Journal of Memory and Language 92 (2017) 202–216

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

The generation effect revisited: Fewer generation constraints enhances item and context memory

Matthew P. McCurdy, Ryan C. Leach, Eric D. Leshikar *

University of Illinois at Chicago, Department of Psychology, 1007 W Harrison St (M/C 285), Chicago, IL 60607, United States

ARTICLE INFO

Article history:
Received 2 January 2016
revision received 8 June 2016
Available online 15 July 2016

Keywords:
Generation effect
Generation constraint
Item memory
Context memory
Recognition
Recall

ABSTRACT

Self-generated information is often better remembered than non-self-generated information. This effect has been robust for item memory (i.e., the content of information) across many different experiments, but inconsistent for context memory (e.g., memory for the extraneous details of information, such as source). Previous studies examining the generation effect, however, have often applied constraints on the generation task possibly limiting the memory benefit from self-generation. In three experiments, we compared item and context memory for a lower-constraint generation task (i.e., free response to a cue word) relative to higher-constraint generation tasks (Exp. 1 & 2: scramble; Exp. 3: word fragment). Results showed that participants had better item and context memory in the lower-constraint compared to higher-constraint generation tasks. Overall, these experiments suggest that that the mnemonic benefits of self-generation depend on the level of task constraint. This study further advances the idea that self-generation is a powerful mnemonic that leads to enriched memory representations for both the item and context, especially when fewer generation constraints are imposed.

Introduction

Self-generated material is better remembered than information that is simply read, or otherwise not generated by the self (Slamecka & Graf, 1978). Since this seminal study, the generation effect has been widely replicated (see Bertsch, Pesta, Wiscott, & McDaniel, 2007, for a review). However, in the majority of these studies there are experimental constraints limiting what participants can produce. As an example, in the typical *generate* condition prior studies have used highly-constrained tasks such as unscrambling a word (e.g., open - coesl; Foley & Foley, 2007; Geghman & Multhaup, 2004), or completing a word fragment (e.g., open - cl∗s∗; Clark, 1995; Hirshman & Bjork, 1988), where there is a single "correct" answer to be generated. Memory for generated items is then compared to a control condition where participants read or listen to experimenter provided material (e.g., open – close). Although this paradigm has been informative in understanding the generation effect, little attention has been given to how differences in the level of constraint provided by different generation tasks might influence memory. Therefore, the aim of this study is to assess whether the magnitude of the generation effect on memory differs as a function of generation constraints (lower versus higher). Specifically, we aim to compare memory performance for materials generated in a "lower-constraint" task, that allows participants to generate any word that comes to mind in response to a cue word (i.e., open - ______), relative to some commonly used "higher-constraint" generation tasks (i.e., scramble/word fragment).

Prior work examining the generation effect has often focused on item memory (i.e., memory for the content of information), and this work has consistently shown robust item memory improvements for self-generated versus

∗ Corresponding author.
E-mail address: leshikar@uic.edu (E.D. Leshikar).

http://dx.doi.org/10.1016/j.jml.2016.06.007
0749-596X/

non-self-generated materials. These effects have been shown across a variety of generation tasks (Bertsch et al., 2007), as well as for both recognition (Graf, 1982; McElroy & Slamecka, 1982; McFarland, Frey, & Rhodes, 1980; Slamecka & Graf, 1978; Slamecka & Katsaiti, 1987) and cued recall memory procedures (Donaldson & Bass, 1980; Jacoby, 1978; Johnson, Raye, Foley, & Foley, 1981). Other work on the generation effect has examined memory for extraneous details of an episode (i.e., context memory), but the results of this work have been mixed. Some have found context memory benefits for generated materials relative to control (Marsh, 2006; Marsh, Edelman, & Bower, 2001), others have found poorer context memory (Mulligan, 2004; Mulligan, Lozito, & Rosner, 2006), while others have found no difference between the two (Mulligan, 2011). Considering item and context memory effects together, researchers have advanced two ways that self-generation may promote memory: The *item-context enhancement* account suggests that encoding for item and context share the same encoding processes, therefore any manipulation that enhances memory for one type should also extend to the other (Geghman & Multhaup, 2004; Marsh, 2006; Marsh et al., 2001). Thus, when item memory improves, context memory should improve as well. In contrast, the *item-context tradeoff* account suggests that encoding item and contextual details requires different processes that may compete for limited cognitive resources at the time of study (Jurica & Shimamura, 1999; Nieznański, 2011). That is, if self-generation recruits more resources for item memory encoding, fewer resources are available to encode peripheral contextual details leading to poorer memory. The present study aims to provide clarity on these accounts.

This study is designed to explore the use of a lower-constraint generation task, where participants are allowed to respond freely to a cue word. To our knowledge, only one study has examined item and context memory using a similar lower-constraint task (Jurica & Shimamura, 1999). Jurica and Shimamura (1999) tested item and context memory in a conversational setting where participants were free to answer questions with any response and found an increase in item memory, but a concomitant decrease in context memory for self-generated material compared to control, consistent with the *item-context trade-off* account. However, this study did not include a manipulation of generation constraint to test memory differences between tasks of differing generation constraint (i.e., lower-constraint versus higher-constraint) precluding any conclusions about how the level of constraint might influence the magnitude of the generation effect. Across three experiments, we aimed to compare memory benefits of a lower-constraint generation task against two commonly used higher-constraint generation tasks (Exp. 1 & 2: scramble; Exp. 3: word fragment) and a read control, as measured by item recognition and cued recall, as well as context recognition.

It is worth noting that in prior work, researchers investigating the generation effect have not typically used a task with lower constraints because of the potential loss of control over the to-be-generated materials, and possible item-selection effects (Hirshman & Mulligan, 1991). The logic behind this concern is that words generated freely from a cue word (e.g., open - _____) may be somehow different, and thus more memorable, than words produced in a higher-constraint procedure where there is an expected, correct response (e.g., open – cosle, restricts participants to generate "close"). Specifically, it is possible that generated words that are not the expected target in the lower-constraint task (e.g., producing open - "door", instead of open - "close") may be more memorable than the expected, normed target word. In this study, we took several steps to limit this concern: First, we used a counterbalancing procedure such that each word pair occurs in all three tasks (lower-constraint, higher-constraint, read control) across participants. This counterbalancing procedure allowed us to conduct an item-analysis where we compared memory performance for each word pair when it occurred in the lower-constraint, higher-constraint, and read tasks. Specifically, in this item-analysis we removed any non-normed target responses for the lower-constraint task. For example if a participant produced "door" in response to "open - ____" instead of "close" as was normed, this trial was removed in the item-analysis. Thus, since the word is the same across all tasks, this reduces the possibility that the memory effects in this item-analysis are due to idiosyncratic differences in the to-be-remembered information. We conducted this item-analysis in addition to our primary analyses to address the possibility that item-selection effects may be confounding our primary findings, and report them after the primary analysis in the results of each experiment.

Given that work on the generation effect has shown that self-generated materials are better remembered than non-self-generated materials, it logically follows that tasks that afford more freedom to produce materials (i.e., more "generation") should lead to bigger memory effects. Therefore, we predict a larger memory benefit for items produced in a lower-constraint versus a higher-constraint generation task. Larger item memory effects in the lower-constraint task would suggest that item memory effects are reduced in generation tasks that are more highly constrained. It would also suggest that the level of generation constraint should be taken into account when assessing these memory benefits. In addition, we expect improved item memory for both generation tasks relative to a non-generate (read) control condition consistent with the typical generation effect (Jacoby, 1978; Slamecka & Graf, 1978).

We see two possible outcomes for our context memory measure. First, if the *item-context enhancement* account is correct, we expect to see more accurate context memory judgments in the lower-constraint relative to the higher-constraint and read control conditions, which would suggest that the generation effect is a robust mnemonic effect that improves memory for the item and context. Alternatively, the accuracy of context memory judgments may be reduced when comparing generate tasks with a read control task, which would support the *item-context tradeoff* account. This finding would suggest that the act of generation might only enhance encoding of the item, leaving fewer resources to encode contextual details. If in fact generation improves item memory and reduces context memory, and we find that the lower-constraint generation task provides larger item memory effects compared to the higher-constraint task, we may expect to see lower accuracy for context memory judgments in the lower-constraint compared to the higher-constraint task.

Finally, although the generation effect is robust across a variety of conditions, some prior work suggests that magnitude of the memory effect can be influenced by the "difficulty" of the generate task (Fiedler, Lachnit, Fay, & Krug, 1992; Foley, Foley, Wilder, & Rusch, 1989; Nieznański, 2011; Taconnat, Froger, Sacher, & Isingrini, 2008; Tyler, Hertel, McCallum, & Ellis, 1979). One way to introduce difficulty is by manipulating the strength of association between cue and target words (as examples, high associate: hot-cold; low associate: glass-window). The principle behind this effect is that some words, like hot, have a single strong associate (cold) whereas other words have multiple weak associates (e.g., glass – break; window; cup, etc.). Therefore, generating a target word from a high-associate cue (e.g., hot), where there is likely to be a strong associate easily produced, is often considered an easier task than generating a target from a low-associate cue (e.g., glass) where multiple words come to mind. Studies that have manipulated difficulty in this way have found that, although high-associates tend to be remembered better overall, the generation effect (i.e., the difference between self-generation and control) for low-associates is larger than high-associates, suggesting greater memory benefits for more difficult tasks (Hirshman & Bjork, 1988; Nieznański, 2011; Taconnat et al., 2008). In this way, the strength of association for the word pair could be considered another manipulation of constraint on the generation task. Specifically, low-associate pairs would be considered less constrained since there are multiple associates that may be brought to mind, whereas high-associates are more constrained since only one strong associate is often brought to mind. This idea is consistent with the idea that fewer constraints may lead to larger memory benefits from self-generation. To this point, in Experiment 1 we include a difficulty manipulation to better understand the memory benefits from self-generation.

Experiment 1

Experiment 1 compared a lower-constraint generation task to a previously used higher-constraint generation task (scramble) and a non-generate (read) control to determine the influence that the level of constraint has on the magnitude of the generation effect. In this experiment, we compared these three tasks on item and context recognition tests for the *cue* words (provided by the experimenter), and we also tested memory for the *target* words (generated by the participant) in a cued recall test. Additionally, in this experiment we used both high- and low-associate word pairs to assess how word association difficulty might also be considered a generation constraint.

Method

Participants

Twenty-four adults (age: 19.9, *SD*: 2.0, range: 18–25, 13 females) were recruited from the University of Illinois at Chicago introductory psychology course subject pool and local community to participate. Participants gave written informed consent in accordance with the University of Illinois at Chicago Institutional Review Board and were compensated with course credit or paid for their participation. All participants were fluent in reading, speaking, and understanding English. Two participants were excluded from the analysis because they did not understand the recognition procedure instructions and showed a 100% false alarm rate on item recognition, leaving a total *N* of 22 for this experiment.

Stimuli

A total of 96 unique cue-target word pairs selected from the University of South Florida Free Association Norms (Nelson, McEvoy, & Schreiber, 1998) were used. Word pairs were selected based on their forward cue-to-target association strength (FSG) so that half of the pairs were low-associates (e.g., glass - window; Mean FSG = .150) and half of the pairs were high-associates (e.g., hot - cold; Mean FSG = .626) to examine effects of association strength on memory. Cues in "high-associate" pairs had a single highly associated target based on the norms, whereas cues in "low-associate" pairs did not. Word pairs were selected so that both cue and target words were between 4–7 letters to keep word length constant across conditions. Across participants, each pair was counterbalanced so that it occurred in every task (generate, scramble, read) and as a new item to control for possible differences between stimuli across the tasks and to allow for our item-analysis described earlier.

Procedure

Experiment 1 consisted of three phases: Encoding, recognition test of *cue* words, cued recall test of *target* words. The encoding and recognition tests were presented on a monitor using E-prime presentation software. Before encoding, participants were trained on shortened versions of the encoding and recognition test phases in white font on a black background. Participants were not informed of the cued recall phase of the target words as done before (Jacoby, 1978; Watkins & Sechler, 1988). In the encoding phase, participants were presented with 72 word pairs across the three task conditions: generate, scramble, read (24 pairs in each task, 12 high-associates and 12 low-associates). In the "generate" task (lower-constraint), participants were shown a cue word followed by a blank line (e.g., brief – _____) and were instructed to generate a word related to the cue word and then say both words out-loud (i.e., the cue and the generated target). The generate task instructions emphasized that the responses were subjective and that there were no correct answers, so that participants were free to generate any word that came to mind. Participant responses during this phase were recorded by the experimenter. In the "scramble" task (higher-constraint), participants were shown a cue word followed by a scrambled target word (e.g., blaze – feri) and instructed to unscramble the word and then say both words out-loud. In the "scramble" task, the first letter of the word was always in its proper place, and the following letters were randomized as done previously (Geghman &

Multhaup, 2004). Participants were not explicitly told that the first letter was always correct for this task, but they did receive practice for the encoding phase where this procedure was also implemented. In the "read" task, participants were shown a cue and target word (e.g., done – finish) and were instructed to read both words out-loud. For all encoding tasks, we chose to have participants say both words out-loud to ensure that the participant was processing the cue word in all tasks. Once the participant said both words out-loud for that trial, the experimenter advanced the screen and the participant was presented with a new word pair after a 500 ms fixation[1] (see Fig. 1). Encoding trials were presented in six blocks of 12 trials per task (6 high-associates, 6 low-associates), and all trials within the same block used the same instructions (generate, scramble, read). The order of the blocks was randomized for each participant with the caveat that no task (generate, scramble, read) was performed back-to-back. Before each block, an instruction prompt appeared for 3000 ms indicating the next task (e.g., "Get ready to do the generate/scramble/read task").

Immediately following the encoding phase, participants were given a recognition test for the experimenter-provided *cue* words they saw during encoding. Participants were shown all 72 cue words from the encoding phase and 24 new words in random order, for a total of 96 recognition trials. Each recognition trial consisted of two self-paced recognition decisions, which mapped onto item and context memory judgments, respectively (see Fig. 1). First, a cue word was presented on the screen and participants judged whether the word was old, new, or whether they didn't know, using their index, middle, and ring finger, respectively, of their right hand. This response served as the item memory measure. Second, following a 500 ms fixation, participants were shown the same cue word and judged whether they saw that cue in the generate, scramble, or read condition, or whether they didn't know using their index, middle, ring, and pinky finger of their right hand, respectively. This response served as the context memory response. For items that were recognized as "new" in the item memory judgment, participants were instructed to make a "don't know" response for this second decision, a procedure used in previous studies (Leshikar & Duarte, 2012, 2014; Leshikar, Dulas, & Duarte, 2015; Leshikar & Gutchess, 2015). The "don't know" response was included as an option to reduce response contamination by guessing (Duarte, Henson, & Graham, 2008; Duarte, Henson, Knight, Emery, & Graham, 2010).

Following the recognition test, participants were given a surprise cued recall test for the *target* words they produced from the encoding phase. Participants were given a pencil and paper that had a list of the 72 cue words seen during encoding (in a randomized order) followed by a blank line. Participants were instructed to write the word that was paired with the cue word at encoding and were reminded that the word could be one they encountered in the generate, scramble, or read task. If they were unsure or could not remember, they were instructed to leave the line blank to reduce the likelihood of guessing.

Results

In Experiment 1, participants gave responses on 100% of the generate and read trials, and successfully unscrambled words 92% of time in the scramble condition. Trials where participants were unable to unscramble the target word successfully were removed from all analyses. The raw recognition responses are presented in Table 1. A 3 (task: generate, scramble, read) $\times$ 2 (association strength: high, low) repeated-measures analysis of variance (ANOVA) was conducted separately for each memory test (item, context, and cued recall). Item recognition scores were calculated by the percent of items seen at study correctly responded to as "old" for each task. Context recognition scores were calculated using the conditional source identification measure (CSIM; Murnane & Bayen, 1996), a procedure that removes the influence of item recognition on source memory performance. Using this measure, our context memory scores are represented by the proportion of correctly recognized items (i.e., item hits) that are also correctly identified with their source (generate, scramble, read).[2] Cued recall scores were calculated as the percent of correctly recalled target words out of all words seen in the encoding phase, and items that were incorrect or left blank were counted as misses. Data for these analyses are graphed in Fig. 2. Additionally, we performed an item analysis that examines memory performance for each word pair as it occurs in each of the three tasks to control for possible item-selection confounds of our subject-based analyses.

Subject-based results

Results of the item recognition analysis revealed a main effect for task, $F(2,20) = 41.40$ $MSE = 1.08$, $p < .001$, $\eta_p^2 = .663$. Planned follow-up comparisons showed that item recognition was better for generate than scramble,[3] $t(21) = 9.01$, $p < .001$, and better for scramble than read, $t(21) = 4.28$, $p < .001$. The analysis also revealed a main effect

[1] Since the encoding phase was self-paced, it could be argued that the amount of time spent at encoding is influencing the memory effects. We have included a supplemental analysis that compared the reaction time data at encoding for each task. For all experiments, this analysis showed significant differences between the generate tasks (lower- and higher-constraint) and read task, but not between the lower- and higher-constraint generation tasks themselves. The data for these analyses are reported in the supplemental materials.

[2] Although the data reported here is the raw "hits", we also performed analyses on corrected recognition scores. Specifically, for item recognition we used "hits" minus total false alarms ("new" words judged to be "old"), and in context recognition we used context "hits" minus context false alarms (e.g., new words falsely identified as occurring in a study task). Performing ANOVAs on these corrected recognition measures fully replicated the results from an analysis of the hits; therefore, we report the raw "hits" as the primary analyses. This is true for all three experiments.

[3] One way to address concerns that the words generated in the unconstrained task were somehow different than the normed words shown in the scramble and read task (item-selection confound) is to perform an item-analysis as we mention in the introduction. Another way to control for item-selection confounds is to separate the free responses from the lower-constraint task into "novel" (words that were different from our expected target) and "matched" (words that matched our expected, normed target). When analyzing the data using only words in the lower-constraint task that were "matched" we found identical significance as the overall analysis (not reported). An analogous analysis was conducted for Experiments 2 and 3, which also supported the main findings of those experiments.

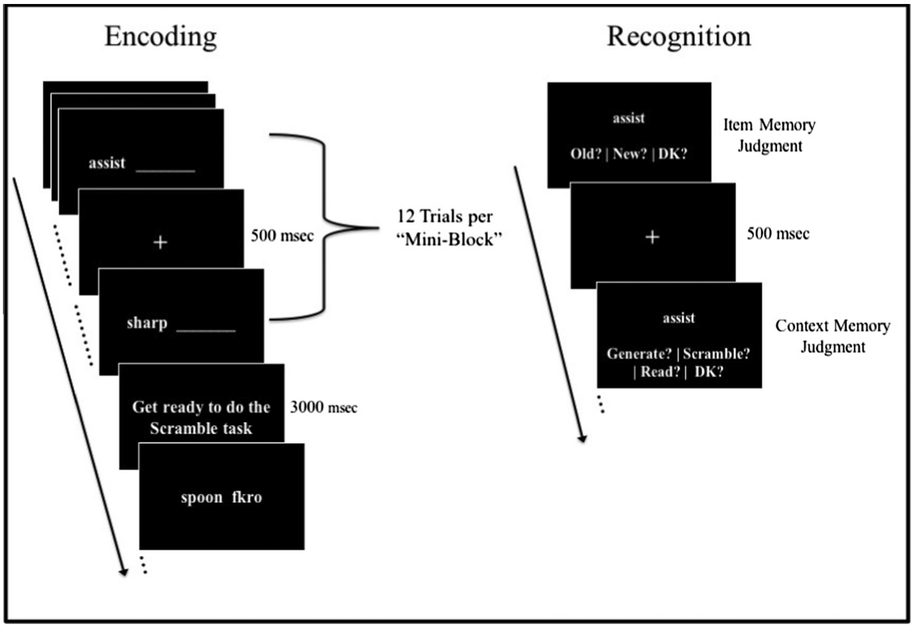

Fig. 1. Trial schematic for the encoding and recognition phases for Experiment 1. Note: For Experiments 2 and 3 the same procedure was used with the exception that the recognition phase tested participants on the *target* word (e.g., fork), instead of the cue word (e.g., spoon).

for association type, $F(1,21) = 13.64$, $MSE = .121$, $p = .001$, $\eta_p^2 = .394$, which was driven by better recognition memory for the high- than low-associates consistent with previous studies (Hirshman & Bjork, 1988; Taconnat et al., 2008). The task by association strength interaction was also significant, $F(2,20) = 4.15$, $MSE = .053$, $p = .023$, $\eta_p^2 = .165$. Planned follow-up analyses revealed that high-associates were recognized more accurately than low-associates, for the scramble, $t(21) = 3.09$, $p = .006$, and read tasks, $t(21) = 2.39$, $p = .026$, but not for the generate task, $t(21) = -1.00$, $p = .329$.

Results of the context recognition analysis revealed a main effect for task, $F(2,20) = 11.90$, $MSE = .517$, $p < .001$, $\eta_p^2 = .362$. Planned follow-up analyses showed this effect was driven by better memory for words in the generate task compared to both the scramble and read tasks, $ts > 7.39$, $ps < .001$. Interestingly, recognition did not differ between scramble and read tasks, $t(21) = -1.19$, $p = .25$. There was no main effect for association type, $F(1,21) = 0.00$, $MSE = 0.00$, $p = .99$, $\eta_p^2 = 0.00$, and no task by associate type interaction, $F(2,20) = 0.02$, $MSE = .001$, $p = .98$, $\eta_p^2 = .001$ in this analysis.

Cued recall results revealed a main effect for task, $F(2,20) = 60.13$, $MSE = 2.03$, $p < .001$, $\eta_p^2 = .741$, which reflected better memory in the generate than the scramble condition, $t(21) = 6.08$ $p < .001$, and better memory in the scramble than read task, $t(21) = 5.39$, $p < .001$. A main effect for association type was also evident, $F(1,21) = 50.83$, $MSE = 1.36$, $p < .001$, $\eta_p^2 = .708$, where high-associates were recalled more accurately than low-associates. The task by association strength interaction was also significant, $F(2,20) = 6.86$, $MSE = .256$, $p = .003$, $\eta_p^2 = .246$ (see Fig. 3). Planned follow-up analyses revealed that, similar to item recognition, the high-associates were recalled more accurately than the low-associates for the scramble $t(21) = 4.60$, $p < .001$, and read, $t(21) = 4.81$, $p < .001$, tasks, but not for the generate task.

To further investigate the task by association type interactions we saw in our ANOVAs and to determine whether the magnitude of the generation effect (i.e., the difference between self-generation and controls) differed for high- and low-associates, we performed a separate 2 (generation task) by 2 (association type) repeated measures ANOVA on the difference scores of the generation tasks to the read control (e.g., generate – read controls v. scramble – read controls). These difference scores represent an estimate of the "generation effect" (the amount of benefit from self-generation over controls) for each generation task. We conducted this separate analysis to determine whether low-associate word pairs (which can be thought of as another type of generation constraint) also led to a greater generation effect compared to high-associates. Since we only saw an interaction for task by association type for both item recognition and cued recall in our original ANOVAs, this analysis was only performed on the difference scores from these two measures. Data for these analyses are graphed in Fig. 4.

Results of this analysis for item memory revealed a main effect for task, $F(1,21) = 81.25$, $MSE = .493$, $p < .001$, $\eta_p^2 = .795$, which was driven by a larger generation effect for the generate task ($M_{diff} = .314$, $SD = .202$) compared to the scramble task ($M_{diff} = .165$, $SD = .183$). There was no main effect for association type, $F(1,21) = 1.12$, $MSE = .058$, $p = .30$, $\eta_p^2 = .051$. The task by association type interaction was significant, $F(1,21) = 10.45$, $MSE = .086$, $p = .004$, $\eta_p^2 = .332$. Planned follow-up analyses revealed that for the generate task, the generation effect was significantly greater when using low-associate word pairs

Table 1
Means and standard deviations (in parentheses) of responses for item and context recognition, and accuracy for cued recall by encoding condition.

Experiment 1: All words (high- and low-associates combined)

Item recognition				Context recognition					Cued recall		
Task	Old	New	Don't Know	Task	Generate	Scramble	Read	Don't Know	Generate	Scramble	Read
Generate	.91 (.09)	.07 (.07)	.02 (.04)	Generate	.89 (.13)	.03 (.06)	.05 (.10)	.03 (.04)	.88 (.10)	.76 (.18)	.46 (.19)
Scramble	.76 (.12)	.18 (.10)	.06 (.09)	Scramble	.09 (.10)	.68 (.20)	.21 (.14)	.02 (.04)			
Read	.60 (.22)	.31 (.19)	.09 (.13)	Read	.10 (.15)	.10 (.09)	.74 (.22)	.06 (.12)			
New	.17 (.23)	.70 (.31)	.13 (.22)	New	.03 (.06)	.05 (.11)	.09 (.18)	.83 (.27)			

Experiment 1: High-associates

Item recognition				Context recognition					Cued recall		
Task	Old	New	Don't Know	Task	Generate	Scramble	Read	Don't Know	Generate	Scramble	Read
Generate	.90 (.08)	.08 (.08)	.02 (.04)	Generate	.89 (.13)	.04 (.06)	.05 (.10)	.02 (.03)	.90 (.11)	.83 (.18)	.62 (.24)
Scramble	.81 (.11)	.14 (.10)	.05 (.08)	Scramble	.10 (.13)	.68 (.23)	.21 (.14)	.01 (.03)			
Read	.64 (.22)	.27 (.18)	.09 (.17)	Read	.10 (.16)	.10 (.15)	.75 (.25)	.05 (.12)			
New	.17 (.22)	.71 (.31)	.12 (.21)	New	.03 (.07)	.05 (.13)	.09 (.18)	.83 (.29)			

Experiment 1: Low-associates

Item recognition				Context recognition					Cued recall		
Task	Old	New	Don't Know	Task	Generate	Scramble	Read	Don't Know	Generate	Scramble	Read
Generate	.92 (.11)	.06 (.08)	.02 (.05)	Generate	.89 (.15)	.03 (.08)	.05 (.12)	.03 (.06)	.87 (.17)	.57 (.27)	.30 (.25)
Scramble	.71 (.17)	.23 (.14)	.06 (.11)	Scramble	.08 (.12)	.67 (.23)	.22 (.17)	.03 (.06)			
Read	.55 (.25)	.36 (.23)	.09 (.11)	Read	.06 (.15)	.11 (.22)	.76 (.29)	.07 (.14)			
New	.17 (.26)	.69 (.32)	.14 (.24)	New	.03 (.08)	.05 (.12)	.10 (.19)	.82 (.27)			

Experiment 2

Item recognition				Context recognition					Cued recall		
Task	Old	New	Don't Know	Task	Generate	Scramble	Read	Don't Know	Generate	Scramble	Read
Generate	.84 (.21)	.13 (.19)	.03 (.08)	Generate	.80 (.20)	.07 (.06)	.08 (.07)	.05 (.18)	.95 (.06)	.91 (.10)	.81 (.16)
Scramble	.86 (.18)	.11 (.14)	.03 (.09)	Scramble	.09 (.09)	.68 (.22)	.17 (.14)	.06 (.18)			
Read	.62 (.24)	.29 (.22)	.09 (.15)	Read	.12 (.17)	.16 (.12)	.63 (.24)	.09 (.22)			
New	.26 (.27)	.60 (.31)	.14 (.24)	New	.04 (.05)	.08 (.11)	.13 (.12)	.75 (.22)			

Experiment 3

Item recognition				Context recognition					Cued recall		
Task	Old	New	Don't Know	Task	Generate	Fragment	Read	Don't Know	Generate	Fragment	Read
Generate	.80 (.22)	.16 (.18)	.04 (.08)	Generate	.79 (.16)	.11 (.09)	.09 (.09)	.01 (.02)	.92 (.08)	.88 (.10)	.75 (.18)
Fragment	.84 (.13)	.12 (.12)	.04 (.07)	Fragment	.13 (.12)	.66 (.16)	.19 (.13)	.02 (.04)			
Read	.60 (.16)	.30 (.15)	.10 (.18)	Read	.12 (.15)	.13 (.11)	.71 (.20)	.04 (.06)			
New	.14 (.12)	.70 (.30)	.16 (.30)	New	.04 (.07)	.04 (.08)	.08 (.11)	.84 (.20)			

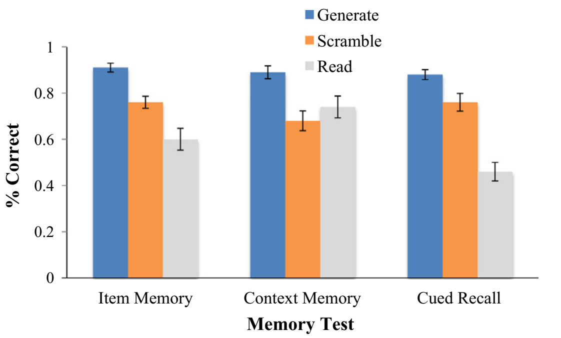

Fig. 2. Experiment 1: Accuracy by task for item recognition, context recognition, and cued recall. Error bars represent standard errors of the mean. Note: For item memory and cued recall all differences between conditions (generate, scramble, read) are significant at $p < .001$. For context memory, the difference between scramble and read was not significant.

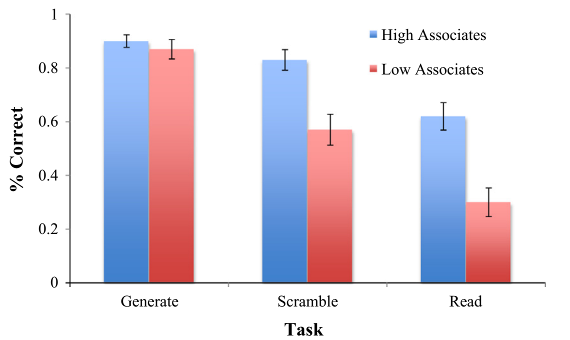

Fig. 3. Experiment 1: Cued recall accuracy by task and association strength. Error bars represent standard errors of the mean. Note: Differences in Generate are not significant. Differences in scramble and read are significant at $p < .001$.

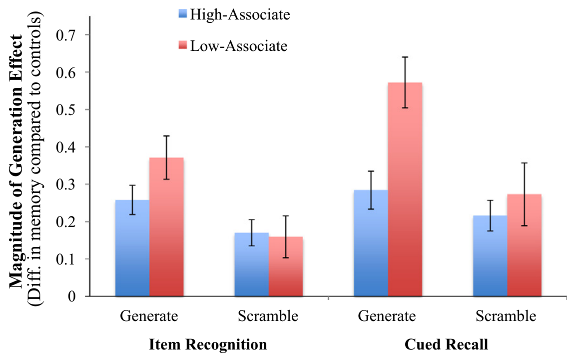

Fig. 4. Experiment 1: Graph of difference scores between generation tasks and read controls by high- and low-associates for item recognition and cued recall. Note: Differences in Generate for both are significant at $p < .05$. Differences in Scramble for both are not significant.

compared to high-associate word pairs, $t(21) = 2.25$, $p = .04$, however, for the scramble task there was no difference, $t(21) = 0.21$, $p = .83$.

For cued recall, this analysis again revealed a significant main effect for task, $F(1,21) = 37.52$, $MSE = .743$, $p < .001$, $\eta_p^2 = .641$, which was driven by a greater generation effect for the generate task ($M_{\text{diff}} = .428$, $SD = .197$) compared to the scramble task ($M_{\text{diff}} = .244$, $SD = .206$). The main effect for association type was not significant, $F(1,21) = 3.86$, $MSE = .653$, $p = .06$, $\eta_p^2 = .155$, but was trending in the direction that overall low-associates ($M_{\text{diff}} = .422$, $SD = .342$) provided a marginally greater generation effect compared to high-associates ($M_{\text{diff}} = .250$, $SD = .197$). The task by association type was significant, $F(1,21) = 16.30$, $MSE = .294$, $p = .001$, $\eta_p^2 = .437$. Planned follow-up comparisons again showed that for the generate task, the generation effect was greater for low-associates compared to high-associates, $t(21) = 3.33$, $p = .003$, but for the scramble task there was no difference, $t(21) = 0.58$, $p = .57$.

Item analysis

To address the concerns of item-selection effects in this experiment, we conducted a 3 (task) by 2 (association type) ANOVA to examine the item and context recognition memory for the cue word of each *word pair* dependent on the encoding condition (generate, scramble, read). Results of both item and context recognition ANOVAs in this analysis provided similar, but not identical results to the between-subjects ANOVAs.

For item recognition, there was a significant main effect for task, $F(2,46) = 102.09$, $MSE = 2.02$, $p < .001$, $\eta_p^2 = .685$. Planned paired-samples *t*-tests revealed higher item recognition between the generate compared to scramble, $t(95) = 7.63$, $p < .001$, and scramble task compared to read $t(95) = 6.36$, $p < .001$. The main effect for association type was no longer significant, but still marginally significant, $F(1,47) = 3.45$, $MSE = .213$, $p = .07$, $\eta_p^2 = .068$. The task by association strength interaction was again significant for item memory, $F(2,94) = 3.89$, $MSE = .094$, $p = .024$, $\eta_p^2 = .076$. Planned follow-up analyses revealed a similar pattern as the between-subjects analyses with the exception of words in the read task. Specifically in this analysis, high-associates were better remembered than low-associates for the scramble, $t(47) = 2.34$, $p = .023$, but not for the generate, $t(47) = -0.67$, $p = .51$, or read tasks, $t(47) = 1.78$, $p = .08$.

For context memory, our results fully replicated the between subjects ANOVAs. We found a significant main effect for task, $F(2,94) = 31.94$, $MSE = 1.29$, $p < .001$, $\eta_p^2 = .405$, which planned paired-samples *t*-tests revealed that context recognition was greater in the generate task compared to both the scramble and read tasks, $ts > 6.47$, $ps < .001$, and again scramble provided no context memory benefit compared to read, $t(95) = -0.61$, $p = .54$. As in the between-subjects analysis, there was no main effect for association type, $F(1,47) = 0.66$, $MSE = .032$, $p = .42$, $\eta_p^2 = .014$, and no task by association type interaction, $F(2,94) = .374$, $MSE = .016$, $p = .69$, $\eta_p^2 = .008$.

Experiment 1 Discussion

Overall, the results of Experiment 1 showed that in nearly all instances, the lower-constraint generation task (generate) led to greater memory improvements compared to the higher-constraint generation task (scramble). These findings suggest that the level of constraint (or "amount" of generation) can influence the magnitude of the memory benefits from self-generation. Specifically, this evidence shows that fewer constraints on the to-be-generated information may improve memory to a greater extent than generation tasks that place higher constraints on what participants can produce. Importantly, the item recognition results replicated prior work showing that both the lower- and higher-constraint generation tasks improved memory relative to a read condition (Foley & Foley, 2007; Foley et al., 1989; Geghman & Multhaup, 2004), but also showed that the lower-constraint task provided an even greater item memory benefit compared to the higher-constraint task.

Prior work on context memory recognition has shown mixed results, with some studies showing a generation effect for both item and context memory (Geghman & Multhaup, 2004; Marsh, 2006; Marsh et al., 2001), and others showing no benefit for context memory despite seeing a benefit for item memory (Jurica & Shimamura, 1999). In this experiment, the higher-constraint task showed no differences in our context memory measure compared to a read control. Importantly, however, there was a context memory benefit for items produced in the lower-constraint task relative to both the higher-constraint and control tasks. If this were a replication study using only a higher-constraint task and read control, no generation effect for context memory would have been evident. However, using a lower-constraint task elicited strong evidence of a generation effect for context memory. Because we found both item and context memory benefits, these results provide support for the *item-context enhancement* account, suggesting that the generation effect is a powerful mnemonic strategy that enhances both item and context memory.

Previous work has shown a robust generation effect for cued recall (Donaldson & Bass, 1980; Graf, 1980; Johnson et al., 1981). Our cued recall analysis replicated this work and showed a similar pattern as our item and context recognition effects, with highest recall for words in the lower-constraint task compared to both the higher-constraint and read tasks. However, these results came with the caveat that when analyzing only high-associate word pairs, there was no longer a significant difference between the lower-constraint and higher-constraint tasks. This suggests that a lower-constraint generation task may only benefit memory for targets to a greater extent than higher-constraint tasks when low-associate word pairs (or more difficult tasks) are used.

We included a difficulty manipulation in our study to determine how the influence of level of constraint on generation tasks might change as a function of difficulty. Prior work has shown that easy generation trials lead to improved memory relative to difficult trials, but that the generation effect (i.e., the difference between the generation task and control) is larger for more difficult tasks (Hirshman & Bjork, 1988; Taconnat et al., 2008). Our results for item memory (recognition and cued recall) support this idea and expand on this prior work by showing that low-associate word pairs provide a greater memory advantage over controls when they were presented in the lower-constraint task, compared to the higher-constraint task. This suggests that difficulty, as evaluated by word association strength, may also be considered a manipulation of constraint, where low-associates are less constrained, and therefore show greater memory benefits from self-generation compared to high-associate word pairs where there is often only a single, strong associate brought to mind. Overall, these findings provide more evidence that the level of constraint is an important consideration when assessing the magnitude of the generation effect.

It is evident that the memory effects shown in Experiment 1 were not always as robust for high-associates compared to low-associates. To further test memory differences between the lower-constraint and higher-constraint generation tasks, we elected to use only high-associates in Experiment 2. Since the memory effects were reduced for high-associate words, using these stimuli is a more conservative test for memory differences between generation tasks. To replicate these effects in additional experiments using only the high-associates would further bolster the claim that generation effects are larger in lower-constraint versus higher-constraint tasks.

One limitation of Experiment 1 was that the recognition tests assessed memory for the *cue* words, and not the *target* words (the word the participant produced). Although it is important to note that cued recall, which tested participants' memory for *target* words, showed similar patterns of results as the recognition test for item and context memory, it was important to replicate these findings in a second experiment where we tested recognition memory for the participant-produced *target* words.

Thus, two changes were made from Experiment 1 to Experiment 2. First, we tested item and context recognition for *target* words and second, we did so using only high-associate word pairs.

Experiment 2

Experiment 1 supported the idea that a lower-constraint generation task leads to larger memory effects compared to both the higher-constraint and control tasks. In Experiment 2, we tested item and context recognition of the *target* words as opposed to the cue words, and used only high-associate word pairs as stimuli. In this experiment we expected similar results to those of Experiment 1, higher item and context memory for the lower- relative to the higher-constraint task.

Method

Participants

Twenty-four adults (age: 19.0, *SD*: 1.0, range: 18–23, 18 females) were recruited from the University of Illinois at Chicago introductory psychology course subject pool and community solicitation, and gave their informed consent and were paid or given course credit as in Experiment 1. In this experiment, all participants reported high fluency in reading, speaking, and understanding English. One participant was excluded from the analysis because they responded "old" to all item memory judgments, indicating a 100% false alarm rate, leaving a total *N* of 23.

Stimuli

A total of 96 cue-target word pairs were used in Experiment 2. We included the 48 high-associate word pairs from Experiment 1 and added an additional 48 word pairs selected from the University of South Florida Free Association Norms (Nelson et al., 1998) that adhered to the same criteria for high-associate word pairs in Experiment 1 (FSG > .50, total Mean FSG = .624).

Procedure

Experiment 2 used the same procedure as Experiment 1, with the exception that the item and context recognition tasks were tested on the *target* words (i.e., the words generated by the participants at study). Responses for the lower-constraint generation task were recorded by the experimenter and uploaded into E-Prime for presentation in the recognition phase immediately following the encoding phase. This process took on average 1 min, and was done while the participant filled out a 6-question demographics questionnaire between the encoding and recognition phases. All other phases were identical to Experiment 1.

Results

For Experiment 2, participants responded to 100% of the generate and read trials, and successfully unscrambled 94% of the words in the scramble condition. The scramble trials that were not successfully unscrambled were removed from all analyses. Means and standard deviations of item and context recognition decisions responses and cued recall accuracy are presented in Table 1. A one-way repeated-measures ANOVA was conducted separately for each memory type (item recognition, context recognition,

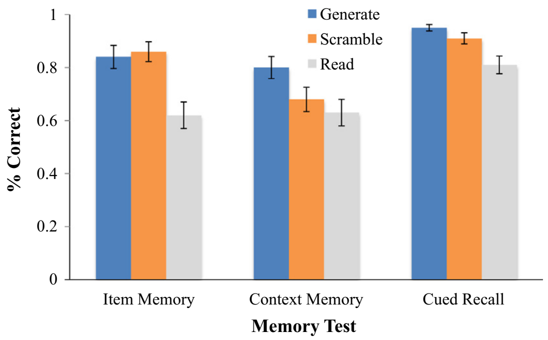

Fig. 5. Experiment 2: Accuracy by task for item recognition, context recognition, and cued recall. Error bars represent standard errors of the mean. Note: For item memory, generate and scramble differed from read at $p < .05$, but difference between generate and scramble was not significant. For context memory difference between scramble and read was not significant. For cued recall, all differences between conditions (generate, scramble, read) are significant at $p < .05$.

cued recall) to compare differences in memory between the three tasks (generate, scramble, read). Item and context recognition and cued recall scores were calculated in the same way as in Experiment 1. The data for these analyses are graphed in Fig. 5. Additionally, we conducted an item analysis to control for item-selection confounds as we did in Experiment 1.

Subject-based results

Results of the item recognition analysis revealed significant differences between the tasks, $F(2,21) = 23.63$, $MSE = .402$, $p < .001$, $\eta_p^2 = .518$. Planned paired-samples *t*-tests interestingly revealed no difference in item recognition between the generate and scramble tasks, $t(22) = -1.18$, $p = .25$. However, both generate, $t(22) = 4.68$, $p < .001$, and scramble, $t(22) = 5.42$, $p < .001$, showed better item recognition than the read task.

The context recognition analysis revealed significant differences between the tasks, $F(2,21) = 7.81$, $MSE = .188$ $p = .001$, $\eta_p^2 = .262$. As predicted, planned follow-up analyses showed that context recognition was significantly better in the generate task than both the scramble and the read tasks, $ts > 3.36$, $ps < .003$. Interestingly, context recognition was not better in the scramble task compared to read, $t(22) = 1.09$, $p = .29$.

Results of the cued recall analysis likewise revealed significant differences between the tasks, $F(2,21) = 16.41$, $MSE = .122$, $p < .001$, $\eta_p^2 = .427$. Planned follow-up analyses showed these differences in the predicted direction: Cued recall was higher in the generate task compared to both the scramble and read tasks, $ts > 2.92$, $ps < .008$, and scramble was recalled better than read, $t(22) = 3.51$, $p = .002$.

Item analysis

As in Experiment 1 we conducted an item analysis to examine item and context memory effects while controlling for item-selection confounds that may be influencing the between-subjects data. Since this analysis was over the target words, we limited our analysis of the generated words to only those where participants produced the expected normed target (e.g., hot – cold), which occurred for 64% of the generate trials in this experiment. For two of the 96 total word pairs included in the experiment, the expected target was never generated by any participant, thus we excluded these word pairs from this analysis. We conducted separate ANOVAs for item and context memory. Results of both item and context recognition fully replicated the between-subjects ANOVAs.

For item recognition, there was a significant main effect for task, $F(2,92) = 54.32$, $MSE = 1.82$, $p < .001$, $\eta_p^2 = .369$. Planned paired-samples *t*-tests revealed no difference in item recognition between the generate and scramble tasks, $t(93) = 0.41$, $p = .69$. However, both generate, $t(93) = 9.04$, $p < .001$, and scramble, $t(93) = 8.32$, $p < .001$, showed better item recognition than the read task.

For context recognition, again there was a significant main effect for task, $F(2,92) = 9.75$, $MSE = .576$, $p < .001$, $\eta_p^2 = .095$. Planned paired-samples *t*-tests revealed that context recognition was greater in the generate task compared to both the scramble and read tasks, $ts > 2.66$, $ps < .009$, and as in the between-subjects analysis, scramble was not better remembered than read, $t(93) = 1.70$, $p = .09$.

Experiment 2 Discussion

The results of Experiment 2 provide more evidence that a lower-constraint task (generate) enhanced memory compared to a higher-constraint generation task (scramble), with some caveats. For context memory recognition and cued recall of the target words, memory was strongest for words in the lower-constraint task, followed by the higher-constraint task and then read controls. These results further support the idea that the level of constraint in generation tasks influences the magnitude of the memory benefits obtained from self-generation. The fact that we found this increased benefit in context memory recognition and cued recall, but not in item memory recognition however, also suggests that there are boundary conditions under which the lower-constraint task provides additional memory benefits. Although we did not see the enhancement in item memory for the lower-constraint generation task compared to the higher-constraint task, both tasks were remembered with greater accuracy than the read task, consistent with the typical generation effect. Importantly, these effects were further substantiated by our item analysis. It should also be noted that memory performance for cued recall improved substantially from Experiment 1 to Experiment 2. This is likely a result of changing the recognition phase of the experiment to be on the *target* words, which are then tested again in the cued recall phase. This is as opposed to Experiment 1, where participants were seeing the *cue* words in the recognition phase, and the target words in the cued recall phase. Having the retrieval practice for the *target* words in the recognition phase in Experiment 2 most likely led to the increase in cued recall performance we saw compared to Experiment 1.

Despite the converging evidence with Experiment 1, there is one limitation to the procedures of Experiment 2 that obscures our interpretation that lower-constraint tasks lead to better memory than the higher-constraint task. Prior work on the generation effect has shown that the type of cognitive processing elicited by the generation task can influence the direction and magnitude of the generation effect, especially for contextual details, which is known as the *processing* account (Mulligan, 2004; Mulligan et al., 2006; Riefer, Chien, & Reimer, 2007). Specifically, these studies have shown that generation tasks that require conceptual processing, in which conceptual relations among the cue-target pair must be made (e.g., our lower-constraint task: open - _____), often lead to a negative or null generation effect for contextual details compared to a read control. In contrast, generation tasks that require perceptual-based processing (e.g., our higher-constraint scramble task: open – cseol) lead to enhanced context memory compared to control. Therefore, since the two generation tasks we compared in Experiments 1 and 2 (scramble, generate) differ in the type of processing they require at encoding, it could be argued that our effects are a result of the differences in the type of processing as opposed to the level of constraint as we have suggested. Although it should be noted that if this were true, then we should have expected to see better context memory in the scramble compared to the read task (Mulligan, 2004), which did not occur in our data. Nonetheless, to account for processing differences between our two generate tasks, we conducted a third experiment using a different higher-constraint generation task that induces more similar processing as our lower-constraint task (e.g., a word fragment: open – cl∗s∗), where participants are required to make a more conceptual connection to the cue word given in order to determine the word. This process could be considered more similar to our lower-constraint task where participants only have the cue word to determine their response. Equating the lower- and higher-constraint tasks on the type of processing allows us to more directly test whether these results are due to differences in the level of generation constraint.

Experiment 3

Experiments 1 and 2 supported our hypothesis that fewer generation constraints improve the generation effect compared to higher-constraint generation tasks often used in prior work. However, it could be argued that the two generation tasks we compared in these experiments may have led to qualitatively different types of processing (i.e., generating a word versus unscrambling letters). We designed Experiment 3 to equate the lower- and higher-constraint tasks on processing demands. Specifically, we compared our lower-constraint task to a word fragment task commonly used in prior studies (Clark, 1995; Gardiner, 1988; Gardiner & Hampton, 1988; Hirshman & Bjork, 1988; McElroy, 1987) that requires similar processing. If we see similar results as our previous two experiments, we can be more confident that our memory effects are due to differences in the level of constraint, and not differences in the type of processing.

Method

Participants

Twenty-four adults (mean age: 19.5, *SD*: 1.4, range: 18–23, 16 females) were recruited from the University of Illinois at Chicago introductory psychology course subject pool and community solicitation, and gave their informed consent and were paid or given course credit. All participants reported high proficiency in English speaking, reading, and understanding.

Stimuli

The same 96 cue-target word pairs were used as in Experiment 2. Word fragments for the higher-constraint task were created by omitting the vowels of each target word (e.g., open – cl∗s∗), a procedure done in previous generation effect studies (Clark, 1995; Gardiner & Hampton, 1988; Hirshman & Bjork, 1988; McElroy, 1987).

Procedure

The procedure for Experiment 3 was identical to Experiment 2 (recognition and recall of *target* words), with the exception that we replaced the scramble task with the word fragment task as our higher-constraint generation task.

Results

For Experiment 3, participants gave responses to 100% of the generate trials and read trials, and 96% of the fragment trials on average. Unsuccessful fragment trials that participants could not successfully complete were removed from all analyses. Means and standard deviations of item and context recognition decisions responses and cued recall accuracy are presented in Table 1. A one-way repeated-measures ANOVA was conducted separately for each memory type (item recognition, context recognition, cued recall) to compare differences in memory between the three tasks (generate, fragment, read). Item and context recognition and cued recall scores were calculated in the same way as in Experiments 1 and 2. The data for these analyses are graphed in Fig. 6.

Subject-based results

Results of the item recognition analysis revealed significant differences between the tasks, $F(2,22) = 25.41$, $MSE = .395$, $p < .001$, $\eta_p^2 = .525$. Planned, paired-samples *t*-tests between the tasks revealed no difference in item recognition between the generate and fragment tasks, $t(23) = -1.28$, $p = .21$. However, both generate, $t(23) = 4.92$, $p < .001$, and fragment, $t(23) = 7.40$, $p < .001$, showed better item recognition than the read task, consistent with the typical generation effect.

The context recognition analysis revealed significant differences between the tasks, $F(2,22) = 5.24$, $MSE = .114$, $p = .009$, $\eta_p^2 = .186$. Our planned follow-up analyses showed that context recognition was significantly better in the generate task than the fragment task, $t(23) = 3.34$, $p = .003$, but interestingly only marginally better than the read task $t(23) = 1.96$, $p = .063$. Again, as in our other

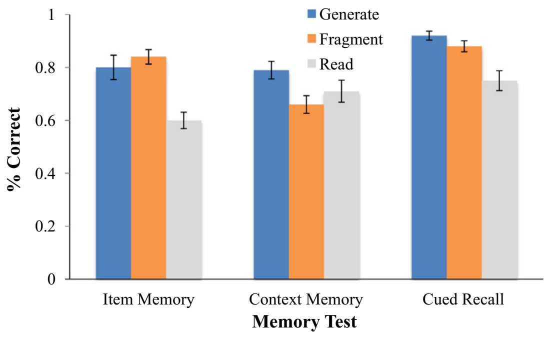

Fig. 6. Experiment 3: Accuracy by task for item recognition, context recognition, and cued recall. Error bars represent standard errors of the mean. Note: For item memory, generate and fragment differed from read at $p < .05$, but difference between generate and fragment was not significant. For context memory, generate differed from fragment at $p < .05$, but all other differences are not significant. For cued recall, all differences significant at $p < .05$.

experiments there was no difference in memory between the fragment and read tasks, $t(23) = -1.17$, $p = .254$.

Results of the cued recall analysis likewise revealed significant differences between the tasks, $F(2,22) = 20.55$, $MSE = .186$, $p < .001$, $\eta_p^2 = .472$. Our planned follow up analyses showed these differences in the predicted direction: Cued recall was higher in the generate task compared to both the fragment and read tasks, $ts > 2.35$, $ps < .028$, and fragment words were recalled better than read words, $t(23) = 3.96$, $p = .001$.

Item analysis

As in Experiments 1 and 2 we conducted an item analysis to examine item and context memory effects. We again limited our analysis of the generated words to only those where participants produced the expected normed target (e.g., hot – cold), which occurred for 61% of the generate trials in this experiment. For four of the 96 total word pairs included in the experiment, the expected, normed target was never generated by any participant therefore we excluded these word pairs from this analysis ($N = 92$). Repeated measures ANOVAs for item and context memory provided similar, but not identical results as the between subject ANOVAs.

For item recognition, there was a significant main effect for task, $F(2,90) = 34.28$, $MSE = 1.58$, $p < .001$, $\eta_p^2 = .274$. Planned paired-samples t-tests revealed no difference in item recognition between the generate ($M = .798$, $SD = .260$) and fragment ($M = .837$, $SD = .158$) tasks, $t(91) = -1.28$, $p = .20$. However, both generate, $t(91) = 5.61$, $p < .001$, and fragment, $t(91) = 9.25$, $p < .001$, showed better item recognition than the read task ($M = .594$, $SD = .243$).

For context recognition, again there was a significant main effect for task, $F(2,90) = 6.97$, $MSE = .476$, $p = .001$, $\eta_p^2 = .071$. Planned paired-samples t-tests in this analysis revealed that context recognition was greater in the generate task ($M = .797$, $SD = .296$) compared to both the fragment ($M = .657$, $SD = .266$) and read tasks ($M = .700$, $SD = 241$), $ts > 2.46$, $ps < .016$, which is consistent with our predictions. Again no difference was found between the fragment and read tasks in this analysis, $t(91) = -1.21$, $p = .23$.

Experiment 3 Discussion

Experiment 3 provided more evidence that the level of constraint can influence the size of the memory benefits from self-generation under some, but not all, circumstances. Furthermore, by using a higher-constraint generation task that is similar in the type of processing required by the lower-constraint task, we can be more confident that our findings were not due to differences in processing demands in the lower- versus higher-constraint tasks.

As in Experiment 2, we did not see any additional benefit for the lower-constraint task compared to the higher-constraint task for item recognition. Broadly, these findings suggest item memory for high-associates may be less affected by the level of constraint given that both Experiments 2 and 3 used only high-associate word pairs. However, this finding should be interpreted cautiously since all three experiments found that the cued recall measure of item memory did show memory enhancement for the lower-constraint compared to the higher-constraint task.

Our findings for context memory again showed improved context memory for a lower-constraint task compared to the higher-constraint task, despite these generation tasks being more similar in the type of processing required. This finding is in contrast to prior work that has shown that unscrambling a word led to improved context memory relative to controls, while word completion task (e.g., open – c____) led to decreased context memory compared to controls (Mulligan, 2004). However, in this prior study, a different type of contextual detail was measured where participants were asked about the font and background color, as opposed to which task was performed as in the present study. Further, Mulligan (2004) tested different types of generation tasks (unscrambling, word completion) separately in different experiments and did not use a free response generation task as in our study. Therefore it is difficult to make a direct comparison between these contrasting results.

Overall, these results from Experiment 3 further support the claim that level of constraint can influence the magnitude of the generation effect and provide evidence against the idea that these improvements are due to qualitative differences in processing. Still, further work is necessary to better understand why lower-constraint tasks lead to improved memory relative to higher-constraint tasks.

General discussion

In this series of experiments we examined whether the level of generation constraint influences the magnitude of the generation effect in memory. Specifically, we tested whether a lower-constraint task led to greater memory improvements for both item and context memory compared to higher-constraint tasks. Overall, we found that context memory was consistently better in a lower-constraint task compared to a higher-constraint task. For item memory, however, we saw improved memory for materials in the lower-constraint relative to a higher-constraint task only under some conditions. Overall, these data support the idea that self-generation is a powerful mnemonic that leads to detail-rich representations of an episode, but that level of constraint of the generation task influences subsequent memory performance.

Prior work has shown that the generation effect is an effective mnemonic that leads to memory improvements relative to non-generated materials (Bertsch et al., 2007; Jacoby, 1978; Slamecka & Graf, 1978). Our results replicate these findings, but also add to this literature by suggesting that generation effects may be larger when participants are more free to generate materials. Interestingly, essentially no studies have scrutinized how differences in constraint might influence subsequent memory performance. This study has shown that the level of generation constraint significantly influences the mnemonic benefits of this phenomenon, especially for context memory. Across three experiments we showed that material produced in a lower-constraint generation task led to improved item and context memory over two higher-constraint generation tasks used in prior studies (scramble, word fragment); however, we also found instances where item memory did not differ, suggesting that item memory may be more protected from varying levels of constraint. The idea that item memory is less effected by generation constraint may partially explain the robust item memory findings in prior work on the generation effect (Bertsch et al., 2007). Nonetheless, our findings suggest that level of constraint dictated by a generation task can influence later memory, which is a factor to account for in future work.

Item memory benefits from self-generation have been consistently reported (Graf, 1982; McElroy & Slamecka, 1982; McFarland et al., 1980; Slamecka & Graf, 1978). This effect has been shown across a variety of manipulations and memory tests (Bertsch et al., 2007). Importantly though, our results from Experiment 1 extend this prior work by suggesting that memory for items produced in an lower-constraint generation task is greater than for items produced in a higher-constraint task. However, in contrast to the item recognition results of Experiment 1, item recognition did not differ between the lower- and higher-constraint tasks in Experiment 2 and 3. We should note, however, that memory performance for both generate tasks was high and significantly better than control, replicating the typical generation effect. Furthermore, item memory as measured by cued recall showed improved memory for the lower-constraint task compared to the higher-constraint task in all experiments. Because we only saw item recognition improvements in Experiment 1, but not in the follow-up studies, our findings suggest that the item memory benefits (as measured by recognition) from self-generation may be more protected against the influence of generation constraint.

One possible reason we did not see item memory improvement for a lower-constraint compared to higher-constraint in Experiments 2 and 3 is our use of only high-associates as stimuli (and not both high- and low-associates as in Experiment 1). This possibility is consistent with prior work showing that targets generated from high-associate cues are better remembered than words generated from low-associate cues (Hirshman & Bjork, 1988; Taconnat et al., 2008). This suggests that when the generation task is easy, constraints may have less impact on subsequent memory. Another possible reason we did not see improved item recognition for the lower-constraint compared to higher-constraint task in these two experiments is because we tested memory for the target word (and not the cue as in Experiment 1). One reason why this factor may have produced the findings in Experiment 1, but not in 2 and 3 could be that in our lower-constraint task (e.g., open – _____) participants spent more time processing the *cue* word in order to freely generate a target word. By contrast, in the higher-constraint tasks (e.g., open – coesl) participants likely spent more time processing the *target* word to determine the "correct" response. By this logic, it follows that memory for the *cue* words would be stronger for the lower-constraint task compared to the higher-constraint task as is supported by the results of Experiment 1. Overall, our study showed that a lower-constraint task produced greater or equal item memory benefits compared to a higher-constraint task suggesting that fewer constraints may not always provide greater memory benefits.

Across all experiments, we consistently found improved context memory in the lower-constraint versus the higher-constraint task, again suggesting that fewer constraints can improve memory for contextual details associated with that item. This finding was the central focus and most important finding of this experimental set, given that prior work on context memory is less clear: some work has supported the idea that self-generation improves context memory (Geghman & Multhaup, 2004; Marsh, 2006; Marsh et al., 2001), whereas other work has not (Jurica & Shimamura, 1999; Mulligan, 2004; Nieznański, 2011). In this prior work however, generation tasks varied in their level of constraint, possibly influencing the magnitude of the memory effects for context memory. Unlike item memory, we found no boundary condition for these added memory effects, that is, we found evidence of context memory improvement for the lower- compared to higher-constraint task across all conditions in all experiments. Importantly, our results showed no context memory benefit for the higher-constraint tasks compared to read, but often large benefits in the lower- compared to both the higher-constraint and read tasks. Together, these findings provide more evidence of the importance of considering the level of constraint in generation tasks when studying the generation effect, especially for contextual details. Furthermore, a lower-constraint generation task likely captures people's day-to-day experiences of generation more accurately compared to higher-constraint tasks. Using lower-constraint tasks in future studies may allow for more accurate generalizations to what generation effects might look like outside of the laboratory setting.

Considering both the item and context memory effects from these experiments together, the data support the idea that self-generation is a powerful mnemonic leading to improved memory for both types of details consistent with the *item-context enhancement* account. These findings are in contrast to the Jurica and Shimamura (1999) findings, whose results supported the *item-context tradeoff* account when comparing a lower-constraint generation task to control. However, their study used a conversational procedure and different stimuli for the generate and control tasks that possibly rely on different encoding processes that could explain why our results support a different conclusion. Importantly, Jurica and Shimamura (1999) did not include a higher-constraint task to compare to their generation task so it is not possible to tell how memory benefits were influenced by level of constraint. In general, our results suggest that prior work in this area may have been mixed due to the use of different generation tasks that vary in the amount of constraints. Our findings potentially provide clarity to the literature and suggest that the generation effect improves memory for both the item and the contextual details of an episode.

Our findings for cued recall showed evidence of a generation effect where generated items (lower- and higher-constraint) were better recalled compared to control which is consistent with previous work (Donaldson & Bass, 1980; Graf, 1980; Johnson et al., 1981). Importantly, our findings extend this work by showing improved memory for a lower-constraint versus the higher-constraint tasks in all comparisons, with the exception of the comparison between high-associates trial types in Experiment 1. In accord with our interpretation of our item recognition findings, this exception provides further support for the idea that when the generation task is easy, level of constraint may not be as impactful. Given that item recognition and cued recall both measure item memory it is interesting that we found improved memory for our lower-constraint task over a higher-constraint task in all of our cued recall tests but not for item recognition in Experiments 2 and 3. Since it is known that cued recall relies on "recollection" processes (as opposed to familiarity) to produce accurate responses (Yonelinas, 2002), it could be that memory for materials generated in lower-constraint task are more supported by recollection processes relative to the higher-constraint tasks. Although speculative, this claim is consistent with prior work suggesting self-generation may more heavily support recollection than familiarity (Dodson & Johnson, 1996; Jacoby, 1991; Jennings & Jacoby, 1993; Verfaellie & Treadwell, 1993; Yonelinas, 2002). More work will be necessary to interrogate this prospect.

Our item analyses in all experiments provide compelling evidence that our memory effects shown in our primary analyses were apparent even when accounting for possible item-selection effects. Specifically, we saw similar item and context memory effects for materials presented in the lower-constraint compared to the higher-constraint and read tasks even when the item itself was controlled to be the same across these conditions. Overall, the item analysis offers further evidence that materials generated in a lower-constraint task are better remembered than in higher-constraint generation tasks.

In prior work there have been many attempts to explain why generated information is more memorable than read controls. One way prior work has suggested that self-generation leads to improved memory is by inducing greater depth of processing relative to control (Graf, 1982; McElroy & Slamecka, 1982). However, this account has been challenged: if depth of processing is at least partially responsible for the generation effect, it should be the case that the cue that one uses to generate a target should also benefit from deeper processing and hence be more memorable. Yet several studies have failed to find improved memory for the cue word relative to control (Donaldson & Bass, 1980; Slamecka & Graf, 1978) arguing against this depth of processing mechanism. In contrast to this prior work, evidence from our Experiment 1 clearly shows an item memory benefit for the cue words in the lower-constraint task compared to both the higher-constraint task and control, consistent with this depth of processing account. Therefore, it could be that a lower-constraint task elicits deeper processing compared to a higher-constraint task. It is possible that lower-constraint task requires a greater "mental search" compared to the scramble or word fragment tasks, leading to more deeper encoding, and thus improving later memory for words generated in that task compared to controls.

Another possible explanation for our observed memory improvement in the lower-constraint compared to higher-constraint task is rooted in transfer appropriate processing (Morris, Bransford, & Franks, 1977). Transfer appropriate processing assumes that memory is enhanced for items that overlap in the type of processing (perceptual or conceptual) engaged at encoding and retrieval. It could be that the lower-constraint task elicits encoding processes that are more closely matched at retrieval compared to the higher-constraint task. This finding is most relevant for our cued recall findings, where the task of freely generating a word from a cue, is akin to the procedure we ask participants to do when recalling a word from a cue (i.e., both task give a cue word and ask participants to produce a target word). Transfer appropriate processing is also mentioned in the processing account put forth by Mulligan (2004) to describe the relationship between generation and context memory improvements. Specifically, this account proposes that generation tasks that are more perceptual (e.g., unscrambling a word) will provide memory benefits for perceptual contextual details (e.g., font color, background color, etc.) whereas a generation tasks that are more conceptual (e.g., word fragment tasks) will provide memory benefits for conceptual contextual details. Our findings of context memory are also in line with this account (i.e., conceptual-based generation task improved conceptual context memory detail, memory for the task); however, it is unknown whether our lower-constraint task also improves memory for perceptual contextual details (i.e., color details, visuospatial details, etc.). Future work will be necessary to test this idea which will allow for more insight into whether the processing account can explain the effects seen in the present study.

An alternative explanation for our context memory findings comes from the work on the production effect (MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010) which posits that items that are encoded in a distinct manner are more memorable at retrieval. For instance, if participants are asked to perform a task in which they read words in a British accent, or blink repeatedly after reading a word, memory for words in these tasks may be more memorable because of the distinctiveness of the task. In the case of the present study, it could be argued that a lower-constraint task is more distinctive compared to the higher-constraint and read tasks, thus leading to greater memory for the task of words encoded in the lower-constraint task at retrieval (Hunt & Worthen, 2006). Future work is necessary to vet this possibility.

A final possible reason we found better memory for materials produced in a lower-constraint task is rooted in the self-reference effect (Leshikar, Cassidy, & Gutchess, 2016; Leshikar, Park, & Gutchess, 2014; Leshikar et al., 2015; Rogers, Kuiper, & Kirker, 1977; Symons & Johnson, 1997). Because the lower-constraint task allows for a free response as opposed to a constrained "correct" response, participants may produce a response that is more personally meaningful and hence more memorable, which is analogous to a self-reference effect. Although possible, the analysis of the "matched" versus "novel" generation trials (see footnote 3) argues against this prospect. Future work is necessary to investigate the question of why lower-constraint generation leads to greater memory benefits compared to higher-constraint generation tasks, especially for context memory where prior work has shown mixed results.

Conclusion

Across all experiments, we consistently found memory improvements for materials generated in a lower-constraint compared to two higher-constraint generation tasks commonly used in prior work. This has the important implication that the amount of constraint a generation task imposes should be taken into account in future work. Particularly, less constrained generation tasks provide greater enhancement of extraneous, contextual details of an episode. Overall, this study suggests the generation effect is a potent mnemonic that leads to rich, detailed memory representations.

Acknowledgments

We thank Dr. Gary Raney for his helpful suggestions to improve the design of this project. We also would like to thank the two anonymous reviewers for their thoughtful feedback that helped strengthen this manuscript.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jml.2016.06.007.

References

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition, 35*(2), 201–210.

Clark, S. E. (1995). The generation effect and the modeling of associations in memory. *Memory & Cognition, 23*(4), 442–455.

Dodson, C. S., & Johnson, M. K. (1996). Some problems with the process-dissociation approach to memory. *Journal of Experimental Psychology: General, 125*(2), 181.

Donaldson, W., & Bass, M. (1980). Relational information and memory for problem solutions. *Journal of Verbal Learning and Verbal Behavior, 19*(1), 26–35.

Duarte, A., Henson, R. N., & Graham, K. S. (2008). The effects of aging on the neural correlates of subjective and objective recollection. *Cerebral Cortex, 18*(9), 2169–2180.

Duarte, A., Henson, R. N., Knight, R. T., Emery, T., & Graham, K. S. (2010). Orbito-frontal cortex is necessary for temporal context memory. *Journal of Cognitive Neuroscience, 22*(8), 1819–1831.

Fiedler, K., Lachnit, H., Fay, D., & Krug, C. (1992). Mobilization of cognitive resources and the generation effect. *The Quarterly Journal of Experimental Psychology, 45*(1), 149–171.

Foley, M. A., & Foley, H. J. (2007). Source-monitoring judgments about anagrams and their solutions: Evidence for the role of cognitive operations information in memory. *Memory & Cognition, 35*(2), 211–221.

Foley, M. A., Foley, H. J., Wilder, A., & Rusch, L. (1989). Anagram solving: Does effort have an effect? *Memory & Cognition, 17*(6), 755–758.

Gardiner, J. M. (1988). Generation and priming effects in word-fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 495.

Gardiner, J. M., & Hampton, J. A. (1988). Item-specific processing and the generation effect: Support for a distinctiveness account. *The American Journal of Psychology*, 495–504.

Geghman, K. D., & Multhaup, K. S. (2004). How generation affects source memory. *Memory & Cognition, 32*(5), 819–823.

Graf, P. (1980). Two consequences of generating: Increased inter-and intraword organization of sentences. *Journal of Verbal Learning and Verbal Behavior, 19*(3), 316–327.

Graf, P. (1982). The memorial consequences of generation and transformation. *Journal of Verbal Learning and Verbal Behavior, 21*(5), 539–548.

Hirshman, E., & Bjork, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 484.

Hirshman, E., & Mulligan, N. (1991). Perceptual interference improves explicit memory but does not enhance data-driven processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(3), 507.

Hunt, R. R., & Worthen, J. B. (2006). *Distinctiveness and memory* : . Oxford University Press.

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior, 17*(6), 649–667.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*(5), 513–541.

Jennings, J. M., & Jacoby, L. L. (1993). Automatic versus intentional uses of memory: Aging, attention, and control. *Psychology and Aging, 8*(2), 283.

Johnson, M. K., Raye, C. L., Foley, H. J., & Foley, M. A. (1981). Cognitive operations and decision bias in reality monitoring. *The American Journal of Psychology*, 37–64.

Jurica, P. J., & Shimamura, A. P. (1999). Monitoring item and source information: Evidence for a negative generation effect in source memory. *Memory & Cognition, 27*(4), 648–656.

Leshikar, E. D., Cassidy, B. S., & Gutchess, A. H. (2016). Similarity to the self influences cortical recruitment during impression formation. *Cognitive, Affective, & Behavioral Neuroscience, 16*(2), 302–314. http://dx.doi.org/10.3758/s13415-015-0390-3.

Leshikar, E. D., & Duarte, A. (2012). Medial prefrontal cortex supports source memory accuracy for self-referenced items. *Social Neuroscience, 7*(2), 126–145.

Leshikar, E. D., & Duarte, A. (2014). Medial prefrontal cortex supports source memory for self-referenced materials in young and older adults. *Cognitive, Affective, & Behavioral Neuroscience, 14*(1), 236–252.

Leshikar, E. D., Dulas, M. R., & Duarte, A. (2015). Self-referencing enhances recollection in both young and older adults. *Aging, Neuropsychology, and Cognition, 22*(4), 388–412.

Leshikar, E. D., & Gutchess, A. H. (2015). Similarity to the self affects memory for impressions of others. *Journal of Applied Research in Memory and Cognition, 4*(1), 20–28.

Leshikar, E. D., Park, J. M., & Gutchess, A. H. (2014). Similarity to the self affects memory for impressions of others in younger and older adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*. gbt132.

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal* *of Experimental Psychology: Learning, Memory, and Cognition, 36*(3), 671–544.

Marsh, E. J. (2006). When does generation enhance memory for location? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(5), 1216.

Marsh, E. J., Edelman, G., & Bower, G. H. (2001). Demonstrations of a generation effect in context memory. *Memory & Cognition, 29*(6), 798–805.

McElroy, L. A. (1987). The generation effect with homographs: Evidence for postgeneration processing. *Memory & Cognition, 15*(2), 148–153.

McElroy, L. A., & Slamecka, N. J. (1982). Memorial consequences of generating nonwords: Implications for semantic-memory interpretations of the generation effect. *Journal of Verbal Learning and Verbal Behavior, 21*(3), 249–259.

McFarland, C. E., Frey, T. J., & Rhodes, D. D. (1980). Retrieval of internally versus externally generated words in episodic memory. *Journal of Verbal Learning and Verbal Behavior, 19*(2), 210–225.

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*(5), 519–533.

Mulligan, N. W. (2004). Generation and memory for contextual detail. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(4), 838.

Mulligan, N. W. (2011). Generation disrupts memory for intrinsic context but not extrinsic context. *The Quarterly Journal of Experimental Psychology, 64*(8), 1543–1562.

Mulligan, N. W., Lozito, J. P., & Rosner, Z. A. (2006). Generation and context memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(4), 836.

Murnane, K., & Bayen, U. J. (1996). An evaluation of empirical measures of source identification. *Memory & Cognition, 24*(4), 417–428.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms.

Nieznański, M. (2011). Generation difficulty and memory for source. *The Quarterly Journal of Experimental Psychology, 64*(8), 1593–1608.

Riefer, D. M., Chien, Y., & Reimer, J. F. (2007). Positive and negative generation effects in source monitoring. *The Quarterly Journal of Experimental Psychology, 60*(10), 1389–1405.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology, 35*(9), 677.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human learning and Memory, 4*(6), 592.

Slamecka, N. J., & Katsaiti, L. T. (1987). The generation effect as an artifact of selective displaced rehearsal. *Journal of Memory and Language, 26*(6), 589–607.

Symons, C. S., & Johnson, B. T. (1997). The self-reference effect in memory: A meta-analysis. *Psychological Bulletin, 121*(3), 371.

Taconnat, L., Froger, C., Sacher, M., & Isingrini, M. (2008). Generation and associative encoding in young and old adults: The effect of the strength of association between cues and targets on a cued recall task. *Experimental Psychology, 55*(1), 23–30.

Tyler, S. W., Hertel, P. T., McCallum, M. C., & Ellis, H. C. (1979). Cognitive effort and memory. *Journal of Experimental Psychology: Human Learning and Memory, 5*(6), 607.

Verfaellie, M., & Treadwell, J. R. (1993). Status of recognition memory in amnesia. *Neuropsychology, 7*(1), 5.

Watkins, M. J., & Sechler, E. S. (1988). Generation effect with an incidental memorization procedure. *Journal of Memory and Language, 27*(5), 537–544.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*(3), 441–517.