition, the ARC scores were submitted to a 2×5 ANOVA, using encoding condition and recall test as factors. The analysis produced a significant effect of test [$F(4,280) = 6.97, MS_e = 0.0510$], indicating that clustering increased across tests. The main effect of encoding approached significance [$F(1,70) = 3.06, p = .08$]. The interaction was not significant ($F = 1.51, p > .15$).

The number of gains and losses (Table 5) were analyzed with separate 2×4 ANOVAs, using encoding condition as a between-subjects factor and recall test (2-5) as a within-subjects factor. For gains, the only significant effect was that of encoding condition [$F(1,70) = 20.62, MS_e = 6.40$], indicating more gains in the generate condition than in the read condition (other $F_s < 1.4, ps > .25$). The analysis of losses produced two significant effects: (1) a main effect of encoding condition [$F(1,70) = 4.27, MS_e = 4.45$], indicating more losses in the generate condition than in the read condition, and (2) a main effect of test [$F(3,210) = 8.34, MS_e = 2.40$], indicating greater losses on the earlier tests. The interaction was not significant ($F_s < 1$).

Recognition results of Experiment 2B. For the generation group, the mean proportions of hits and false

alarms were .86 and .05, respectively. The corresponding means for the read group were .74 and .07. An analysis of d' recognition accuracy indicated that recognition was significantly better in the generate condition than in the read condition [2.95 vs. 2.33; $t(30) = 2.51$]. The same results were obtained when recognition accuracy was assessed with corrected *old* scores.

Discussion

The central focus of this experiment was to determine whether the negative generation effect is eliminated over multiple-recall tests. Thus, the most important results were net recall in Experiment 2A, several aspects of which should be noted. First, a negative generation effect was obtained on the initial recall test. This was expected and represents a replication of Steffens and Erdfelder (1998; see also Burns, 1990, 1992). Second, the generation condition produced greater hypermnesia than did the read condition, consistent with earlier results (Mulligan, 2001, in press). It should be noted that the present experiment yielded significant hypermnesia for both the generate and the read conditions, whereas Experiment 1 yielded hypermnesia only for the generate condition. The difference in results may be due to the use of categorically related target items in the present experiment (Experiment 1 used unrelated target items). Specifically, Mulligan (2001, Experiment 4) found that with categorized lists (i.e., sets of target items were members of common categories), both generation and reading at encoding produced hypermnesia, with the generate condition producing more hypermnesia (see Mulligan, 2001, for a full discussion). With lists of unrelated target items, only the generation condition produced hypermnesia (as in Experiment 1; Mulligan, 2001).

Third and most important, the negative generation effect observed on the initial test was eliminated on later tests, with the generate condition producing nonsignificantly greater recall by the final test. Thus, an important aspect of the emergent generation effect generalized to the conditions that produced the negative generation effect. Specifically, over repeated recall tests (or extended recall time), the generation condition produced greater increases in recall than did the read condition, inducing a positive generation effect under conditions that initially produced a null result and equivalent recall under conditions initially producing a negative generation effect. It is tempting to extrapolate the present results to additional recall tests and wonder whether a significantly positive generation effect would be obtained on a sixth or seventh recall test. The literature on hypermnesia suggests that this is unlikely, however, because recall of word lists is typically near asymptotic levels after 20 or 25 min of retrieval (see Roediger & Challis, 1989).

As a secondary point, it should be noted that several aspects of the results conformed to the item-specific-relational analysis of generation. This view argues that generation is expected to enhance item-specific encoding, even in the present paradigm in which its overall effect on

recall is negative (at least on an initial test). In addition, generation is predicted to disrupt relational encoding. In the present experiments, item-specific encoding may be evaluated via recognition performance (in Experiment 2B) or by the number of gains across recall tests (in Experiment 2A). In Experiment 2B, recognition accuracy was greater in the generate condition than in the read condition. In Experiment 2A, gains were greater in the generate condition than in the read condition. In both cases, the results were consistent with greater item-specific processing in the generate condition.

Intertarget relational encoding may also be evaluated in two ways, via ARC scores or intertest losses, both of which indicated disrupted relational encoding in the generate condition. Specifically, the ARC scores on Test 1 were lower for the generate group than for the read group. Likewise, the number of losses was greater in the generate group. The measures of relational information also make a second, interesting point. ARC scores significantly increased across tests, whereas losses significantly decreased. These results indicated an increasing use of relational (i.e., categorical) information during retrieval, replicating earlier studies that have reported that measures of objective or subjective organization increased with recall trials (e.g., Mulligan, 2001, 2002a; Payne & Wenger, 1992).

GENERAL DISCUSSION

The present experiments were motivated by the emergent generation effect and further delineated this phenomenon. Experiment 1 demonstrated that multiple-recall tests are not necessary to produce this effect; the generation effect emerges on a single test of relatively long duration. Experiment 2 extended the phenomenon to conditions initially producing a negative generation effect. The negative generation effect was abolished by repeated recall trials. Of course, the results of Experiment 1 (see also Payne, 1986; Roediger & Thorpe, 1978; Turtle & Yuille, 1994) imply that the negative generation effect would also be abolished with a single, long-duration test. Coupled with prior research (Mulligan, 2001, *in press*), these results imply that the greater hypermnesia in the generation condition than in the read condition induces positive generation effects when the two conditions are initially equal (e.g., Experiment 1) or equivalent recall when the generation condition initially produces a negative generation effect (Experiment 2). Consistent with these results, it should be noted that in within-subjects designs (Mulligan, 2001), the generation effect has been found on the initial test and then increased in size on later tests (alternatively stated, generating produced greater hypermnesia than did reading in within-subjects as well as between-subjects designs).

The present results are theoretically important. The results of Experiment 1 (along with the results of Mulligan, 2001) indicate that experimental design is not a true limiting condition for the generation effect, either across multiple-recall tests or even for a single-recall test. Likewise, the

encoding conditions of Experiment 2 did not produce a true negative generation effect; asymptotic levels of recall in the generate and the read conditions did not differ. Theories of the generation effect have focused heavily on the moderating effects of experimental design (e.g., Begg et al., 1991; Grosofsky et al., 1994; Hunt & McDaniel, 1993; McDaniel et al., 1988; Nairne, Reigler, & Serra, 1991; Slamecka & Graf, 1978; Steffens & Erdfelder, 1998; see Mulligan, 2001, for a review) and need to take these findings into consideration.

One potential account of these results is based on the item-specific-relational account of hypermnesia developed by McDaniel et al. (1998; Hunt & McDaniel, 1993; see the discussion in Mulligan, 2001). According to this view, relational information plays an important role in guiding retrieval strategies and selecting potential responses. If targets (or subsets of targets) are related to one another during encoding, it is expected that stable retrieval strategies will emerge quickly and be applied consistently on successive recall trials (Hunt & McDaniel, 1993; McDaniel et al., 1998). In this case, items associated to the consistent relational cues used at retrieval will be consistently generated, which should result in few item losses from one recall trial to the next. In contrast, if targets have not been encoded relationally, it may take longer for the subject to develop effective retrieval routes or strategies. In this case, item losses should especially raise problems over the first few tests, but less so on later tests as more effective search routines are employed (Hunt & McDaniel, 1993; McDaniel et al., 1998; Payne, 1986; Payne & Wenger, 1992). Thus, this framework suggests that the disrupted relational encoding in the generate condition raises the most problems initially, but less so later, as retrieval strategies become more routinized. Item-specific encoding is hypothesized to lead to richer, more extensive encoding of item attributes, leading to a more distinctive trace and facilitating discrimination among potential responses at retrieval. In this framework, greater item-specific encoding is proposed to enhance item gains over tests (see McDaniel et al., 1998, for details). Consequently, as retrieval strategies stabilize and the initial impairments in relational encoding become less determinative of recall, the enhanced item-specific processing in the generate condition asserts greater influence on recall. This produces greater increases in recall across tests (i.e., greater hypermnesia), with the concomitant emergence of the generation effect in Experiment 1 or the elimination of the negative generation effect in Experiment 2.

This account is consistent with several findings. First, generation generally enhances item-specific encoding, as evidenced by the generation effect in recognition in between-subjects designs (e.g., Experiment 2B; Begg et al., 1989; Kinoshita, 1989; Mulligan, *in press*). Second, two pieces of evidence are consistent with the notion of increasingly stable retrieval strategies: Category clustering scores increased across recall tests in Experiment 2A (Mulligan, 2001; see Mulligan, 2002a, and Payne & Wenger, 1992, for similar results), and item losses were greater across earlier

tests than across later tests in both Experiments 1 and 2A (a typical result; e.g., McDaniel et al., 1998; Mulligan, 2001, 2002a; Payne & Wenger, 1992; Roediger & Payne, 1985). The final issue raised by this account is whether the changes in retrieval strategies occurring across multiple tests are similar to changes in retrieval strategies occurring across time within a single, long test. To apply the item-specific-relational analysis to the present results, this is assumed, consistent with conclusions from the hypermnnesia literature (e.g., Payne, 1987; Roediger & Challis, 1989; Roediger & Thorpe, 1978; Turtle & Yuille, 1994). However, this assumption could bear more direct research.

Finally, several other encoding manipulations produce opposing effects on item-specific and relational information similar to those produced by generation. These manipulations include orthographic distinctiveness, bizarre imagery, enactment effects, perceptual interference, and word frequency. It has been proposed that in each of these manipulations, one of the encoding conditions is relatively unusual (e.g., the bizarre condition, the enacted condition, low-frequency words) and attracts greater processing of item-specific features at the cost of interitem relational encoding. Like generation, the effects of each of these variables on recall is moderated by experimental design (e.g., DeLosh & McDaniel, 1996; Engelkamp & Dehn, 2000; Hunt & Elliot, 1980; Hunt & McDaniel, 1993; McDaniel & Einstein, 1986; Mulligan, 1999). Such findings have played a critical role in theoretical accounts of item-specific encoding effects (e.g., Begg & Snider, 1987; DeWinstanley & Bjork, 1997; Engelkamp & Dehn, 2000; Hirshman & Bjork, 1988; Hunt & McDaniel, 1993; Mulligan, 1999; Nairne et al., 1991; Serra & Nairne, 1993; Slamecka & Katsaiti, 1987; Steffens & Erdfelder, 1998). The present results, especially those of Experiment 1, indicate that between- and within-subjects designs may not produce such starkly different results if extended recall time or multiple recall trials are provided. The results of Mulligan (2002a) indicate just this. Relative to reading words, interfering with perception during encoding (by backward masking) enhances recall in a within-subjects design. In a between-subjects design, Mulligan (2002a) found that perceptual interference produced no advantage on an initial recall test. However, the perceptual interference effect emerged on subsequent recall tests, producing the same sort of emergent effect as that found with generation.

REFERENCES

- BATTIG, W. F., & MONTAGUE, W. E. (1969). Category norms of verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monographs*, **80**(3, Pt. 2).
- BEGG, I., & SNIDER, A. (1987). The generation effect: Evidence for generalized inhibition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 553-563.
- BEGG, I., SNIDER, A., FOLEY, F., & GODDARD, R. (1989). The generation effect is no artifact: Generating makes words distinctive. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 977-989.
- BEGG, I., VINSKI, E., FRANKOVICH, L., & HÖLGATE, B. (1991). Generation makes words memorable, but so does effective reading. *Memory & Cognition*, **19**, 487-497.
- BELMORE, S. M. (1981). Imagery and semantic elaboration in hypermnnesia for words. *Journal of Experimental Psychology: Human Learning & Memory*, **7**, 191-203.
- BJORK, R. A. (1988). Retrieval practice and the maintenance of knowledge. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues. Vol. 1: Memory in everyday life* (pp. 396-401). New York: Wiley.
- BURNS, D. J. (1990). The generation effect: A test between single- and multi-factor theories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 1060-1067.
- BURNS, D. J. (1992). The consequences of generation. *Journal of Memory & Language*, **31**, 615-633.
- BURNS, D. J., & GOLD, D. E. (1999). An analysis of item gains and losses in retroactive interference. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 978-985.
- DELOSH, E. L., & MCDANIEL, M. A. (1996). The role of order information in free recall: Application to the word-frequency effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 1136-1146.
- DEMPSTER, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. L. Bjork & R. A. Bjork (Eds.), *Memory: Handbook of perception and cognition* (pp. 317-344). San Diego: Academic Press.
- DEWINSTANLEY, P. A., & BJORK, E. L. (1997). Processing instructions and the generation effect: A test of the multifactor transfer-appropriate processing theory. *Memory*, **5**, 401-421.
- ENGELKAMP, J., & DEHN, D. A. (2000). Item and order information in subject-performed tasks and experimenter-performed tasks. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 671-682.
- ERDELYI, M. H. (1996). *The recovery of unconscious memories: Hypermnnesia and reminiscence*. Chicago: University of Chicago Press.
- ERDELYI, M. [H.], BUSCHKE, H., & FINKELSTEIN, S. (1977). Hypermnnesia for Socratic stimuli: The growth of recall for an internally generated memory list abstracted from a series of riddles. *Memory & Cognition*, **5**, 283-286.
- GREENE, R. L. (1992). *Human memory: Paradigms and paradoxes*. Hillsdale, NJ: Erlbaum.
- GROSOFSKY, A., PAYNE, D. G., & CAMPBELL, K. D. (1994). Does the generation effect depend upon selective displaced rehearsal? *American Journal of Psychology*, **107**, 53-68.
- 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.
- HUNT, R. R., & EINSTEIN, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning & Verbal Behavior*, **19**, 497-514.
- HUNT, R. R., & ELLIOT, J. M. (1980). The role of nonsemantic information in memory: Orthographic distinctiveness effects on retention. *Journal of Experimental Psychology: General*, **109**, 49-74.
- HUNT, R. R., & MCDANIEL, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory & Language*, **32**, 421-445.
- KINOSHITA, S. (1989). Generation enhances semantic processing? The role of distinctiveness in the generation effect. *Memory & Cognition*, **17**, 563-571.
- KLEIN, S. B., LOFTUS, J., KIHLMSTROM, J. F., & ASERON, R. (1989). Effects of item specific and relational information on hypermnnesic recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 1192-1197.
- KUČERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- LATOUR, M. A., & MCKELVIE, S. J. (1994). Effects of type of input and type of processing on recall hypermnnesia. *Canadian Journal of Experimental Psychology*, **48**, 451-459.
- MASSON, M. E. J., & MACLEOD, C. M. (1992). Reenacting the route to interpretation: Enhanced perceptual identification without prior perception. *Journal of Experimental Psychology: General*, **121**, 145-176.
- MCDANIEL, M. A., & EINSTEIN, G. O. (1986). Bizarre imagery as an ef-

- fective memory aid: The importance of distinctiveness. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 54-65.
- MCDANIEL, M. A., MOORE, B. A., & WHITEMAN, H. L. (1998). Dynamic changes in hypermnnesia across early and late tests: A relational/item-specific account. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 173-185.
- MCDANIEL, M. A., & WADDILL, P. J. (1990). Generation effects for context words: Implications for item-specific and multifactor theories. *Journal of Memory & Language*, **29**, 201-211.
- 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.
- MORRIS, P. E., & FRITZ, C. O. (2000). The name game: Using retrieval practice to improve the learning of names. *Journal of Experimental Psychology: Applied*, **6**, 124-129.
- MULLIGAN, N. W. (1999). The effects of perceptual interference at encoding on organization and order: Investigating the roles of item-specific and relational information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 54-69.
- MULLIGAN, N. W. (2000). Perceptual interference at encoding enhances item-specific encoding and disrupts relational encoding: Evidence from multiple recall tests. *Memory & Cognition*, **28**, 539-546.
- MULLIGAN, N. W. (2001). Generation and hypermnnesia. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 436-450.
- MULLIGAN, N. W. (2002a). The emergence of item-specific encoding effects in between-subjects designs: Perceptual interference and multiple recall tests. *Psychonomic Bulletin & Review*, **9**, 375-382.
- MULLIGAN, N. W. (2002b). *Hypermnnesia, total retrieval time and delayed recall: Benefits of multiple recall tests apparent at a delay*. Manuscript submitted for publication.
- MULLIGAN, N. W. (in press). The emergent generation effect and hypermnnesia: Influences of semantic and non-semantic generation tasks. *Journal of Experimental Psychology: Learning, Memory, & Cognition*.
- NAIRNE, J. S., REIGLER, G. L., & SERRA, M. (1991). Dissociative effects of generation on item and order retention. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 702-709.
- OLOFSSON, U. (1997). Win some, lose some: Hypermnnesia for actions reflects increased item-specific processing. *Memory & Cognition*, **25**, 797-800.
- PAYNE, D. G. (1986). Hypermnnesia for pictures and words: Testing the recall level hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 16-29.
- PAYNE, D. G. (1987). Hypermnnesia and reminiscence in recall: A historical and empirical review. *Psychological Bulletin*, **101**, 5-27.
- PAYNE, D. G., & WENGER, M. J. (1992). Repeated recall of pictures, words, and riddles: Increasing subjective organization is not sufficient for producing hypermnnesia. *Bulletin of the Psychonomic Society*, **30**, 407-410.
- ROEDIGER, H. L., III, & CHALLIS, B. H. (1989). Hypermnnesia: Improvements in recall with repeated testing. In C. Izawa (Ed.), *Current issues in cognitive processes: The Tulane Floweree Symposium on Cognition* (pp. 175-199). Hillsdale, NJ: Erlbaum.
- ROEDIGER, H. L., III, & PAYNE, D. G. (1985). Recall criterion does not affect recall level or hypermnnesia: A puzzle for generate/recognize theories. *Memory & Cognition*, **13**, 1-7.
- ROEDIGER, H. L., III, & THORPE, L. A. (1978). The role of recall time in producing hypermnnesia. *Memory & Cognition*, **6**, 296-305.
- ROENKER, D. L., THOMPSON, C. P., & BROWN, S. C. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin*, **76**, 45-48.
- SCHMIDT, S. R., & CHERRY, K. (1989). The negative generation effect: Delineation of a phenomenon. *Memory & Cognition*, **17**, 359-369.
- SERRA, M., & NAIRNE, J. S. (1993). Design controversies and the generation effect: Support for an item-order hypothesis. *Memory & Cognition*, **21**, 34-40.
- 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.
- STEFFENS, M. C., & ERDFELDER, E. (1998). Determinants of positive and negative generation effects in free recall. *Quarterly Journal of Experimental Psychology*, **51A**, 705-733.
- TURTLE, J. W., & YUILLE, J. C. (1994). Lost but not forgotten details: Repeated eyewitness recall leads to reminiscence but not hypermnnesia. *Journal of Applied Psychology*, **79**, 260-271.

NOTES

1. Despite the fact that generation can impair intertarget relational encoding, there are conditions under which the multifactor theory predicts that generation will enhance this type of processing. If the study list is structured so that information from earlier targets is useful in the generation of subsequent targets, intertarget relational processing will be enhanced. For instance, if the target items are all members of a common category and the required cue-target processing does not obscure this relationship (more on this in the introduction to Experiment 2), processing these common characteristics assists generation. Consequently, when structured lists are used, generation effects are often obtained in free recall for between-subjects designs (e.g., deWinstanley & Bjork, 1997; McDaniel et al., 1988). Also, under these conditions, generation leads to higher levels of category clustering than does the read condition, providing additional evidence for enhanced intertarget processing (deWinstanley & Bjork, 1997; McDaniel et al., 1998; McDaniel & Waddill, 1990). With unrelated targets, intertarget processing is not useful and is neglected in favor of item-specific and cue-target relational processing.

2. There are two other relevant studies that produced results similar to those of Mulligan (2001). Both studies (Erdelyi, Buschke, & Finkelstein, 1977; Payne & Wenger, 1992) used a between-subjects design and the same atypical manipulation of read versus generate. In both studies, a list of target items was presented as words (e.g., *bat*), as pictures (e.g., a sketch of a bat), or as riddles whose answers were the target words (although the words themselves were not presented; e.g., "This longish wooden object is used by baseball players to hit the ball. What is it?"). In both studies, hypermnnesia occurred for both the picture and the riddle conditions, but not for the word condition. Conceiving of the riddle and the word conditions as generate and read conditions, respectively, the results were consistent with Mulligan's (2001) finding that generation produced hypermnnesia, whereas reading did not. In addition, and again consistent with Mulligan (2001), recall in the riddle and the word conditions did not differ on the initial tests, but the riddle condition appears to have produced higher recall on later tests (statistical comparisons within tests were not conducted). Neither Erdelyi et al.'s nor Payne and Wenger's study was designed to isolate the effects of generation proper. For example, the presence of the riddle at encoding was confounded with encoding condition, appearing in the riddle condition, but not in the word condition. In standard generation manipulations, when words are generated from meaningful cues, the cues are presented in both the generate and the read conditions. This necessitated Mulligan's (2001) study, which, using a more standard generation manipulation, produced consistent results.

3. A reviewer raised the point that the "read" condition is actually a "copy" condition. Many studies of the generation effect, including the present experiment, required subjects to write down the items in the "read" condition (as well as items in the generate condition; e.g., Burns, 1990, 1992; Hirshman & Bjork, 1988; Kinoshita, 1989; McDaniel et al., 1988; Mulligan, 2001). The traditional labels (*read* vs. *generate*) are used in the present study, both to maintain consistency with the prior literature and because the traditional labels denote the operative difference between the conditions.