Examining the relationship between generation constraint and memory

Matthew P. McCurdy¹ · Andrea N. Frankenstein¹ · Allison M. Sklenar¹ · Pauline Urban Levy¹ · Eric D. Leshikar¹

Accepted: 16 November 2020 / Published online: 7 January 2021
© The Psychonomic Society, Inc. 2021

Abstract

Self-generated information is often better remembered than read information (the generation effect). Recent research, however, has shown that generating information under fewer experimental constraints (i.e., fewer limitations on what can be generated) can increase the magnitude of the generation effect. This study systematically varied generation constraint to better understand the effects of constraint on memory. Participants encoded associated cue–target word pairs (above–below) on either the left or right side of a computer monitor. At encoding, generation constraint was manipulated by systematically varying the number of letters given to participants to generate the target word (i.e., above–below; option–choic_ ; bank–mon_ ; etc.). At retrieval, participants were given either a recognition, cued recall, or free recall test measuring both item (target word) and context memory (location on the computer monitor). Using mixed-effects logistic regression analyses to control for item-selection effects (e.g., participants producing idiosyncratic targets in some conditions relative to others), results indicated that generation constraint significantly influenced item, but not context (location) memory. The relationship between generation constraint and memory performance, however, differed by the type of memory test administered: Recognition data revealed a curvilinear relationship; cued recall showed a negative, linear relationship; and free recall showed no significant relationship. Overall, these findings provide more evidence that generation constraint has a strong yet complex effect on different aspects of memory, and further delineates some boundary conditions of the influence of generation constraint on memory.

Keywords Generation effect · Generation constraint · Item memory · Context memory

Information that is self-generated is often better remembered than information that is read, or provided from another source (Jacoby, 1978; Slamecka & Graf, 1978). This memory benefit, known as the *generation effect*, has been studied extensively over the years. A meta-analysis on the generation effect found that self-generating provides more than a 10% improvement in memory performance compared with reading across a variety of experimental conditions (McCurdy, Viechtbauer, Sklenar, Frankenstein, & Leshikar, 2020). Despite the general robustness of this effect, research has revealed several experimental factors that strongly influence the magnitude of generation effects (e.g., within-subjects/between-subjects design, semantic/lexical activation; McCurdy, Viechtbauer, et al., 2020). In particular, one factor known as *generation constraint* strongly influences the size of the generation effect

(see McCurdy, Viechtbauer, et al., 2020), yet comparatively little research has been devoted to understanding *how* differences in generation constraint act to influence the memory benefits from self-generation. Given the relative dearth of research on this influential factor, we examine the relationship between generation constraint and memory performance across a variety of memory tests to better understand the underlying mechanisms by which constraint influences the magnitude of generation effects.

Generation constraint refers to the amount of information that limits what participants can self-generate. For example, in a typical generation effect study, participants are often given a cue word, and asked to generate a target word from a fragment (e.g., above–bel_). Throughout the literature examining generation effects, researchers have used generation tasks that vary in the amount of constraint they provide without consideration of how this factor impacts the size of the memory benefits across studies. A growing body of research, however, suggests that generation tasks that place less constraints on what participants can generate (by providing less information about the target word) often lead to larger generation effects,

✉ Eric D. Leshikar
leshikar@uic.edu

¹ Department of Psychology, University of Illinois at Chicago, 1007 W Harrison St (MC 285), Chicago, IL 60607, USA

relative to tasks with more constraints (Fiedler, Lachnit, Fay, & Krug, 1992; Gardiner, Smith, Richardson, Burrows, & Williams, 1985; McCurdy, Leach, & Leshikar, 2017, 2019; McCurdy, Sklenar, Frankenstein, & Leshikar, 2020). For example, we observed improved memory in a lower-constraint generation task where participants were given less information about the target word (e.g., above—) compared with a higher-constraint generation task where participants were given more information about the target item (e.g., completing a word fragment: above—bel_, or solving an anagram of the target: above—blweo; McCurdy et al., 2017; McCurdy, Sklenar, et al., 2020).

A major critique of studying the influence of generation constraint on memory, however, is that in lower-constraint generation tasks, participants are more likely to generate an idiosyncratic response compared with a higher-constraint task where an expected target item is more likely to be generated (Slamecka & Graf, 1978). As an example, in a lower-constraint generation task (e.g., open—), a subject is more likely to generate an idiosyncratic response because there is less information guiding what participants should generate, whereas in a higher-constraint generation task (e.g., open—cl_), participants are less likely to produce an idiosyncratic target response because there is more information guiding the participant to generate the normed, expected response. This creates greater potential for item-selection effects in lower-constraint generation tasks where information generated across different generation conditions is not the same, which is potentially problematic because differences in memory performance may be due to generation constraint manipulations (i.e., lower versus higher constraint) or to differences in the memorability of the idiosyncratic items themselves (Slamecka & Graf, 1978). Although avoiding lower-constraint tasks is understandable given potential item-selection confounds, this has created a barrier to producing new knowledge on the generation effect, which is especially important given that past work has shown that less constraints can increase the magnitude of the generation effect (Fiedler et al., 1992; Gardiner et al., 1985; McCurdy et al., 2017, 2019; McCurdy, Sklenar, et al., 2020). Devising ways to understand the influence of generation constraint while controlling for item-selection effects is a potential avenue to make continued progress. In this study, we use a mixed-effects regression technique to model the unique influence of generation constraint on memory performance while controlling for item-selection effects.

Although past work has shown lower-constraint compared with higher-constraint tasks can improve memory, there may be important boundary conditions, or situations when lower-constraint generation does not increase the generation effect relative to higher constraint. To understand the potential reasons for these boundary conditions, it is first useful to understand mechanisms thought to underlie the generation effect in general. Theoretical work suggests that generating (relative to

reading) enhances two types of processes at encoding: item-specific and relational processing (Hirshman & Bjork, 1988; McDaniel, Waddill, & Einstein, 1988). Broadly, *item-specific processing* is the processing of information unique to the target item (e.g., the sound of the word; the letters that make up the word). *Relational processing* is the processing of information presented (or associated) with the target item. More specifically, relational processing can refer to the processing of the relationship between the target word and several different types of details, such as the cue word (i.e., cue–target relational processing; Hirshman & Bjork, 1988), other target words presented in the same list of stimuli (i.e., target–target relational processing; McDaniel, Riegler, & Waddill, 1990; McDaniel et al., 1988), as well as contextual details (e.g., spatial location, source) presented with the target item (i.e., context–target relational processing; Greenwald & Johnson, 1989; Marsh, 2006). Because research has shown that different memory tests are sensitive to different types of memory representations (Burns, 2006), it is possible to understand the influence of item-specific and relational processing on memory (and in turn generation constraint) by using different memory tests (see Table 1). Specifically, item recognition tests are sensitive to item-specific information, as these tests require distinguishing a previously studied item among distractor items (Begg, Snider, Foley, & Goddard, 1989; Burns, 2006). In contrast, cued recall tests are sensitive to cue–target relational information, as they rely on remembering which target went with a particular cue (i.e., remembering the cue–target relationship; Hirshman & Bjork, 1988). Free recall tests are thought to be sensitive to target–target relational information (i.e., the relationship among multiple target items in a list; Burns, 2006; Einstein & Hunt, 1980; Hunt & Einstein, 1981; McDaniel et al., 1990). Given that there is no cue word provided to aid recall, participants often rely on previously recalled target words as a cue for remembering additional target words. Thus, a stronger memory representation of the relationship among different target words in a list is often associated with increased free recall performance (Burns, 2006; Einstein & Hunt, 1980; Hunt & Einstein, 1981; McDaniel et al., 1990). As for context memory tests, given that participants are asked to remember the relationship between the target word and the context in which it was encoded, context memory tests are thought to be sensitive to the context–target relational information (Chalfonte & Johnson, 1996; Greenwald & Johnson, 1989).

Turning back to generation constraint, this principle (different encoding tasks induce different types of processing at encoding, which in turn are measurable by different memory tests) can be used to better understand conditions under which generation constraint might affect memory. Prior studies examining generation constraint have shown the strongest effects (lower greater than higher constraint) for measures of cued recall and context (source) memory, relative to item recognition (McCurdy et al.,

Table 1 Type of memory tests included in the study, test sensitivities to the processing of different types of information, and predictions of generation constraint effects for each test

Memory test type	Processing type	Generation constraint effect prediction
Recognition	Item specific	No relationship
Cued recall	Cue–target relational	Negative relationship (higher constraint = lower memory)
Free recall	Target–target relational	No prediction
Context memory	Context–target relational	Negative relationship (higher constraint = lower memory)

2017; McCurdy, Sklenar, et al., 2020), suggesting that differences in generation constraint may primarily act on relational processing at encoding (e.g. cue–target and context–target relational processing). Indeed, other work using free recall tasks have also shown that generation tasks with fewer constraints yield larger generation effects, although these studies were not designed to examine generation constraint, explicitly (Fiedler et al., 1992; Gardiner et al., 1985). Thus, based on this work, we expect that lower-constraint tasks should improve memory on tasks that are sensitive to relational information (e.g., cued recall, context memory recognition). In contrast, we expect that generation constraint effects may be reduced on tasks sensitive to item-specific processing (e.g., item memory as measured through recognition; McCurdy, Sklenar, et al., 2020). In the present study, we extend prior work by measuring the relationship between generation constraint and memory performance on three different types of memory tests thought to be sensitive to different memory representations. Specifically, we examine differences in memory performance for information learned under varying generation constraints as measured by recognition, cued recall, free recall memory tests, and for both item and context (spatial location) memory, to better understand the memory mechanism(s) underlying the effects of generation constraint.

Although prior work has shown that generation constraint has a strong influence on memory, many of the studies investigating this factor have relied on dichotomous manipulations of constraint (lower versus higher constraint). A more rigorous approach is to systematically vary generation constraint, as we do in this study, because it is possible to observe more nuanced memory effects that may not be captured by a dichotomous manipulation of constraint. Specifically, a systematic approach provides a way to examine the relationship between constraint and memory performance more precisely, such as whether the relationship is linear or nonlinear. As an example, research on the concept of desirable difficulties (Bjork, 1994) has suggested possible inverted U-shape functions in memory. The term desirable difficulty refers to the common yet counterintuitive finding that manipulations that make initial learning more difficult often improve long-term memory. However, introducing too much difficulty can have the opposite effect where learning becomes too effortful and long-term memory is reduced. It is possible that generation constraint might have a similar effect on memory, where there is an

“optimal” level of constraint (not too much, not too few) that could lead to increased memory performance, but dichotomous manipulations are not precise enough to detect such effects. To our knowledge, only one other study has examined a continuous relationship between constraint and memory, which showed a negative, linear relationship between constraint and memory performance (Gardiner et al., 1985). Critically though, memory performance was only measured through a free recall test; therefore, it is unknown how constraint might influence different measures of memory, such as item recognition. Given that the continuous relationship between constraint and memory is essentially unknown, we make no strong predictions, but the systematic manipulation of constraint used in the present study should produce important new knowledge about the generation effect and generation constraint.

Present study

In this study, we systematically vary generation constraint by removing letters of the to-be-generated target word to investigate the relationship between generation constraint and memory. We use three different memory tests in a between-subjects design (item recognition, cued recall, free recall) that examine both item and context memory, tapping into different types of memory representations (item-specific, cue–target relational, target–target relational, context–target relational, respectively) to examine potential memory mechanisms that underlie the effects of generation constraint. Importantly, we do so while measuring and controlling for differences in the responses participants generate at encoding (i.e., idiosyncratic responses). We make three predictions (see Table 1 for a summary of predictions for each memory test). First, for item recognition, we expect no relationship between generation constraint and memory, as prior work has shown that generation constraint does not strongly influence item-specific processing (McCurdy, Sklenar, et al., 2020). Second, for cued recall, we expect to find a negative relationship between constraint and memory, such that as constraint decreases, memory performance will increase, in line with prior work showing the strongest effects of generation constraint on cued recall (McCurdy et al., 2017; McCurdy, Sklenar, et al., 2020). For free recall, given the limited work using this measure,

we make no predictions regarding the effects of generation constraint; however, findings from this measure will provide more evidence regarding the specificity of generation constraint effects on relational processing, particularly of target–target relationships. Third, for context memory, prior work has suggested fewer generation constraints provide robust improvements in source memory performance (McCurdy et al., 2017, 2019; McCurdy, Sklenar, et al., 2020), thus, we expect a negative relationship between generation constraint and context memory for spatial location. However, prior work has yet to examine the influence of generation constraint on memory for spatial location, therefore the results of this study will provide evidence about the extent generation constraint enhances context–target relational processing broadly, or more specifically for only certain types of contextual details.

Method

Participants

A total of 107 University of Illinois at Chicago undergraduate students participated in this study. Participants were randomly assigned to one of three memory test conditions (recognition, cued recall, free recall), with 36, 38, and 33 participants in each, respectively. An a priori power analysis was conducted using G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) for logistic regression (two-tailed z test). We estimated an effect size from a recent meta-analysis reporting the probability of remembering items from lower-constraint compared with higher-constraint generation tasks (lower-constraint = .642, higher-constraint = .528; see McCurdy, Viechtbauer, et al., 2020). This power analysis recommended a total of 1,511 observations to detect an effect for the generation constraint predictor using a power level (1- error prob.) of .95. From our samples, we collected a total of 1,588; 1,695; 1,531 usable observations¹ from the recognition ($n = 36$), cued recall ($n = 38$), and free recall ($n = 33$) groups, respectively, suggesting that our samples were adequate to detect an effect of generation constraint in each group. Each participant gave informed written consent in accordance with the University of Illinois at Chicago Institutional Review Board, and received course credit for their participation.

Stimuli

Seventy-two highly associated cue–target word pairs were used as stimuli. Word pairs were selected from the University of South Florida Free Association Norms (Nelson, McEvoy, & Schreiber,

2004) based on the following criteria: the forward cue–target association strength (FSG) was greater than .40 ($M = .60$, $SD = .12$, range: .40–.89), all normed target words were exactly five letters long, and the target words were uniquely associated with their respective cue (i.e., a target word of any pairing was not a strong associate of more than one cue in the list). The word pairs were fully counterbalanced across conditions to create 18 different word lists. This counterbalance procedure ensured that each word pair occurred in all six missing letter conditions (0–5 target letters missing) exactly once in each of the location conditions (left, right) and as a “new” distractor item for the recognition memory test. The cued recall and free recall groups did not use the “new” distractor items in their respective word list versions. We used the same counterbalance list versions regardless of the type of memory test to minimize differences between the materials across the different memory test groups.

Procedure

The study procedure consisted of three phases: training, encoding, and retrieval. In the training phase, participants were trained on *both* the encoding and retrieval phases of the experiment to ensure participants were aware of their task instructions and the nature of the memory test (i.e., intentional encoding). Training consisted of three examples explaining the encoding instructions, showing an example trial with zero letters (read), three letters, and five (all) letters missing. After the examples, participants went through a practice encoding phase with a total of six practice trials mimicking the actual experiment (one trial of each letters missing condition, in a random order). The examples and training word pairs were the same for all participants, and the words were not strongly related to any of the items used in the actual experiment. Training for the retrieval phase differed based on the memory test group the participant was assigned to (e.g., participants assigned to the recognition test group were trained on the retrieval procedure for the recognition test). All participants were given one example trial, with instructions on how to respond (based on the type of memory test). The practice for the recognition memory test group consisted of twelve practice trials in the recognition group (six “old” trials, six “new” trials). The cued recall group was given six practice trials (using the practice trials from encoding). The free recall group was given 1 minute to recall as many words as possible from the encoding practice phase. Both item and context memory measures were included in the example and practice trials to make participants aware of the details they would be asked to remember. Following the training, participants were given the opportunity to ask questions and gave verbal confirmation that they understood the task instructions. No participants required additional instructions.

After training, participants began the encoding phase. In the encoding phase, participants saw a total of 48 cue–target

¹ This represents the number of observations after removing “don’t know” responses at retrieval, which were not analyzed in this study (see Footnote 3). This includes “blank” responses for the cued recall memory test group, where participants were instructed to leave their response blank if they were unsure.

word pairs across six conditions (eight word pairs each condition). The six conditions corresponded to the generation constraint manipulation. For each word pair, the cue word was shown on the screen and the target word was manipulated to show all (5), some (1–4), or none (0) of the letters of the word. For example, eight word pairs were shown with zero letters given (e.g., open-____),² eight word pairs were shown with one letter given (e.g., back-f____), and so on, including eight word pairs with all five letters given of the target word (e.g., saddle–horse), serving as the “read” control task. Word pairs with a partial number of letters given of the target word were always created by replacing letters with an underscore, starting from the end of the word moving toward the beginning of the word, to make a word stem of the target (e.g., bank–mo____, for a target word with two letters given). Word pairs were presented in a pseudorandom order, such that consecutive items were never from the same condition (e.g., a trial with zero target letters given was never followed by another trial with zero target letters given). After encoding, participants were given a nonverbal filler task (digit symbol replacement) for 2 minutes before moving on to the retrieval phase. During the filler task for the recognition memory test group, the experimenter uploaded participant-generated responses into E-Prime to be used in the recognition test. Importantly, this process allowed the recognition test to display participant-generated responses as “old” words at recognition, even when the participant’s response did not match the normed target. To be clear, for all memory tests, participants were explicitly instructed in the training phase (and reminded again at the beginning of the retrieval phase) that the words they self-generated or read in the encoding phase were the to-be-remembered items (i.e., should be classified as “old,” or recalled during the memory test).

In the retrieval phase, participants were given either a recognition, cued recall, or free recall memory test. Participants in the *recognition test group* were shown a total of 72 words (48 “old” target words from the experiment, and 24 “new” distractor words not shown at encoding), one at a time, in a different random order than they saw at encoding. For each word, participants were asked to make two judgments (item, context). For the item memory judgment, they were asked to decide whether the word was old, new, or whether they were unsure (“don’t know”), by pressing the V, B, or M key on a keyboard, respectively, as we have done before (Leshikar, Cassidy, & Gutchess, 2016; Leshikar & Duarte, 2012, 2014; Leshikar, Dulas, & Duarte, 2015; Leshikar & Gutchess, 2015;

Leshikar, Park, & Gutchess, 2015). If participants responded “old” to the item memory judgment, they were asked to make a context memory judgment for that same word. For the context memory judgment, participants decided whether the word was presented on the left or right of the screen at encoding, or whether they were unsure (“don’t know”), by pressing the V, B, or M key on a keyboard, respectively.

Participants in the *cued recall group* were shown all 48 cue words from the encoding phase, one at a time, in a different random order from the encoding phase. In the cued recall test, participants were instructed to “type the word that was paired with this word in the experiment” and to press the ENTER key to submit their response. If they were unsure, participants were instructed to leave their response blank and press the ENTER key. After giving their item memory response, they were shown the same cue word, and asked to make the same context memory judgment as the recognition group. Specifically, they were asked to respond whether that cue word (and its paired target word) was presented on the left, right, or whether they were unsure (“don’t know”), by pressing the V, B, or M key on a keyboard, respectively. Both the item memory and context memory responses were self-paced.

Participants in the *free recall group* were given a paper and pencil with two columns of blank lines, and the letters “L” or “R,” or “don’t know” to the right of each blank line. Participants were instructed to write down as many of the *target* words they could remember from the encoding phase, and for each word to circle which side of the computer monitor they thought the word was presented on, left (“L”), right (“R”), or whether they were unsure (“don’t know”). Participants were given as long as they needed to recall as many words as they could, with the caveat that they were required to attempt recall for at least 5 minutes. After 5 minutes, they were instructed to inform the experimenter whenever they felt they could no longer remember any more words (only one participant used more than 5 minutes, but did not recall any additional words in that time). Following the retrieval phase, participants were given credit for their participation and debriefed.

Analyses

We ran a total of six mixed-effects logistic (logit) regression models: one for item memory for each memory test (recognition, cued recall, free recall), and one for context (location) memory for each memory test (recognition, cued recall, free recall), separately. The dependent variable for the item memory analyses was a binary memory performance variable for each target word, coded as either 0 (incorrect) or 1 (correct). For the context (location) memory analyses, given that item and context memory are known to be confounded, we used a conditional measure that reduces the influence of item memory performance on context memory by only considering

² Participants were not aware that the target word was to be five letters long, however, the length of the underscores were shortened by the number of letters missing (five underscores for five letters missing; four underscores for four letters missing, etc.). Given that a goal of this study was to measure whether fewer constraints enhance memory, we opted to not explicitly state that each target must be five letters in length because this procedure would introduce more “constraints” on the participant.

context memory performance for items that were correctly remembered (CSIM; Murnane & Bayen, 1996). This measure has been widely used in prior studies examining the generation effect for context memory (Hashtroudi, Johnson, & Chrosniak, 1989; Johnson, Raye, Foley, & Foley, 1981; Marsh, 2006; Marsh, Edelman, & Bower, 2001; Mulligan, 2004; Rabinowitz, 1989). Using this conditional measure, the dependent variable for the context memory analyses was a binary memory performance variable for the location of the word pair, conditional upon that trial being correctly remembered in the item memory measure, coded as either 0 (incorrect) or 1 (correct). We report both the conditional and unconditional context (location) memory response rates in Tables 2 and 3. Trials where the participant gave a “don’t know” response were removed from all analyses (item and context memory).³

For each of the six models, we included three predictors based on our a priori theoretical interests: *Condition* (categorical: read, generate), *GenConstraint* (continuous: 0–5 letters given), *NormMatchTarget* (categorical: idiosyncratic responses, norm-matched responses). Including the *NormMatchTarget* predictor importantly allows us to parse out the variance in memory performance that is due to generation constraint, and the variance that is due to differences in generated items. In other words, including this predictor in the model with the generation constraint predictor enables the assessment of generation constraint effects on memory while controlling for item-selection effects. The regression equation for this model is shown in Equation 1. Unless otherwise noted, this model was used to assess both item and context memory effects for each memory test group (item recognition, cued recall, free recall), respectively:

$$\text{logit}(p) = \beta_0 + \beta_1 \text{Condition} + \beta_2 \text{GenConstraint} + \beta_3 \text{NormMatchTarget} + \varepsilon. \quad (1)$$

Principle components analysis (PCA) was used to determine the best-fitting random structure for our models. This process revealed that allowing random intercepts for both participants and items was the best-fitting and most parsimonious model in each data set; therefore, we used this random structure for each of the models we report. This random structure (random-intercepts only) allows each participant and each word pair to have a unique intercept, which captures variability in overall memory performance across different participants and word pairs. The benefit of using a mixed-effects model is that these models account for variance across the

different participants and individual stimuli (i.e., word pairs). In contrast, in standard ordinary least squares (OLS) regressions, this variance is considered fixed for all participants and word pairs, requiring the assumption that all participants and word pairs are equal in their baseline memory performance. This assumption would rarely be true, of course, therefore mixed-effects models often provide a more nuanced estimate of the memory effects of interest by accounting for this random variance that comes from participants and stimuli (word pairs).

Before reporting the results, we now describe what each coefficient in the model (Equation 1) tells us about the data to help interpretation. The *Condition* (β_1) coefficient indicates the difference in log odds of correctly remembering an item (or location) between all “generated” items compared with the read items (i.e., the standard generation effect), while controlling for both generation constraint (*GenConstraint*) as well as whether the participant-generated response matched the expected normed target (*NormMatchTarget*). We can use this coefficient to investigate the standard generation effect (generate greater than read). The *GenConstraint* (β_2) coefficient indicates the change in log odds of correctly remembering an item (or location) uniquely associated with a one unit increase in the number of target letters given (i.e., the effect of increasing generation constraint), controlling for the standard generation effect (*Condition*) and also controlling for whether the participant-generated response matched the expected normed target (*NormMatchTarget*). The *NormMatchTarget* (β_3) coefficient indicates the change in log odds of correctly remembering an item (or location) for participant-generated responses that *matched* the expected normed target relative to participant-generated responses that *did not match* the normed target (i.e., an idiosyncratic response). This variable is critical because it allows us to measure and control for the influence of item-selection effects in our data, and to more accurately model the unique effects of the *Condition* and *GenConstraint* variables. Finally, the ε represents the residual error between observed and predicted values.

All analyses were conducted in R (Version 3.6.1; R Core Team, 2018), using the *lme4* (Version 1.1-21; Bates, Mächler, Bolker, & Walker, 2015), *effects* (Version 4.1-0; Fox & Weisberg, 2019), *ggplot2* (Version 3.1.1; Wickham, 2016), *rePsychLing* (Version 0.0.4; Baayen, Bates, Kliegl, & Vasisht, 2015), and *emmeans* (Version 1.3.4; Lenth, 2019) packages.

Results

The observed mean response rates for each memory test are reported in Tables 2 and 3. In the following sections, we first report a correlation analysis to examine the relationship between generation constraint and the likelihood of generating an idiosyncratic target item. The proportion of idiosyncratic responses for

³ “Don’t know” responses made up a small proportion of the total number of responses (see Table 2), thus for simplicity we removed these trials instead of including them, which would require a nonbinary (multinomial) regression model.

Table 2 Observed means and standard deviations (in parentheses) of response rates for item, conditional context (location), and unconditional context (location) memory by number of target letters given (ltrs given) at encoding (0–5) and memory test (recognition, cued recall)

Recognition											
Item memory				Conditional context (location) memory				Unconditional context (location) memory			
Ltrs given	Old	New	DK	Ltrs given	Correct	Incorrect	DK	Ltrs given	Correct	Incorrect	DK
0	.83 (.20)	.15 (.19)	.02 (.05)	0	.69 (.23)	.28 (.24)	.03 (.06)	0	.59 (.22)	.24 (.18)	.17 (.19)
1	.86 (.19)	.12 (.18)	.02 (.04)	1	.69 (.26)	.26 (.27)	.05 (.10)	1	.63 (.23)	.23 (.21)	.14 (.18)
2	.89 (.13)	.09 (.12)	.02 (.07)	2	.74 (.20)	.23 (.18)	.03 (.10)	2	.66 (.21)	.21 (.17)	.13 (.16)
3	.89 (.15)	.09 (.12)	.02 (.08)	3	.66 (.21)	.30 (.21)	.04 (.10)	3	.59 (.23)	.28 (.20)	.13 (.18)
4	.84 (.19)	.13 (.17)	.03 (.07)	4	.67 (.20)	.29 (.20)	.04 (.12)	4	.57 (.21)	.25 (.19)	.18 (.21)
5 – read	.62 (.22)	.35 (.21)	.03 (.08)	5 – read	.59 (.30)	.38 (.31)	.03 (.18)	5 – read	.37 (.21)	.24 (.22)	.39 (.24)
New	.13 (.13)	.83 (.15)	.04 (.06)								
Cued recall											
Item memory				Conditional context (location) memory				Unconditional context (location) memory			
Ltrs given	Correct	Incorrect	Blank	Ltrs given	Correct	Incorrect	DK	Ltrs given	Correct	Incorrect	DK
0	.92 (.12)	.07 (.12)	.01 (.03)	0	.76 (.21)	.19 (.03)	.05 (.10)	0	.75 (.19)	.19 (.18)	.06 (.09)
1	.85 (.16)	.10 (.13)	.05 (.11)	1	.80 (.20)	.15 (.19)	.05 (.09)	1	.76 (.20)	.17 (.19)	.07 (.11)
2	.87 (.16)	.07 (.11)	.06 (.12)	2	.77 (.19)	.18 (.16)	.05 (.08)	2	.74 (.20)	.18 (.16)	.08 (.12)
3	.89 (.14)	.07 (.12)	.04 (.09)	3	.77 (.22)	.17 (.17)	.06 (.11)	3	.73 (.19)	.18 (.17)	.09 (.11)
4	.88 (.16)	.07 (.12)	.05 (.11)	4	.72 (.19)	.20 (.19)	.08 (.13)	4	.68 (.19)	.20 (.17)	.12 (.16)
5 – read	.81 (.21)	.13 (.19)	.06 (.12)	5 – read	.61 (.23)	.24 (.18)	.15 (.22)	5 – read	.55 (.23)	.23 (.17)	.23 (.24)

Note. “Ltrs given” represents generation constraint manipulation (more letters given = higher generation constraint). “5” target letters given represents the standard read (control) condition because each target word was 5 letters in length. DK = “don’t know.” Conditional location = proportion of location memory responses out of correctly remembered items. Data are aggregated at the participant level (recognition: $n = 36$; cued recall: $n = 38$)

each generation constraint condition by memory test group is shown in Table 4. Then, we report the logistic regression analyses for each memory test group (recognition, cued recall, free recall), separated by memory type (item memory, context [location] memory). The coefficient summary tables and corresponding odds ratios from the logistic regression analyses are reported in Table 5. We report item and context memory for each memory test group starting first with the recognition group, followed by the cued recall, and then free recall groups.

Correlation analysis

We conducted three point-biserial correlation analyses to examine the relationship between generation constraint and the number of idiosyncratic target items generated in each data set (recognition, cued recall, free recall). These analyses showed a significant negative relationship in all three of our data sets (recognition: $r = -.336$, $p < .001$; cued recall: $r = -.384$, $p < .001$; free recall: $r = -.351$, $p < .001$), indicating that

Table 3 Observed means and standard deviations (in parentheses) of accuracy for the free recall memory test, by number of target letters given at encoding, and memory type (item memory, conditional context [location] memory, unconditional context [location] memory)

Free recall			
Ltrs given	Item memory	Conditional context (location) memory	Unconditional context (location) memory
0	.23 (.14)	.94 (.14)	.21 (.14)
1	.23 (.18)	.91 (.14)	.20 (.17)
2	.23 (.13)	.89 (.32)	.20 (.14)
3	.21 (.14)	.76 (.38)	.15 (.12)
4	.17 (.12)	.87 (.30)	.14 (.11)
5 – read	.10 (.11)	.80 (.35)	.08 (.09)

Note. “Ltrs given” represents generation constraint manipulation (more letters given = more generation constraint). “5” letters given is equivalent to the standard read control condition because each target word was 5 letters in length. Conditional Location = the proportion of correctly remembered locations out of items that were correctly recalled. Data are aggregated at the participant level ($n = 33$)

Table 4 Mean proportion and standard error (in parenthesis) of idiosyncratic target responses by memory test group and missing letter condition

Ltrs given	Recognition	Cued recall	Free recall
0	.32 (.03)	.40 (.03)	.35 (.03)
1	.19 (.02)	.24 (.03)	.23 (.03)
2	.10 (.02)	.13 (.02)	.19 (.02)
3	.04 (.01)	.08 (.02)	.07 (.02)
4	.02 (.01)	.02 (.01)	.01 (.01)
5 – read	.00 (.00)	.00 (.00)	.00 (.00)

decreasing generation constraints was associated with an increase in generating an idiosyncratic target at encoding, as expected.

Logistic regression analyses

Recognition

Item memory For the recognition group data, we ran two regression models: We first ran the full model with *Condition*, *GenConstraint* and *NormMatchTarget* as predictors (Equation 1). However, the observed mean proportions shown in Table 2 suggested a possible quadratic relationship between generation constraint and memory performance in these data. Therefore, we ran an additional model on the item recognition data to examine this relationship by adding a quadratic generation constraint predictor to the model (see Equation 2):

$$\begin{aligned} \text{logit}(p) = & \beta_0 + \beta_1 \text{Condition} + \beta_2 \text{GenConstraint} \\ & + \beta_3 \text{GenConstraint}^2 + \beta_4 \text{NormMatchTarget} \\ & + \varepsilon. \end{aligned} \quad (2)$$

The model with the linear *GenConstraint* predictor (Equation 1) provided a significantly better fit than the null model (a model with no predictors), $\chi^2(3) = 105.89$, $p < .001$ (see Supplemental Table S1 for fixed effects summary table of this model). However, the model including a quadratic predictor of generation constraint (Equation 2) fit significantly better than both the null model, $\chi^2(4) = 110.57$, $p < .001$, and the linear model, $\chi^2(1) = 4.69$, $p = .030$. Therefore, given the significantly better fit to the data, we interpret the model with the quadratic predictor (see Table 5 for fixed effects summary table). The results showed a significant effect for *Condition* ($\beta = 0.90$, $SE = 0.43$, $p = .038$), with a positive coefficient indicating that generated items were more likely to be remembered relative to read controls, in line with the standard generation effect. The effect for *NormMatchTarget* ($\beta = 0.91$, SE

$= 0.25$, $p < .001$) was also significant, with a positive coefficient indicating that participant-generated responses that matched the normed expected target (norm-matched targets) were significantly more likely to be remembered compared with participant responses that did not match the normed target (idiosyncratic targets). Interestingly, for *GenConstraint*, both the linear coefficient ($\beta = 1.68$, $SE = 0.22$, $p = .045$) and quadratic coefficient ($\beta = -0.12$, $SE = 0.05$, $p = .027$) were significant. This finding indicates a positive relationship between memory and generation constraint, but with a significant downward curve in the data. Figure 1 shows the predicted probability of a correct response based on this model as a function of generation constraint (number of target letters given). The figure depicts an inverted U-shape trend, with a large drop off for the read items (i.e., 5 letters given) indicative of the robust standard generation effect. This finding suggests that generation constraint influences memory, even when controlling for whether the participant generated the normed target, but the relationship between generation constraint and memory on a recognition test is curvilinear. Given that recognition tests are thought to tap into item-specific memory representations, these data suggest that generation constraint influences item-specific processing at encoding.

Context (location) memory For context (location) memory, the full model (Equation 1) did not fit the data better than a null model (with no predictors), $\chi^2(3) = 6.33$, $p = .096$, indicating that these three predictors did not account for a significant amount of variance in the data.⁴ In other words, we found no significant generation effect, and no significant relationship between the *GenConstraint* or *NormMatchTarget* variables and the likelihood of remembering the encoded location of a word pair for this model. The predicted probabilities from this model are shown in Fig. 1. In the figure, the flat line across all levels of generation constraint (i.e., target letters given) shows the lack of relationship between generation constraint and context memory. This finding suggests that generation constraint did not influence context–target relational processing in these data.

Cued recall

Item memory In the cued recall data, the full model (Equation 1) fit significantly better than a null model with no predictors, $\chi^2(3) = 105.45$, $p < .001$. The model showed that when controlling for *GenConstraint* and *NormMatchTarget*, the effect of *Condition* was nonsignificant ($\beta = 0.33$, $SE = 0.31$, $p = .293$), indicating

⁴ Like item memory, we also ran a quadratic model (Equation 2) for the context memory data, to test for the presence of a quadratic relationship between constraint and context memory. The quadratic model did not fit better than the null model (with no predictors), $\chi^2(4) = 7.37$, $p = .118$, providing converging evidence that there were no significant effects for context memory in these data.

Table 5 Logistic (logit) regression fixed effects summary tables by memory test (recognition, cued recall, free recall) and memory type (item, conditional context [location]memory)

Recognition									
Item memory					Conditional context (location) memory*				
Predictor	β (SE)	z	p	OR (e^β)	Predictor	β (SE)	z	p	OR (e^β)
Intercept	0.44 (0.47)	0.94	.349	1.55	Intercept	0.66 (0.37)	1.77	.076	1.93
Condition	0.90 (0.43)	2.08	.038	2.46	Condition	0.36 (0.24)	1.53	.126	1.44
GenConst	1.68 (0.22)	2.00	.045	1.55	GenConst	−0.04 (0.05)	−0.67	.501	0.96
GenConst ²	−0.12 (0.05)	−2.21	.027	0.89	NMT	0.14 (0.23)	0.61	.541	1.15
NMT	0.91 (0.25)	3.70	<.001	2.49					
Cued recall									
Item memory					Conditional context (location) memory				
Predictor	β (SE)	z	p	OR (e^β)	Predictor	β (SE)	z	p	OR (e^β)
Intercept	1.52 (0.46)	3.30	.001	4.60	Intercept	1.08 (0.36)	2.99	.003	2.95
Condition	0.33 (0.31)	1.05	.29	1.39	Condition	0.33 (0.24)	1.38	.168	1.40
GenConst	−0.39 (0.09)	−4.16	<.001	0.67	GenConst	−0.08 (0.06)	−1.38	.167	0.92
NMT	2.68 (0.29)	9.34	<.001	14.58	NMT	0.40 (0.24)	1.68	.094	1.49
Free recall									
Item memory					Conditional context (location) memory*				
Predictor	β (SE)	z	p	OR (e^β)	Predictor	β (SE)	z	p	OR (e^β)
Intercept	−1.91 (0.36)	−5.38	<.001	0.15	Intercept	2.93 (0.99)	2.96	.003	18.68
Condition	0.71 (0.27)	2.61	.009	2.03	Condition	−0.36 (0.69)	−0.52	.605	0.70
GenConst	−0.10 (0.05)	−1.86	.062	0.91	GenConst	−0.33 (0.16)	−2.14	.033	0.72
NMT	0.02 (0.20)	0.08	.937	1.02	NMT	0.25 (0.57)	0.43	.668	1.28

Note. “Condition” = effect of generate compared with read items; “GenConst” = generation constraint; “GenConst²” = quadratic effect of generation constraint. “NMT” = norm-matched target; SE = standard error; OR = odds ratio. Betas (β) represent change in log odds of correctly remembering an item, or location. *Models for conditional location did not fit data significantly better than a null model with no predictors and thus should be interpreted with caution

that generated and read items did not significantly differ in their likelihood of being recalled (i.e., no significant generation effect). This finding, however, was likely due to the robust effects of the *GenConstraint* and *NormMatchTarget* variables in the model. The effect of *GenConstraint* was significant ($\beta = -0.39$, $SE = 0.09$, $p < .001$), indicating a negative relationship between generation constraint and the likelihood of correctly recalling that item (i.e., more constraint yields poorer memory), even after controlling for whether the participant-generated response matched the normed target or not. The *NormMatchTarget* coefficient was also significant ($\beta = 2.68$, $SE = 0.29$, $p < .001$), indicating that generating the norm-matched target led to a greater likelihood of remembering that item compared with idiosyncratic targets. It is worth noting, however, that memory performance in these data was near ceiling, thus the effects in this model may be muted to some degree. Although, given that we have seen this effect consistently for cued recall in prior work (McCurdy et al., 2017; McCurdy, Sklenar, et al., 2020) and the model captured a significant amount of variance in the data, the influence of any potential ceiling effects seems minimal on the interpretations made here. Figure 2 shows the predicted probabilities of correctly recalling

an item based on this model, and depicts the strong downward sloping effect of increasing generation constraint on memory performance in cued recall.⁵ This finding is in line with previous work showing that generation constraint increases the generation effect through enhanced cue–target relational processing at encoding, as measured by cued recall (McCurdy, Sklenar, et al., 2020).

Context (location) memory For the context (location) memory data in the cued recall group, the full model (Equation 1) fit significantly better than the null model, $\chi^2(3) = 10.78$, $p = .013$; however, the model revealed no significant effects for *Condition* ($\beta = 0.33$, $SE = .24$, $p = .168$), *GenConstraint* ($\beta = -0.08$, $SE = .06$, $p = .167$), or *NormMatchTarget* ($\beta = 0.40$, $SE = .24$, $p = .094$). The predicted probabilities from this

⁵ To examine whether the effect of generation constraint differs between norm-matched and idiosyncratic targets, we ran an additional model including the interaction of the *GenConstraint* and *NormMatchTarget* variables. This model revealed the interaction was not significant ($\beta = 0.12$, $SE = .19$, $p = .52$), indicating that the influence of generation constraint did not significantly differ based on whether the participant-generated response matched the normed target in these data.

Fig. 1 Predicted probability of correctly remembering an item and location for the *recognition* memory test group as a function of generation constraint (i.e., number of letters given of the target word at encoding). Context (location) memory is conditional upon correct recognition of the item. Error bars represent 95% confidence intervals. *Note.*

The conditional context (location) memory model did not significantly fit the data; thus, the pattern of results should be interpreted with caution. More letters given = more generation constraint. Generation constraint level “5” represents the standard “read” control condition

model are shown in Fig. 2. This finding provides converging evidence with our recognition data suggesting that generation constraints did not strongly influence context–target relational processing.

Free recall

Item memory In the free recall data, the full model (Equation 1) fit significantly better than a null model with no predictors, $\chi^2(3) = 27.58$, $p < .001$. This model revealed a significant effect of *Condition* ($\beta = 0.71$, $SE = .27$, $p = .009$), indicating that generated items were significantly more likely to be recalled than read items (i.e., a generation effect). The effect of *GenConstraint* just missed significance ($\beta = -0.10$, $SE = .05$, $p = .062$), but showed the trend that the likelihood of recalling an item decreased for each additional letter given of the target word (i.e., increasing constraint). The effect of *NormMatchTarget* was not significant ($\beta = 0.02$, $SE = .20$, $p = .937$), indicating that there were no significant differences in the likelihood to recall an item between norm-matched and idiosyncratic target items in these data. Figure 3 plots the predicted probability of correctly recalling an item as a function of generation constraint, which shows a nonsignificant negative trend of increasing generation constraint on memory performance. Given that free recall tests are thought to be sensitive to target–target relational information (Burns, 2006), this finding suggests that generation constraints did

not strongly influence target–target relational processing at encoding in these data.

Context (location) memory For the context (location) memory data in the free recall group, the full model (Equation 1) was not a significantly better fit than a null model with no predictors, $\chi^2(3) = 5.79$, $p = .122$, indicating that there were no significant effects of the *Condition*, *GenConstraint*, or *NormMatchTarget* variables on context memory in these data. The predicted probabilities shown in Fig. 3 depict a negative trend between generation constraint and the likelihood of recalling the encoded location of an item, but the variance around these data are too large to interpret this effect. Similar to the recognition and cued recall data, this finding suggests generation constraint did not significantly influence context–target relational processing of location information.

Discussion

The primary goal of this study was to examine the relationship between generation constraint and memory performance on three different memory tests (measuring both item and context memory), as a way to investigate the memory mechanisms underlying the effects of generation constraint. There were three major findings in this study. First, we found that generation constraint influences memory performance even after

Fig. 2 Predicted probability of correctly recalling an item and location for the *cued recall* memory test group as a function of generation constraint (i.e., letters given of the target word at encoding). Context (location) memory is conditional upon correctly recalling the item. Error bars

represent 95% confidence intervals. *Note.* More letters given = higher generation constraint. Generation constraint level “5” represents the standard “read” control condition

controlling for item-selection effects. Second, we found that the relationship between generation constraint and memory is

dependent on the type of memory test used to assess performance. Specifically, we found a curvilinear relationship

Fig. 3 Predicted probability of correctly recalling an item and location for the *free recall* memory test group as a function of generation constraint (i.e., letters given of the target word at encoding). Context (location) memory is conditional upon correctly recalling the item. Error bars represent 95% confidence intervals. *Note.* The conditional context (location)

memory model did not significantly fit the data; thus, the pattern of results should be interpreted with caution. More letters given = higher generation constraint. Generation constraint level “5” represents the standard “read” control condition

between constraint and memory in a recognition test, a negative, linear relationship in a cued recall test, and no significant relationship in a free recall test. Third, we found no significant evidence that generation constraint influenced context memory for spatial location in this study. We discuss the implications of these findings in the sections that follow.

Prior work has typically avoided the use of lower-constraint generation tasks because fewer constraints increase the likelihood a participant will generate idiosyncratic responses that are different from the responses being learned in higher-constraint and read control tasks. This creates the potential for item-selection effects, making it difficult to determine whether any effect on memory is due to generation constraints or due to the differences in memorability of the (idiosyncratic) items themselves. Indeed, our correlation analysis between generation constraint and whether the participant generated an idiosyncratic response confirmed this idea in our data (fewer generation constraints was associated with a greater number of idiosyncratic responses). In this study, however, we found evidence that generation constraint affects memory even after controlling for item-selection effects, which argues against the idea that item-selection effects fully obscure the ability to empirically study the influence of generation constraint. This is an important finding, and one that has implications for the field. First, prior work has generally avoided studying generation constraints, leaving this influential factor relatively underexamined compared with other notable moderators of the generation effect (McCurdy, Viechtbauer, et al., 2020). The results of the current study, however, provide compelling data that generation constraint has a measurable influence on the magnitude of the generation effect, which is consistent with past work (Fiedler et al., 1992; Gardiner et al., 1985; McCurdy et al., 2017, 2019; McCurdy, Sklenar, et al., 2020; McCurdy, Viechtbauer, et al., 2020). To better understand the generation effect, future work should carefully consider the role of generation constraint in designing empirical studies on the generation effect. Second, our results generally showed that generating an idiosyncratic response at encoding *reduced* the likelihood of that item being subsequently remembered. This effect (poorer memory for idiosyncratic responses) was sometimes quite large, such as in the cued recall data where norm-matched responses were nearly 15 times more likely to be recalled than idiosyncratic responses (see odds ratios in Table 5). Thus, although it may seem that generating an idiosyncratic response at encoding could make that item more memorable, our data strongly refute this idea. In conjunction with our data showing that fewer generation constraints are associated with more idiosyncratic responses, this finding (idiosyncratic responses are *less* likely to be remembered) importantly suggests that the effects of generation constraint (fewer constraints increase the generation effect) may be underestimated in prior work that has not controlled for potential item-selection effects (Fiedler et al., 1992; Gardiner

et al., 1985; McCurdy et al., 2017, 2019; McCurdy, Sklenar, et al., 2020). As a whole, the present data highlight that fewer generation constraints can increase the generation effect, and that constraint is a factor worth considering in future work. Importantly, this study demonstrates a way to examine the unique effect of generation constraints on memory while controlling for potential item-selection effects by using a mixed-effects logistic regression analysis, which may allow more progress in this domain.

Another novel contribution of this study is the delineation of a curvilinear relationship between generation constraint and item recognition performance. Prior investigations examining the effects of generation constraint have consistently reported that constraint does not strongly influence memory on item recognition measures (McCurdy et al., 2017, 2019; McCurdy, Sklenar, et al., 2020). Critically, though, these studies only compared a lower-constraint generation task to a higher-constraint generation task. In some ways, our current data concur with this prior work by showing that the lowest generation constraint condition (i.e., zero target letters given) led to nearly identical memory performance as the highest generation constraint condition (i.e., four target letters given) with observed means of .83 and .84, respectively (see Table 2). However, the current study also provides data on a critical component that is missing in prior studies—memory performance on tasks with intermediate levels of constraint. Medium-constraint generation conditions (i.e., two and three letters given) showed a clear advantage on the recognition memory test over both lower-constraint and higher-constraint generation. This finding suggests that generation constraint influences memory performance as measured by a recognition test, but dichotomous manipulations of generation constraint were too imprecise to detect this relationship before now. The inverted U-shape trend in the item recognition data depicted in Fig. 1 is in line with the concept of desirable difficulties (Bjork, 1994). This concept represents one way to understand the relationship of generation constraint and memory, at least in our recognition data. Generation constraint is an encoding difficulty that, when dosed appropriately, can lead to measurable long-term memory improvements (see McDaniel & Butler, 2011, for a related discussion). Based on the evidence provided in this study, using a (word stem) generation task with two or three letters given (when five letter target words are used) may be the optimal dose of constraint when using a recognition memory test (see Table 2). These findings have implications for real-world applications of the generation effect, such as in the classroom (Metcalf & Kornell, 2007; Richland, Bjork, Finley, & Linn, 2005). For example, our data suggest that an effective study strategy might be to use an “optimal” level of constraint when employing the generation effect to improve long-term retention of educational materials. These data may also have implications for the retrieval practice effect (or the testing effect;

Roediger & Karpicke, 2006; Rowland, 2014), an increasingly popular learning strategy that is recommended for enhancing long-term retention in the classroom (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). It is possible that *constraints* on a practice test strongly influence the memory benefits of the testing effect as well, which could be an exciting area of future research.

Another goal of this study was to examine the relationship between generation constraint and performance on different memory tests to better understand the underlying memory mechanism(s) of generation constraint effects. Prior work has attributed the standard generation effect to enhanced item-specific and relational processing at encoding for generate tasks relative to read tasks (Hirshman & Bjork, 1988; McDaniel et al., 1988). Overall, in our item memory measures we found a curvilinear relationship between constraint and memory in item recognition, a negative, linear relationship in cued recall, and no significant relationship in free recall. Given that the memory tests used in this study are thought to tap into different types of encoding processes (Begg, 1978; Begg et al., 1989; Burns, 2006; Hirshman & Bjork, 1988), the different pattern of results across the different memory tests allows us to make some inferences about whether generation constraint effects on memory can be tied to item-specific or relational processing.

Starting with item recognition (a measure thought to be sensitive to item-specific information; Burns, 2006), our results suggest that increased generation constraints generally led to increased item-specific processing. One likely explanation is that when participants were given more letters of the target item (i.e., increasing constraints), participants shifted their attention more toward processing features of the target item itself (i.e., increased item-specific processing), leading to increased item recognition performance. The curvilinear relationship found in the recognition data showing a drop-off in performance for the highest generation constraint condition (i.e., four target letters given), however, suggests that this increase in item-specific processing only occurs up to a certain point of constraint, in line with desirable difficulties (Bjork, 1994). Giving all but one of the letters of the target word may have led to shallower processing overall, similar to the read control condition (i.e., five letters given). This idea that generation constraint acts to shift the focus of participants processing is further supported by the finding of a noticeable drop in recognition performance in the lowest constraint condition (i.e., zero target letters given). When there is less information given about the target item (i.e., fewer generation constraints), participants must rely more heavily (if not exclusively) on the cue word to help them generate a target, thus leading to less reliance on item-specific information, and more on “cue–target relational” information.

The idea that generation constraints influence memory by shifting the focus of encoding processing is further supported by our findings from the cued recall memory test (a measure

thought to be sensitive to cue–target relational processing; Hirshman & Bjork, 1988). We found a significant negative, linear relationship between generation constraint and cued recall, indicating that fewer constraints generally led to increased cue–target relational processing. This result is in line with the theory that generation constraints shift the focus of the information participants rely on to help them generate at encoding. As the number of letters given for the target word decreased, participants become increasingly likely to rely on the cue word to decide which word to generate. In other words, reducing generation constraints shifts participants to focus more on the relationship between the cue and target word (i.e., cue–target relationship), and less on the target item (i.e., item-specific processing). Taken together with the recognition data discussed above, our findings provide strong evidence that generation constraint influences memory through a shift in encoding processing—increasing generation constraint acts to enhance item-specific processing, while decreasing generation constraint acts to enhance cue–target relational processing (relative to reading).

As for free recall, the absence of a significant relationship between generation constraint and memory suggest that constraint did not strongly influence “target–target relational” processing at encoding, at least in these data. Free recall tests are thought to be sensitive to relational information among multiple items in a list (Burns, 2006; McDaniel et al., 1990). One possible explanation for our data is that because there was no noticeable preexisting relationship among the different target words in this study, participants did not focus their attention on the relationship among different target items to help them generate (regardless of constraint). Future studies using related lists of stimuli (where the target words are related to one another) and a similar manipulation of generation constraint would provide valuable information about the influence of generation constraint on target–target relational processing. We can turn to prior work on the generation effect, however, to offer some speculation on why we did not find a relationship between generation constraint and memory in the free recall group. Early work on the generation effect found that generation effects were often reduced or eliminated in measures of free recall, which was attributed to the idea that generation tasks enhanced item-specific processing at the expense of processing relations between target items (Hirshman & Bjork, 1988; Slamecka & Katsaiti, 1987). Later work however, showed that generation effects could occur in free recall if the generation task induced participants to focus on the relationship among the target items to help them generate. Critically though, increased generation effects on free recall measures under these conditions led to a concomitant decrease in performance on cued recall performance (deWinstanley & Bjork, 1997; deWinstanley, Bjork, & Bjork, 1996). This work convincingly demonstrates that generating acts to enhance memory for information that is useful in helping to generate the target item at encoding, but this heightened processing for certain information sometimes comes

at the expense of other, less relevant information (deWinstanley & Bjork, 1997; deWinstanley et al., 1996; McDaniel et al., 1990; McDaniel et al., 1988). Thus, although our data showed no evidence that generation constraint significantly influences target–target relational processing, future work using different procedures or stimuli may find that constraints can influence whether participants focus more (or less) on the relationship among target items. This idea is in line with the more general view that generation effects, as with many memory phenomena, are “context-sensitive” and depend heavily on the experimental context under which they are studied (Jenkins, 1979; McDaniel & Butler, 2011; Roediger, 2008).

Turning to our context (location) memory data, we found no evidence of a significant effect of generation constraint on memory for the spatial location of an item. This finding was somewhat surprising, given that previous studies have shown that some of the strongest effects of generation constraint are on context memory (McCurdy et al., 2017, 2019; McCurdy, Sklenar, et al., 2020). This prior work, however, has primarily focused on memory for source information (e.g., which task did you perform with this word pair?) and memory for font color of encoded words. Specifically, source memory was stronger under lower-constraint generation compared with higher-constraint generation across all three studies (McCurdy et al., 2017, 2019; McCurdy, Sklenar, et al., 2020), while font color showed no significant generation constraint effects in the lone study it has been investigated in (McCurdy, Sklenar, et al., 2020). Previously, it has been argued that certain contextual details may be more difficult to bind into thus it could be that generation constraint has a weaker influence on memory for spatial location. Indeed, all three of our models showed a negative relationship between generation constraint and location memory, but failed to reach significance. In general, research on the generation effect for context memory has often produced mixed results. Some studies have found that generation improves context memory over reading (Geghman & Multhaup, 2004; Marsh, 2006; Marsh et al., 2001), whereas others have found that generation provides no benefits or even hurts memory for some contextual details (Jurica & Shimamura, 1999; Mulligan, 2004, 2011; Mulligan, Lozito, & Rosner, 2006). The most widely accepted explanation for the mixed findings in the literature is known as the *processing account*, which proposes that the extent generation will improve memory for contextual details depends on the type of context detail and the processing required to encode that information (Mulligan, 2004, 2011). Interestingly, spatial location of the sort we tested in this study has been shown to yield enhanced memory from generating relative to reading (Marsh, 2006; Marsh et al., 2001). In line with this idea, the observed means in Tables 2 and 3 show that, relative to the read control task, generation conditions overall seemed to provide a memory boost for the location of the item. To confirm this, we ran three additional logistic regression models that included *Condition* and *NormMatchTarget* as predictors, which revealed a significant effect of *Condition* in both the recognition

($\beta = 0.46$, $SE = .19$, $p = .014$), and cued recall data ($\beta = 0.55$, $SE = .18$, $p = .003$), but not the free recall data ($\beta = 0.55$, $SE = .57$, $p = .330$). Thus, despite finding that generation constraint does not seem to influence memory for this detail, our data generally support prior work showing that generating improves memory for location over reading (Marsh, 2006; Marsh et al., 2001). It is worth noting that compared with the item recognition and cued recall data, the free recall data showed larger differences between conditional context memory and unconditional context memory performance (see Tables 2 and 3).⁶ This larger difference occurs because fewer items were remembered for free recall, and the conditional context memory measure adjusts for this poor item memory to allow for a purer comparison of context memory performance that is less confounded with item memory performance. Taken as a whole, the research on generation constraint and context memory suggests that constraint likely influences memory for some contextual details more than others, further highlighting the need for more research to identify the boundary conditions of generation constraint effects. Understanding how to improve memory is an important scientific goal (Frankenstein et al., 2020; Leach, McCurdy, Trumbo, Matzen, & Leshikar, 2018; Leshikar, Duarte, & Hertzog, 2012; Leshikar et al., 2017; Matzen, Trumbo, Leach, & Leshikar, 2015; Meyers, McCurdy, Leach, Thomas, & Leshikar, 2020), and this work adds to that endeavor.

To close the discussion, we recognize some limitations of the present study that prime future research. Keeping with context memory, in the present study we only investigated one type of contextual detail (spatial location). Prior work on the generation effect has examined generation effects for several different types of contextual details (e.g., source, font color, font type), and it is known that the generation effect depends on the type of contextual detail being tested (Marsh, 2006; Marsh et al., 2001; Mulligan, 2004, 2011; Mulligan et al., 2006). We chose to examine spatial location in the current study because previous work has examined the effects of generation constraint on source memory and font color (McCurdy et al., 2017, 2019; McCurdy, Sklenar, et al., 2020). These studies, however, utilized a dichotomous manipulation of generation constraint, and are subject to potential item-selection confounds.⁷ Thus, an opportunity exists for future studies to examine other types of contextual details using the experimental design and analysis technique put forth in this study to provide a more nuanced understanding of the relationship between generation constraint and different types of context memory. A second avenue for future research is to continue studying the relationship between generation constraint and free recall memory performance. A growing body of work has increased understanding of generation constraint effects for

⁶ A similar effect occurs for the read condition in item recognition, where item memory performance is lower, thus conditional context memory is inflated.

⁷ Although steps were taken to account for item-selection effects in these studies, the method did not model and control for them as we do in this study.

recognition and cued recall (McCurdy et al., 2017, 2019; McCurdy, Sklenar, et al., 2020). Yet less work has directly examined generation constraint effects for free recall (Fiedler et al., 1992; Gardiner et al., 1985). The current findings are limited in the amount of evidence provided about the relationship between generation constraint and free recall, and more research is necessary to more fully understand this relationship. Specifically, previous research has shown that the type of materials used (particularly whether the materials are related in some way or not) strongly influences whether a generation effect occurs in free recall tests (deWinstanley & Bjork, 1997; deWinstanley et al., 1996; McDaniel et al., 1990; McDaniel et al., 1988). Thus, it may be beneficial to use the design put forth in the present study to continue to investigate the relationship between generation constraint and free recall memory tests across different sets of materials or stimuli. Another important area of future research might be using different methods to examine generation constraint effects on encoding processes. In the current study, we use different memory tests to examine different types of encoding processes (item specific, relational), but another way to study the effects of generation constraint on encoding processes is through cumulative recall curves for free recall (Burns, 2006). Future work using this method could provide converging evidence for the findings put forth in the current study. Finally, the present study used intentional encoding procedures. It is well known that intentional versus incidental encoding procedures can influence how participants engage in certain encoding tasks (Begg, Vinski, Frankovich, & Holgate, 1991; McCurdy, Viechtbauer, et al., 2020). Thus, it is possible that knowing which type of memory test they would be given may have influenced participants encoding processes in this study. Future work using an incidental encoding procedure would provide more evidence for the findings in the current study.

In conclusion, this study shows that the generation effect is still a powerful and robust memory effect and we have provided more evidence that generation constraint influences the magnitude of generation effects, even after controlling for item-selection effects. Generation constraint is important to consider because, as shown here and in prior work, fewer constraints may be an important way to maximize the memory benefits from self-generation. This study should be informative to future work on the generation effect because it delineates what might be optimal experimental designs to maximize the benefits from the generation effect. Furthermore, this study provides a framework for future investigations examining how generation constraint influences memory, while controlling for the potential confound of item-selection effects that have previously discouraged the study of this influential factor.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13421-020-01119-0>.

Author note The data and analysis code supporting the findings in this manuscript are openly available at: <https://osf.io/3vqju/>

Open practices statement The data and analysis code for this article are available (<https://osf.io/3vqju/>). This article was not preregistered.

References

- Baayen, H., Bates, D., Kliegl, R., & Vasishth, S. (2015). RePsychLing: Data sets from psychology and linguistics experiments (R Package Version 0.0.4) [Computer software]. Retrieved from <https://rdrr.io/github/dmbates/RePsychLing/>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Begg, I. (1978). Similarity and contrast in memory for relations. *Memory & Cognition*, 6(5), 509–517.
- 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, and Cognition*, 15(5), 977–989.
- Begg, I., Vinski, E., Frankovich, L., & Holgate, B. (1991). Generating makes words memorable, but so does effective reading. *Memory & Cognition*, 19(5), 487–497.
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.
- Burns, D. J. (2006). Assessing distinctiveness: Measures of item-specific and relational processing. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 109–130). Oxford, UK: Oxford University Press.
- Chalfonte, B., & Johnson, M. (1996). Feature memory and binding in young and older adults. *Memory & Cognition*, 24(4), 403–416.
- deWinstanley, P. A., & Bjork, E. L. (1997). Processing instructions and the generation effect: A test of the multifactor transfer-appropriate processing theory. *Memory*, 5(3), 401–422.
- deWinstanley, P. A., Bjork, E. L., & Bjork, R. A. (1996). Generation effects and the lack thereof: The role of transfer-appropriate processing. *Memory*, 4(1), 31–48.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4–58.
- Einstein, G. O., & Hunt, R. R. (1980). Levels of processing and organization: Additive effects of individual-item and relational processing. *Journal of Experimental Psychology: Human Learning and Memory*, 6(5), 588–598.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160.
- 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.
- Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Thousand Oaks, CA: SAGE Publications.
- Frankenstein, A. N., McCurdy, M. P., Sklenar, A. M., Pandya, R., Szpunar, K. K., & Leshkar, E. D. (2020). Future thinking about social targets: The influence of prediction outcome on memory. *Cognition*, 204, 104390.
- Gardiner, J. M., Smith, H. E., Richardson, C. J., Burrows, M. V., & Williams, S. D. (1985). The generation effect: Continuity between

- generating and reading. *The American Journal of Psychology*, 98(3), 373–378.
- Geghan, K. D., & Multhaup, K. S. (2004). How generation affects source memory. *Memory & Cognition*, 32(5), 819–823.
- Greenwald, A. G., & Johnson, M. M. (1989). The generation effect extended: Memory enhancement for generation cues. *Memory & Cognition*, 17(6), 673–681.
- Hashtroudi, S., Johnson, M. K., & Chrosniak, L. D. (1989). Aging and source monitoring. *Psychology and Aging*, 4(1), 106.
- 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), 484–494.
- Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior*, 20(5), 497–514.
- 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.
- Jenkins, J. J. (1979). Four points to remember: A tetrahedral model of memory experiments. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 429–446). Hillsdale, NJ: Erlbaum.
- 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*, 94(1), 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.
- Leach, R. C., McCurdy, M. P., Trumbo, M. C., Matzen, L. E., & Leshikar, E. D. (2018). Differential age effects of transcranial direct current stimulation on associative memory. *The Journals of Gerontology: Series B*, 74(7), 1163–1173.
- Lenth, R. (2019). emmeans: Estimated marginal means, aka least-squares means [Computer software]. Retrieved from <https://cran.r-project.org/package=emmeans>
- Leshikar, E. D., Cassidy, B. S., & Gutches, A. H. (2016). Similarity to the self influences cortical recruitment during impression formation. *Cognitive, Affective, & Behavioral Neuroscience*, 16(2), 302–314.
- 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., Duarte, A., & Hertzog, C. (2012). Task-selective memory effects for successfully implemented encoding strategies. *PLOS ONE*, 7(5), e38160. <https://doi.org/10.1371/journal.pone.0038160>
- 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., & Gutches, 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., Leach, R. C., McCurdy, M. P., Trumbo, M. C., Sklenar, A. M., Frankenstein, A. N., & Matzen, L. E. (2017). Transcranial direct current stimulation of dorsolateral prefrontal cortex during encoding improves recall but not recognition memory. *Neuropsychologia*, 106, 390–397.
- Leshikar, E. D., Park, J. M., & Gutches, A. H. (2015). Similarity to the self affects memory for impressions of others in younger and older adults. *The Journals of Gerontology: Series B*, 70(5), 737–742.
- Marsh, E. J. (2006). When does generation enhance memory for location? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(5), 1216–1220.
- Marsh, E. J., Edelman, G., & Bower, G. H. (2001). Demonstrations of a generation effect in context memory. *Memory & Cognition*, 29(6), 798–805.
- Matzen, L. E., Trumbo, M. C., Leach, R. C., & Leshikar, E. D. (2015). Effects of non-invasive brain stimulation on associative memory. *Brain Research*, 1624, 286–296.
- McCurdy, M. P., Leach, R. C., & Leshikar, E. D. (2017). The generation effect revisited: Fewer generation constraints enhances item and context memory. *Journal of Memory and Language*, 92, 202–216.
- McCurdy, M. P., Leach, R. C., & Leshikar, E. D. (2019). Fewer constraints enhance the generation effect for source memory in younger, but not older adults. *Open Psychology*, 1(1), 168–184.
- McCurdy, M. P., Sklenar, A. M., Frankenstein, A. N., & Leshikar, E. D. (2020). Fewer generation constraints increase the generation effect for item and source memory through enhanced relational processing. *Memory*, 28(5), 598–616.
- McCurdy, M. P., Viechtbauer, W., Sklenar, A. M., Frankenstein, A. N., & Leshikar, E. D. (2020). Theories of the generation effect and the impact of generation constraint: A meta-analytic review. *Psychonomic Bulletin & Review*, 1–27.
- McDaniel, M. A., & Butler, A. C. (2011). A contextual framework for understanding when difficulties are desirable. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork* (pp. 175–198). New York, NY: Psychology Press.
- McDaniel, M. A., Riegler, G. L., & Waddill, P. J. (1990). Generation effects in free recall: Further support for a three-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(5), 789.
- McDaniel, M. A., Waddill, P. J., & Einstein, G. O. (1988). A contextual account of the generation effect: A three-factor theory. *Journal of Memory and Language*, 27(5), 521–536.
- Metcalfe, J., & Kornell, N. (2007). Principles of cognitive science in education: The effects of generation, errors, and feedback. *Psychonomic Bulletin & Review*, 14(2), 225–229.
- Meyers, Z. R., McCurdy, M. P., Leach, R. C., Thomas, A. K., & Leshikar, E. D. (2020). Effects of survival processing on item and context memory: Enhanced memory for survival-relevant details. *Frontiers in Psychology*, 11, 2244.
- Mulligan, N. W. (2004). Generation and memory for contextual detail. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(4), 838–855.
- 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), 836–846.
- 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. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, 36(3), 402–407.
- R Core Team. (2018). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing.
- Rabinowitz, J. C. (1989). Judgments of origin and generation effects: Comparisons between young and elderly adults. *Psychology and Aging*, 4(3), 259–268.
- Richland, L. E., Bjork, R. A., Finley, J. R., & Linn, M. C. (2005, July). *Linking cognitive science to education: Generation and interleaving effects*. Paper presented at the Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society, Stresa, Italy.
- Roediger, H. L. (2008). Relativity of remembering: Why the laws of memory vanished. *Annual Review of Psychology*, 59, 225–254.

- Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological science*, 17(3), 249–255.
- Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140(6), 1432.
- Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, 4(6), 592–604.
- 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.
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*: New York, NY: Springer-Verlag.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.