THE INFLUENCE OF CUE GENERATION AND WORKING MEMORY ON RECALL FROM LONG-TERM MEMORY

by

Amanda L. Fordyce

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Psychological Sciences

West Lafayette, Indiana

August 2026

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Prof. Thomas Redick, Chair
Department of Psychological Sciences

Prof. Darryl Schneider, Member
Department of Psychological Sciences

Prof. Jeffrey Karpicke, Member
Department of Psychological Sciences

Prof. Hongmi Lee, Member
Department of Psychological Sciences

Approved by:
Dr. Margo Monteith

ACKNOWLEDGMENTS

I would like to thank my committee members, Tom Redick, Darryl Schneider, Jeff Karpicke, and Hongmi Lee for their feedback on this project. Additionally, I would like to thank my lab members, Alexa Giebel, Maddie Hoffman, and Harun Yoruk for their feedback. Finally, I would like to thank the undergraduate research assistants that helped with data collection: Carly Booth, Caleb Roberts, Atmika Ajay, Andjela Zanza, Callie Clodfelter, and Emily Baker.

TABLE OF CONTENTS

TABLE OF CONTENTS..**Error! Bookmark not defined.**
LIST OF TABLES 7
LIST OF FIGURES 8
LIST OF ABBREVIATIONS 9
ABSTRACT 10
INTRODUCTION 12
Long-Term Memory 12
Encoding 13
Retrieval 15
Working Memory 18
Encoding 19
Retrieval 23
Encoding and Retrieval 28
The Current Study 30
EXPERIMENT 1 34
Method 34
Participants 34
Design and Materials 35
Procedure 35
Generation Task 35
Reading Span 37
Change Detection 38
Analyses 39

Results 42
Proportion Correct 42
Recall Latency 44
Recall Errors 45
Cue Properties 48
Discussion 55
EXPERIMENT 2 59
Method 59
Participants 59
Design and Materials 60
Procedure 61
Generation Task 61
Reading Span 63
Change Detection 63
Analyses 64
Results 65
Proportion Correct 65
Recall Latency 68
Recall Errors 71
Cue Properties 73
Discussion 79
GENERAL DISCUSSION 83
Limitations and Future Directions 86
Conclusion 89

APPENDIX A. EXPERIMENT 1 SUPPLEMENTAL 91
APPENDIX B. EXPERIMENT 2 SUPPLEMENTAL... 115
REFERENCES 137
PUBLICATION 147

LIST OF TABLES

Table 1. Summary of Previous WM Retrieval Dynamics Research 26

Table 2. Experiment 1 Descriptive Statistics. 51

Table 3. Experiment 1 Cue Properties Descriptive Statistics.. 52

Table 4. Experiment 1 Correlation Matrix. 53

Table 5. Experiment 2 Descriptive Statistics. 75

Table 6. Experiment 2 Cue Properties Descriptive Statistics.. 76

Table 7. Experiment 2 Generate Cued Recall Condition Correlation Matrix. 77

Table 8. Experiment 2 Allocate Cued Recall Condition Correlation Matrix. 78

Table 9. Experiment 2 Generate Free Recall Condition Correlation Matrix. 78

Table 10. Experiment 2 Allocate Free Recall Condition Correlation Matrix. 79

LIST OF FIGURES

Figure 1. Experiment 1 Generation Task Procedure. 37

Figure 2. Reading Span Procedure. 38

Figure 3. Change Detection Procedure. 39

Figure 4. Experiment 1 Proportion Correct by Condition. 43

Figure 5. Experiment 1 Proportion Correct by Condition and WM. 44

Figure 6. Experiment 2 Generation Task Procedure. 63

Figure 7. Experiment 2 Proportion Correct by Condition. 67

Figure 8. Experiment 2 Proportion Correct by Condition and WM. 68

Figure 9. Experiment 2 Mean Recall Latency by Condition. 70

Figure 10. Experiment 2 Mean Recall Latency by Condition and WM. 71

LIST OF ABBREVIATIONS

ANCOVA: Analysis of covariance

ANOVA: Analysis of variance

ERP: Event-related potential

IRT: Inter-response time

Lag-CRP: Lag-conditional response probability

RT: Reaction time

SAM: Search of Associative Memory

WM: Working memory

ABSTRACT

Working memory (WM) refers to the ability to maintain and manipulate information in mind while simultaneously inhibiting distracting information. WM ability has been shown to correlate with measures of attentional control as well as measures of long-term memory. Given this, multiple studies have investigated WM-related differences in attention at encoding and in the retrieval cues used during recall. This research has provided evidence that high WM individuals tend to be better at sustaining attention, encoding only relevant or highly important information, and using effective encoding strategies such as sentence generation or grouping. During recall, high WM individuals tend to use more effective cues (e.g., semantic category cues or temporal-contextual cues) which allow them to recall more items, make fewer intrusions, and recall items faster than low WM individuals. However, encoding and retrieval processes are not independent. Where attention is allocated during encoding and what cues are generated during encoding will determine what cues are possible to use during retrieval. The current study addresses this issue and examines the extent that encoding and retrieval strategies account for WM-related differences in long-term memory. The current study examines WM-related differences in performance following the use of techniques that may facilitate more effective encoding. In Experiment 1, participants either self-generated descriptive cues, self-generated any cue, or were allocated cues during study. They were later tested on their memory with a cued recall test. In Experiment 2, participants either self-generated cues or were allocated cues during study. They were later tested on their memory with either a cued recall test or a free recall test. In both experiments, WM was measured using a reading span task and a change detection task. Proportion correct, recall latencies, recall errors, and properties of generated cues were examined in both experiments to provide insight on the effect of WM and self-generation on encoding and

retrieval processes. Overall, self-generating cues during encoding improved proportion correct in cued and free recall, reduced recall latencies in free recall, and had mixed results on recall errors. Additionally, more distinctive and unique cues were generated in the generated descriptive cue condition relative to the generated cue condition. Individuals with high WM recalled a higher proportion of words correctly and made fewer recall errors in both cued and free recall. There was partial evidence that individuals with high WM generated more distinctive and unique cues in Experiment 1, but this effect was not replicated in Experiment 2. Across all dependent variables, only the interaction in omission errors between WM and condition in Experiment 1 was significant. WM-related differences in omission errors were reduced in the generated descriptive cue condition relative to the generated cue and allocated cue conditions. Together, this suggests that while both self-generating cues and WM largely had a positive effect on long-term memory performance, neither self-generation of cues at encoding (Experiments 1 and 2) nor providing cues during retrieval (Experiment 2) helped reduce WM-related differences in long-term memory performance.

INTRODUCTION

Decades of research on long-term memory have provided evidence to the cue-dependent nature of memory. This research has been used to explain and identify situations where memory retrieval is likely to succeed and likely to fail. Additionally, understanding the role of cues in memory provides insight into what strategies will be most effective during encoding and how we search through memory representations during retrieval. However, there are still numerous questions about how and why different individuals are better at retrieving information from memory. Processes at encoding such as allocation of attention and cue generation, as well as processes at retrieval such as search set formation, recovery ability, and monitoring ability can all contribute to performance differences. The current study aimed to examine these factors as well as to apply principles of effective encoding and retrieval from the long-term memory literature to the WM literature.

Long-Term Memory

Memory refers to the ability to use the past in service of the present. Long-term memory specifically refers to the system used to store and retrieve information over extended periods of time. Previous research on long-term memory has demonstrated that probability of recall is typically improved when the conditions at retrieval are similar to the conditions during encoding (Thomson & Tulving, 1970). Tulving and Pearlstone (1966) was one of the first studies to demonstrate this idea. Participants studied multiple lists of categorical words. At retrieval, participants were either provided category names to aid retrieval, or were not given any cues. Performance was better in the cued recall condition relative to the non-cued condition. Thomson and Tulving (1970) expanded on this finding and coined the term *encoding specificity*.

Participants studied multiple words either presented alone, with a strong semantic associate, or with a weak semantic associate. At retrieval, participants either engaged in non-cued recall, or were given the strong or weak semantic associate as a cue to help recall the target word. Performance was best when the cue at encoding matched the cue at retrieval (i.e., strong/strong or weak/weak). Nairne (2002) further expanded on this concept, proposing that it was not the absolute match that was important, but rather the diagnosticity of the cue. In other words, what drives successful retrieval in such situations is the extent that the cue matches the target to the exclusion of other items in memory. Together, this work provides insight into the characteristics which make a retrieval cue effective. Unique retrieval cues that are present at both encoding and retrieval will best aid recall.

Encoding

Previous research has identified a variety of encoding tasks that improve retention of information by building on principles of effective cues. One such method is self-generation of cues (for review, see Wheeler & Gabbert, 2017). For example, Mäntylä (1986) used this method and demonstrated high levels of recall at immediate and delayed retention intervals (i.e., 1 day, 2 days, and 7 days). Participants began by viewing 500-600 words. They were not informed that they would later be tested on their memory for the words. Participants were asked to generate 1 or 3 words that could be used to describe the target word. At retrieval, they were then given the cues that they generated (Experiments 1-3), or the cues of another participant (Experiment 2). Across all experiments, high recall performance was observed when cues were generated during study. For example, in Experiment 2, near perfect recall was observed with 3 self-generated cues at the immediate test ($M = .91$) and remained high 1 week later ($M = .65$). Recall with allocated cues was much lower for both the immediate test ($M = .55$) and the 1 week delayed test ($M =$

.43). The high performance in the self-generated condition can be attributed to the characteristics of effective retrieval cues. Self-generation of multiple cues at encoding ensured that the cues were present at encoding and, when participants were given their own cues at retrieval, that the cues at encoding were also present at retrieval. Additionally, the generation of multiple cues created unique and item-specific combinations that were distinct relative to other items.

Subsequent research has examined the properties of generated cues to determine which cues were most effective for guiding later retrieval. For example, Mäntylä and Nilsson (1988) asked participants to either generate three cue words that they felt appropriately described the item (spontaneous condition) or to generate three cue words that they felt were distinctive and that they would most likely generate at a later time in response to the target word (focused condition). Higher intersubject overlap (i.e., output of the same cue by different participants) was observed for the spontaneous condition relative to the focused condition, suggesting that cues generated in the focused condition were more unique. Additionally, when presented with the generated cues again after 3 weeks, recall was better in the focused condition ($M = .94$) than in the spontaneous condition ($M = .55$).

Tullis and Benjamin (2015) also reported a benefit in performance when more distinctive cues were generated at encoding. Germane to the current research, in Experiments 1 and 3 of Tullis and Benjamin (2015), participants were instructed to either generate a cue word that described the target word (*description generation* condition) or to generate any cue that would help them remember the target word (*cue generation* condition). Distinctiveness and uniqueness of generated cues was compared between conditions. Overall, cues from the cue generation condition were associated with fewer targets, had lower cumulative cue-to-target associative strength, and were less likely to be in the University of South Florida Free Association Norms database (Nelson et al., 1998) than cues from the description generation condition. Additionally, across participants, more unique cues were generated for target words in the cue generation condition relative to the description generation condition. Finally, cued recall performance following a 10 minute delay was better for the cue generation condition relative to the description generation condition. Together, the self-generation literature provides evidence for improved memory performance when cues match at encoding and retrieval, as well as when cues are distinctive for a given target word.

Retrieval

In addition to encoding processes, retrieval processes also contribute to successful recall from memory. One model of long-term memory detailing these retrieval processes is SAM (Search of Associative Memory), proposed by Raaijmakers & Shiffrin (1980). The SAM model proposes that information is represented in memory as a memory trace or "memory image". This image contains item information (features of the item such as semantic or phonetic properties), contextual information (features of the environmental context the item occurred in), and associative information (relationships between the item and other memory images). The SAM model also assumes that retrieval is cue-dependent. The degree to which a retrieval cue is associated with a specific target memory image, as well as the degree the retrieval cue is associated with other images, determines the probability of recall. Memory will be most successful when the cue is highly associated with the target image and poorly associated with other images. This association between the retrieval cue and the image is referred to as the associative strength. The SAM model specifies three main retrieval processes involved in memory search. The first is selection of the search set. Rather than searching the entirety of memory, only a few items are selected to sample from, based on their relevance to the retrieval cue. The next process is sampling. Sampling is determined by the associative strength of the item relative to the associative strength of the other items in the search set. Finally, after an item is sampled from the search set, it may be recovered. Recovery is determined by the absolute associative strength of the item. After these search processes occur and an item is recovered, individuals then engage in an evaluation process to determine if the recovered item was the target item. Deficits in any of these processes could result in failure to recall an item, or lead to recall of an incorrect item.

One challenge of memory research is approximating memory search processes from empirical data. To address this issue, Bousfield & Sedgwick (1944) used an approach examining the time course of recall from semantic memory. This approach was later expanded upon by Wixted & Rohrer (1994) and applied to episodic memory. Wixted & Rohrer reported that cumulative recall could be plotted as a negatively accelerated exponential function. This function could be characterized by asymptotic recall and the rate of approach to asymptote. Asymptotic recall can be approximated by the total number of items recalled. The rate of approach to asymptote can be measured by examining recall latencies (the time when a given item is recalled). Cumulative recall curves can then be expressed by the following exponential:

$$F(t) = N(1 - e^{-\lambda(t-c)})$$

where $F(t)$ represents the cumulative number of items recalled by time t, N represents asymptotic recall, λ represents the rate of approach to asymptote, and c represents the initiation pause preceding recall. A larger λ value indicates faster recall, while a smaller λ indicates slower recall. Wixted & Rohrer (1994) noted that for episodic memory, the relationship between asymptotic recall (N) and rate of approach to asymptote (λ) will vary depending on the size of the search set. This was supported by Wixted & Rohrer (1993), which varied the amount of proactive interference by exposing participants to multiple lists of items from the same semantic category.

As proactive interference increased, rate of approach to asymptote (λ) decreased. Buildup of proactive interference can be conceptualized as including more items in the search set. Thus, as more items are included in the search set, it will take longer to sample a given item. This supports the idea that recall latency can be used as an index of search set size. Another approach to calculating recall latencies involves measuring inter-response times (IRTs) rather than λ, as parameter estimates derived from fitting data to a cumulative recall curve can lead to low statistical power from increased variability, making it more difficult to detect differences between individuals and conditions (Unsworth, 2007).

Total number of items recalled can also provide information about retrieval processes, namely the number of recoverable items in the search set. A study by Rohrer & Wixted (1994) showed that as presentation duration increased, the total number of items recalled also increased. Increasing presentation duration can improve encoding and subsequently, the absolute strength of the item. This would increase the probability of recall.

Finally, recall errors can provide insight into monitoring processes of retrieval (Unsworth, 2007). Recall errors in multi-list free recall can be categorized as prior-list or extra-list intrusions. Prior-list intrusion errors would occur when an item from a previous studied list was recalled. Extra-list intrusion errors would occur when an item not presented in the experiment was recalled. These errors reflect a failure of monitoring. The item recalled was included in the search set and recovered, but then incorrectly identified as a to-be-recalled item.

Retrieval dynamics can also be analyzed in cued recall, though there are a few differences. In a typical cued recall paradigm, participants will be provided with a cue corresponding to a single target which they are expected to output. However, in free recall, multiple target words are expected to be output during a recall trial. Given this, cued recall requires the search set to be narrower than in free recall. However, the processes involved in retrieval largely remain the same. Recall latency can be calculated by measuring the time between presentation of a cue and output of a response. This would again be an index of search set size. Additionally, recall errors such as intra-list intrusions (an incorrect response with a studied target item) and extra-list intrusions can be measured in cued recall as an index of monitoring ability. Finally, proportion correct and omission errors (no response) can be calculated to provide information about the success or failure of recall. These could occur due to a combination of retrieval processes, however, and are not often used as a pure index of recovery ability like the total number of words recalled in free recall. Additionally, cumulative recall curves are not fit to cued recall data as they are with free recall data.

Working Memory

WM refers to the ability to maintain and manipulate information in mind while simultaneously inhibiting distracting information. WM ability has been shown to predict multiple measures of higher-order cognition including fluid intelligence, attentional control, reading comprehension, academic performance, and episodic memory (for review, see Kane et al., 2007; Unsworth, 2019; Unsworth & Redick, 2017). Although there is debate about the structure and processes involved in WM, many researchers propose some interaction between attention and memory as a key characteristic (Baddeley, 2012; Cowan, 1999; Cowan, 2017; Engle et al., 1999; Oberauer, 2009). Certain models emphasize specific components over others, however. For example, the executive attention account of WM proposes that the ability to control attention and inhibit irrelevant information is the primary driver of WM-related differences (Engle et al., 1999). The embedded processes model proposes that WM is an activated portion of long-term memory. This model emphasizes the role of primary memory and the amount of information that can be maintained in the spotlight of attention (Cowan, 1999). Multicomponent (Baddeley, 2012) and multifaceted accounts (Robison et al., 2024; Unsworth et al., 2014) propose that WM involves multiple components including attentional control, short-term memory storage/capacity, and long-term memory. The combination of these components can be used to explain the relationship between WM and higher-order cognition. This combination of attention and memory components is also reflected in WM-related differences in encoding and retrieval during episodic memory tasks.

Encoding

Previous research on WM-related differences in encoding processes for long-term memory tasks has largely focused on attention at encoding. For example, previous work has suggested that high WM individuals are better at sustaining attention relative to low WM individuals (e.g., McVay & Kane, 2009; Miller et al., 2019). McVay and Kane (2009) demonstrated this by examining mind-wandering rates, reaction time (RT) variation, and performance errors on the Sustained Attention to Response Task. This task is a variation of a go/no-go task which requires participants to make responses to all stimuli except certain target stimuli that are presented infrequently. McVay and Kane observed that high WM individuals reported fewer task-unrelated thoughts, showed less RT variation, and showed larger signal detection sensitivity on the Sustained Attention to Response Task, indicating that they were better at sustaining their attention to task demands. Multiple correlational, experimental, and factor analytic studies have similarly shown significant correlations between WM and performance on attentional control tasks (for review, see Kane et al., 2007; Unsworth et al., 2024). In addition to attentional control tasks, Miller et al. (2019) demonstrated WM-related differences in sustained attention on an immediate free recall task. Miller et al. examined pupil dilation during encoding of multiple words. Previous research has demonstrated that pupil dilation changes with attentional demands of a cognitive task. For example, the pupil dilates with increased memory load in a short-term memory task (e.g., Kahneman & Beatty, 1966; Peavler, 1974), and narrows when memory load exceeds capacity limits (e.g., Granholm et al., 1996; Granholm et al., 1997; Van Gerven et al., 2004). Miller et al. (2019) found that pupil dilation for high WM individuals increased with serial position, whereas pupil dilation for low WM individuals decreased with serial position. In their Experiment 2, a value-directed remembering paradigm was applied, which has been used in memory research to examine how individuals strategically allocate encoding (Watkins & Bloom, 1999). Words were assigned point values which were awarded to participants upon successful recall. Participants were told to study the words with the goal of maximizing their total points. When values were presented in ascending order, such that words presented later in the list were labeled as more important than items presented earlier in the list, low WM individuals no longer showed a decline in pupil dilation with serial position. Instead, pupillary responses increased with serial position, though they were still smaller relative to high WM individuals. Together these results suggest that high WM individuals are typically better at sustaining attention across a list, though certain encoding strategies may be able to influence how attention is allocated during encoding. Individual differences in attention at encoding may still persist with these strategies, however.

Other research has also suggested that there may be differences in how attention is allocated during encoding. For example, Vogel et al. (2005) provided evidence that there were not differences in the absolute capacity of WM. Instead, low WM individuals were more likely to encode irrelevant information than high WM individuals. Vogel et al. measured event-related potentials (ERPs) when encoding a visual array. The visual array consisted of either 2 or 4 red rectangles at various orientations (distractor absent condition). On a third of the trials, 2 red rectangles and 2 blue rectangles at various orientations were presented (distractor present condition). Following a brief display of the items, participants were then shown another array with the red rectangles in the same orientations as the first display, or with one rectangle in a different orientation from the first display. Participants were then asked if the second display of red rectangles was the same or different from the first display. Vogel et al. compared the amplitude of the contralateral delay activity, which increases with the number of items maintained in memory, for distractor present and distractor absent conditions. If the amplitude was similar for the distractor absent condition with 2 red rectangles and the distractor present condition, then participants would be encoding the items with maximum efficiency, and irrelevant blue rectangles were not maintained in WM. However, if the amplitude was similar for the distractor absent condition with 4 red rectangles and the distractor present condition, participants would not be encoding the items with maximum efficiency. Overall, Vogel et al. observed that high WM participants showed greater encoding efficiency than low WM participants.

Robison & Unsworth (2017) observed complementary findings in a value-directed remembering paradigm. In this paradigm, participants engaged in self-directed study of words which varied in assigned point value. They were instructed to study with the goal of maximizing their total points at retrieval. Following the study period, participants completed a free recall test. In Experiment 1, when participants were allowed to use any strategy, high WM individuals were more likely to ignore low value items and only study high value items (i.e., the ignore low strategy) for one or more lists. Overall, this was the most effective strategy for optimizing total score. High WM individuals had higher point totals, recalled more words overall, recalled a higher proportion of the words they studied, and were more selective in recalling high-value words than low WM individuals. In Experiment 2, participants were instructed to use the 'ignore low' strategy. Following use of this strategy, WM-related differences in point totals, number of words recalled, and memory selectivity were eliminated. These results again suggest that despite initial differences in where attention is being allocated during encoding, WM-related differences in performance can be reduced or eliminated with effective encoding strategies.

However, the literature on WM-related differences in encoding strategies and the training of effective encoding strategies is somewhat limited. The majority of this research has focused on strategies for effective performance in WM tasks (e.g., complex span tasks). For example, prior work has suggested that high WM individuals are more likely to spontaneously use strategies such as sentence generation, mental imagery, or grouping on WM span tasks (Dunlosky & Kane, 2007; Turley-Ames & Whitfield, 2003). Research examining the relationship between WM and reading comprehension has reported that low WM individuals are more likely to use rehearsal (Kaakinen & Hyona, 2007) relative to high WM individuals, who are more likely to use semantic elaboration strategies. Whitney et al. (1991) observed that low WM individuals made more specific elaborations throughout the text, while high WM individuals made more specific elaborations at the end of the text. Elaboration at the end of the text can help avoid misinterpretations and aid in integrating content. Other research has examined the influence of training specific strategies, though these results have been somewhat mixed. Some work has observed a benefit of sentence generation training (Bartsch & Oberauer, 2021; Turley-Ames & Whitfield, 2003) and mental imagery training (Bartsch & Oberauer, 2021) on WM span performance, while other work has not (Bartsch et al., 2018, 2019; McNamara & Scott, 2001).

Similarly, some work has observed a benefit of rehearsal on WM performance (McNamara & Scott, 2001), while other work has not (Souza & Oberauer, 2018).

Less research has examined WM-related differences in encoding strategies for long-term memory tasks. Bailey et al. (2008) observed that effective strategy use (e.g., sentence generation, interactive imagery) accounted for significant variance in the WM/long-term memory relationship, but did not fully mediate the relationship. Effective strategy use did not account for significant variance in the WM/reading comprehension relationship. Unsworth (2016) similarly observed that effective strategy use partially mediated the relationship between WM and long-term memory. In contrast, a model taking into account strategy use, study time, search efficiency (IRTs), and monitoring ability (intrusions) fully mediated the relationship between WM and long-term memory. Together, Bailey et al. (2008) and Unsworth (2016) suggest that, while encoding strategy may be an important factor for explaining the WM/long-term memory relationship, there may still be WM-related differences in retrieval processes that account for variation in performance.

Retrieval

A number of studies have examined WM-related differences in retrieval processes. Much of this work has suggested that high WM individuals use more effective retrieval cues to guide recall. For example, Unsworth and Engle (2006) examined recall errors across serial positions. They found that low WM individuals were more likely to make omission errors at the beginning of the list and transposition errors at the end of the list relative to high WM individuals. This suggests that low WM individuals had more difficulty reinstating relatively distant context (i.e., the beginning of the list) in order to recall the target items. Additionally, when items were recalled, they were more likely to be placed in the incorrect serial position, suggesting that low

WM individuals used noisier temporal contextual cues during recall than high WM individuals. Spillers and Unsworth (2011) also found WM-related differences in the use of temporal-contextual cues. They observed that both high and low WM individuals began recall similarly (with an item at the beginning of the list); however, high WM individuals showed more temporal contiguity, as measured by lag-conditional response probabilities (lag-CRPs). Lag-CRPs reflect the probability that, when an item is recalled, the next item recalled will be one that was presented at a presentation lag of *x* (Howard & Kahana, 1999). This can provide insight into the cues being encoded and used during retrieval. For example, higher temporal contiguity in recall would suggest that temporal-contextual cues were encoded and used to drive recall. High WM individuals were more likely to recall items presented in succession, thus using the products of retrieval as retrieval cues for items presented in nearby serial positions.

WM-related differences in semantic retrieval cues have also been observed. Using a category fluency task which required participants to name animals which were then sorted into categories (e.g., pet, farm, aquatic) and clusters (items of the same category recalled in succession), Rosen and Engle (1997) and Unsworth et al. (2013) observed that high WM individuals recalled more animal names, recalled more clusters of animal names, had larger cluster sizes, and recalled at a faster rate than did low WM individuals. This again suggested that there may be WM-related differences in cue generation at retrieval. This idea was also supported by Unsworth (2009a), which reported higher correlations between WM and episodic memory performance for tasks that required more self-initiated processes (e.g., free recall) relative to tasks that required less self-initiated processes (e.g., cued recall, recognition). In free recall, participants self-generate retrieval cues, whereas in cued recall, a cue is provided to participants to aid recall.

Analyses of WM-related differences in retrieval dynamics have also been conducted for free recall (Unsworth, 2007, 2009c; Miller & Unsworth, 2018) and cued recall (Unsworth, 2009b; see Table 1 for summary). First, looking at the free recall studies, Unsworth (2007) examined proportion correct, previous list intrusions, extra list intrusions, and cumulative latency distributions on a free recall task. High WM individuals recalled more total items and more items correctly, made fewer previous list intrusions, and recalled items quicker (lower mean recall latency) than low WM individuals. Unsworth (2009c) similarly found that WM ability was positively correlated with proportion correct and negatively correlated with previous list intrusions and mean recall latency on free recall tasks. The strongest correlation was between WM and recall latency (r = -.46). This would suggest that WM-related differences in recall are primarily related to the ability to focus the search set; however, retrieval processes related to recovery and monitoring ability may also be important to consider. Miller & Unsworth (2018) observed a similar pattern of WM-related differences in retrieval dynamics measures, but also noted that proactive interference impaired recall accuracy and slowed recall latency for low WM individuals to a greater extent than high WM individuals. This finding is consistent with Kane & Engle (2000), who also reported greater impairment in performance with the buildup of proactive interference for low WM individuals. Miller & Unsworth (2018) also noted that WM-related differences were eliminated in IRTs and recall accuracy with repeated practice of the word list. Finally, Unsworth (2009b) investigated WM-related differences in proportion correct, recall latencies, omission errors, intra-list intrusions, and extra-list intrusions on a cued recall task. High WM individuals recalled more items correctly, recalled items faster, and made fewer intrusions than low WM individuals.

Table 1.

Summary of Previous WM Retrieval Dynamics Research

Citation	Long-Term Memory Task	Experimental Manipulation	Long-Term Memory Dependent Variables	WM Results
Miller & Unsworth (2018) Exp 1	FR (semantically related)	-	(a) Proportion Correct, (b) Mean Recall Latency, (c) IRT, (d) Intrusions, (e) Repetitions	(a) H > L; (b, c, d) H < L; (e) H = L
Miller & Unsworth (2018) Exp 2	Multi-trial FR	-	(a) Proportion Correct, (b) Mean Recall Latency, (c) IRT, (d) Intrusions, (e) Repetitions, (f) Organization	(a, d, f) H = L; (b, c) H > L, but H = L on later trials; (e) H < L
Spillers & Unsworth (2011)	FR	-	a) Proportion Correct, (b) PFR, (c) lag-CRP	(a) H > L; (b) H = L; (c) Nearby lags and forward transitions: H > L
Unsworth (2007) Exp 1	FR	List Length (6, 9, 12)	(a) Proportion Correct, (b) List Intrusions, (c) Extra-list Intrusions, (d) Repetitions, (e) N, (f) , (g) Mean Recall Latency, (h) IRT	(a, e, f) H > L; (b, g, h) H < L; (c, d) H = L
Unsworth (2007) Exp 2	Continuous Distractor FR	List Length (6, 9, 12)	(a) Proportion Correct, (b) List Intrusions, (c) Extra-list Intrusions, (d) Repetitions, (e) N, (f) , (g) Mean Recall Latency, (h) IRT	(a, e, f) H > L; (b, g, h) H < L; (c, d) H = L

Table 1 (*continued*)

Unsworth (2009b) Exp 1	Cued Recall	-	(a) Proportion Correct, (b) Omissions, (c) List Intrusions, (d) Extra-list Intrusions, (e) Correct Recall Latency, (f) Incorrect Recall Latency	(a) H > L; (b, f) H = L; (c, d, e) H < L
Unsworth (2009b) Exp 2	Cued Recall (phrase)	-	(a) Proportion Correct, (b) Omissions, (c) List Intrusions, (d) Extra-list Intrusions, (e) Correct Recall Latency, (f) Incorrect Recall Latency	(a) H > L; (b, e) H < L; (c, d, f) H = L
Unsworth (2009c)	FR (unrelated), FR (semantically related), List-before-last	-	(a) Proportion Correct, (b) Mean Recall Latency, (c) Intrusions	(a) H > L; (b, c) H < L
Unsworth (2016)	FR	Presentation Rate (1s, 4s, unlimited)	(a) Proportion Correct, (b) Strategy, (c) Study Time, (d) IRT, (e) Intrusions	(a) H > L; (b) Effective: H > L; (c) H = L; (d, e) H < L
Unsworth et al. (2012) Exp 1	Categorized Recall	Encoding Condition (Blocked vs. Random), Retrieval Condition (Cued vs. Free)	(a) Proportion Correct, (b) Number of categories FR, (c) Number of Words per Category FR, (d) IRT	(a) H > L, smaller difference in the cued condition than free condition; (b, c) H > L; (d) H < L
Unsworth et al. (2012) Exp 2	Categorized Cued Recall	Encoding Condition (Blocked vs. Random), Number of Words per Category (3 vs. 9)	(a) Proportion Correct, (b) IRT	(a) List Length 9: H > L, List Length 3: H = L; (b) H < L

NOTE: Overview of past research examining WM-related differences in retrieval dynamics on long-term memory tasks. All studies measured WM as a composite of Operation Span, Symmetry Span, and Reading Span scores. FR = Free Recall. IRT = Inter-response Time. H = High WM. L = Low WM. PFR = Probability of First Recall. Lag-CRP = Lag-Conditional Response Probability. N = Asymptotic Number of Items Recalled.

Encoding and Retrieval

As noted above, some previous research has suggested that WM-related differences in recall performance are due to both encoding processes (e.g., strategy) and retrieval processes (e.g., search set efficiency; Unsworth, 2016). Research attempting to examine the differential impact of these processes is limited. Unsworth & Spillers (2010) addressed this question by requiring participants to complete a continuous-distractor free recall task under incidental or intentional encoding conditions. High WM individuals recalled more items correctly in both encoding conditions. However, there was a significant interaction between WM and condition, such that the effect of WM was larger in the intentional encoding condition. Together this suggests that high WM individuals engage in more strategic encoding processes; however, differences in performance are not solely due to those differences in encoding.

Unsworth et al. (2012) examined WM-related differences in performance for cued and free recall of categorized lists. Encoding and recall conditions were manipulated such that words were either studied in a blocked or randomized order and recalled under free recall or cued recall conditions. In the blocked encoding condition, participants were presented with a category label followed by the exemplars of the category. In the randomized encoding condition, exemplars were presented intermixed with other category exemplars and no category labels were presented. In the free recall condition, participants were asked to recall as many items as they could remember. Finally, in the cued recall condition, participants were given a category label and asked to recall the exemplars of that category. Overall, high WM individuals recalled more items correctly than low WM individuals. This difference was reduced in the cued recall condition, and cued recall benefited low WM individuals more than high WM individuals. The three-way interaction between WM, encoding condition, and retrieval condition was not statistically significant. However, high WM individuals recalled significantly fewer items in the random-cued condition than the random-free condition, while there was no difference in performance between these conditions for low WM individuals. This suggests that high WM individuals may be more susceptible to mismatches between encoding and retrieval, consistent with the findings of Unsworth et al. (2011).

Unsworth et al. (2011) required participants to make a rhyme or semantic judgment during encoding. On a later cued recall test, participants were either presented with a matching or mismatching cue (i.e., match: rhyme/rhyme, mismatch: rhyme/semantic). When encoding and retrieval conditions matched, high WM individuals outperformed low WM individuals. However, there were no WM-related differences in performance when encoding and retrieval conditions mismatched. This was primarily due to an impairment in performance for high WM individuals. Related findings can be found in the part-list cuing literature. Part-list cuing paradigms require participants to study a set of stimuli. At recall, some of the studied stimuli are provided and participants are asked to recall the remaining studied stimuli. Typically, participants will recall a smaller proportion of words when given the part-list cues (relative to recall without the cues) due to the cues disrupting the way that participants would naturally recall information. However, Cokely et al. (2006) observed that performance for low WM individuals was not affected by part-list cuing, but high WM individuals did show the typical impairment. If high WM individuals typically engage in more effective encoding strategies, retrieval conditions which do not allow them to effectively use their self-generated cues at study may be more detrimental to performance than low WM individuals who did not engage in those encoding strategies initially.

The Current Study

The current study examined how cue generation at encoding impacted WM-related differences in cued and free recall performance. This was done with the goal of providing insight on which encoding and retrieval processes accounted for WM-related differences in long-term memory. Additionally, the current study examined if there were WM-related differences in the effectiveness of generated cues (i.e., distinctiveness and uniqueness). Given that prior research has provided evidence for WM-related differences in attention at encoding and use of effective cues at retrieval, a self-generation encoding strategy was selected with the goal of facilitating attentional processes at encoding for low WM individuals who may not engage in these processes typically. This strategy requires participants to attend to every to-be-remembered item. Given this, if WM-related differences in long-term memory performance are driven by intensity of attention at encoding (e.g., Miller et al., 2019), these differences should be reduced when participants are asked to generate a cue at encoding and studied items are actively attended to. Additionally, this self-generation task requires a participant to generate a cue with each item. Previous research has proposed that individuals with high WM use more effective cues at retrieval (e.g., Unsworth & Engle, 2007); however, the use of effective cues at retrieval may be due to high WM individuals initially generating more effective cues at encoding to be used later. If this explanation is driving WM-related differences in long-term memory performance, these differences should also be reduced, particularly when more effective cues are generated. The role of WM in retrieval processes was also examined by varying the level of cue support/self-initiated processing at retrieval through the use of free recall and cued recall tests.

This study aimed to address a number of limitations in the self-generation and WM/long-term memory literature as well. Previous studies that have compared memory performance between self-generated and allocated cue conditions typically manipulate condition at retrieval, not encoding. For example, all participants may self-generate cues at encoding, but at retrieval, they are either given their own cues or another participants' cues. This confounds the effect of self-generation versus allocation with the effect of an encoding/retrieval mismatch. Given this, the current study manipulated self-generation versus allocation at encoding to better isolate this effect. Additionally, to my knowledge, prior literature has not examined recall errors or recall latencies following self-generation. These measures can provide insight on how self-generation affects memory search processes such as search set size, recovery ability, and monitoring ability, which have been examined in the retrieval dynamics literature (e.g., Wixted & Rohrer, 1994).

In the WM literature, the effect of encoding strategies has primarily been examined in WM tasks, not in long-term memory tasks. These assessments have different task demands that make a given encoding strategy more or less successful. The current study aimed to provide additional evidence on how directed use of an effective encoding strategy would impact WM-related differences in long-term memory performance. Additionally, to my knowledge, no prior work has examined WM-related differences in cue properties at encoding. This can provide insight on WM-related differences in the effectiveness of generated cues.

Experiment 1 compared cued recall performance for items studied with self-generated cues (descriptive cue vs. any cue) and items studied with allocated cues. This manipulation was inspired by Tullis and Benjamin (2015; Experiments 1 and 3), and aims to expand on the results they found while also considering individual differences in WM. Experiment 2 compared cued recall performance and free recall performance for items studied with self-generated cues and items studied with allocated cues. Various properties of the generated cues were also examined in both experiments to determine if there were WM-related differences in the distinctiveness and uniqueness of generated cues.

Overall, in both experiments, I predicted that recall performance would be better (higher proportion correct, shorter recall latency, fewer recall errors) for the generated cue conditions relative to the allocated cue condition. This result would be consistent with prior literature on cue generation (e.g., Tullis & Benjamin, 2015); however, previous studies have only manipulated allocation of cues at retrieval and examined the effect on proportion correct. If allowing participants to generate their own cue results in more distinct cues, as has been proposed in the self-generation literature (e.g., Wheeler & Gabbert, 2017), these cues would theoretically be better for narrowing the search set (indexed by recall latency). The generated cues would match fewer target words and would likewise lead to fewer confusions with other targets (i.e., intrusion errors). Consistent with Tullis & Benjamin, I also predicted that generated cues would be more distinct and unique when participants could generate any cue (generated cue condition) relative to when they were specifically required to generate a descriptive cue (generated descriptive cue condition).

Additionally, I predicted that recall performance would be better for high WM individuals relative to low WM individuals, consistent with prior literature (Unsworth, 2019). This would support the idea that high WM individuals are better at focusing the search set, recovery ability, and monitoring ability. I also predicted that there would be an interaction with encoding condition such that WM-related differences were reduced in the generated cue conditions. In Experiment 1, if WM-related differences were equally reduced in both the generated descriptive cue and the generated cue condition, this would suggest that WM-related differences in recall are primarily due to attention at encoding, as both of these conditions would require participants to actively attend to the target word. If WM-related differences had a larger reduction in the generation condition that produced more distinct/unique cues, this would suggest that recall differences are primarily due to WM-related differences in generating an effective cue at encoding. Additionally, in this case, I would expect to see a positive relationship between WM and cue effectiveness, such that high WM individuals are generating more distinct and unique cues relative to low WM individuals. Finally, in Experiment 2, I predicted that there may be a significant three-way interaction between encoding, retrieval, and WM. Specifically, a three-way interaction could indicate that there is a smaller difference between high and low WM individuals in the generate-cued relative to generate-free conditions than the difference between high and low WM individuals in the allocated-cued relative to allocated-free conditions. This would suggest that WM-related differences in cue generation at encoding are at least partially responsible for differences in WM- related differences in cue generation at retrieval.

Overall, the current study aimed to examine if WM-related differences in long-term memory performance were due to attention at encoding, the generation of distinctive cues at encoding, the generation of cues at retrieval, or a combination of these processes.

EXPERIMENT 1

This experiment was pre-registered (https://aspredicted.org/np3yz5.pdf). Materials, data, and analysis files for this and subsequent experiments are available on OSF (https://osf.io/wkb2x/overview?view_only=8e1331a867a74416a0def423ff71f4cc).

Experiment 1 compared cued recall performance for items studied with self-generated cues (descriptive cue vs. any cue) and items studied with allocated cues. Additionally, WM was measured to examine WM-related differences in cued recall performance and to determine if self-generating cues during encoding could reduce these differences in performance. Finally, the effect of encoding condition and WM on the effectiveness of generated cues was examined.

Method

Participants

Three hundred and three participants were recruited for Experiment 1. A power analysis was conducted on G*Power 3.1 to detect a small effect size of $\eta_p^2 = .04$ for the interaction between condition and WM in proportion correct with 80% power in an analysis of covariance (ANCOVA). With 244 participants, an effect size of $\eta_p^2 = .04$ would be significant at $p < .05$ with 80% power. Additional participants were added to this estimate to account for potential exclusions, outliers, or issues with data collection.

The participants were undergraduate students recruited through Sona and compensated with course credit. Twenty participants were excluded from the final sample due to program crashes, reporting of colorblindness, or generating more than 50% of invalid cues (e.g., outputting multiple words or the same word as the target word) in one or more conditions. This produced a final sample of N = 283 (52.3% women, 45.6% male, 1.4% nonbinary, 0.7% no response; M_{age} = 18.60, SD_{age} = 0.86).

Design and Materials

One hundred and twenty concrete nouns from the University of South Florida Free Association Norms (Nelson et al., 1998) were used for the experiment. The independent variable, Condition (generated cue vs. generated descriptive cue vs. allocated cue), was manipulated within subjects and blocked. The highest semantic associate to each target word in the association norm database was selected as the cue word in the allocated cue condition. Stimuli were further divided into 3 sets of 40 target words equated on average forward associative strength between the cue word and the target word (Nelson et al., 1998). The stimuli set was counterbalanced across conditions, and word order within a condition was randomized. Words in the cue generation condition were presented first during encoding, followed by the description generation and allocated conditions, respectively, to avoid conflating order variance with variance related to individual differences in WM ability (Goodhew & Edwards, 2019). Additionally, presentation of the generated cue condition first would avoid issues where the generated descriptive cue or allocated cue conditions may influence the types of cues generated in the unconstrained generated cue condition.

Procedure

Generation Task

The procedure for the generation task is depicted in Figure 1. All instructions and stimuli were presented on the computer. Participants were instructed that they would study multiple target words and be tested on their memory for those words. Participants first received instructions for the cue generation condition followed by the study of 40 words. After completing this block, participants received instructions for the generated descriptive cue condition and studied the next 40 words. Finally, participants received instructions for the allocated cue condition followed by study of the final 40 words. The instructions were a modified version of the instructions given in Tullis and Benjamin (2015, Experiments 1 and 3) and are provided below:

Generated Cue: "For each target word, you will generate some type of cue to help you remember the target word at a later test. At the time of the test, you will be given the cue word you generated and will be asked to remember the specific target word for each cue. Please generate the cues that will be most useful to you in remembering the target words. You can use any one word cue. However, the cue word cannot be the same as the target word."

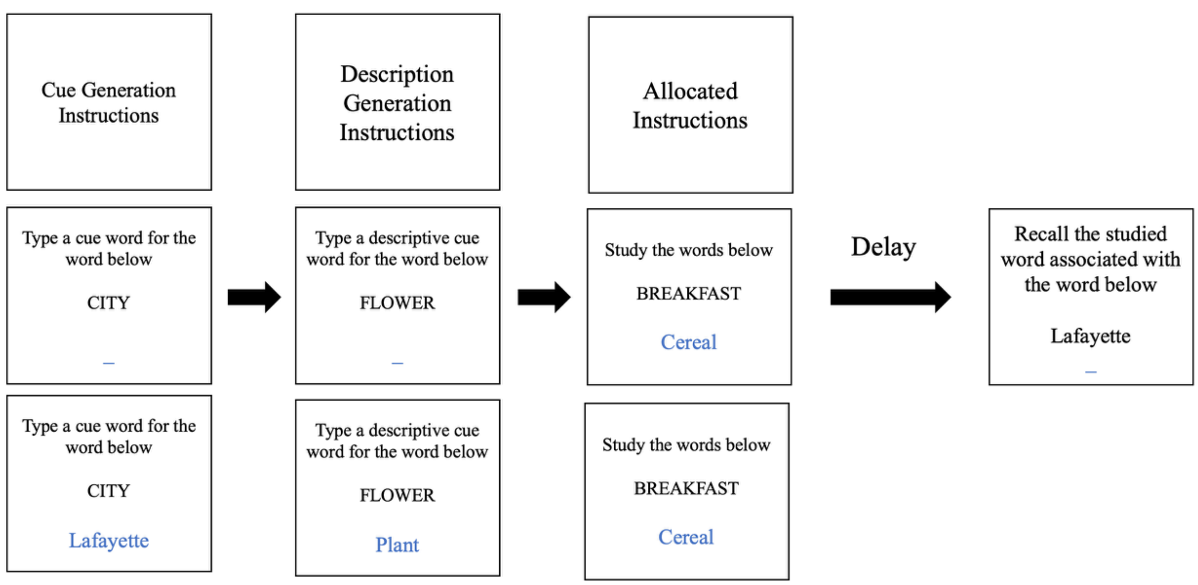
Generated Descriptive Cue: "For each target word, you will generate a cue word that constitutes an appropriate description of the target item or is an aspect of the target word. You will later be tested on your memory for the target words. At the time of test, you will be given the cue word you generated and will be asked to remember the specific target word for each cue. Once again, for each target in the list, we ask that you type in one word that, according to your own experiences, describes the target word or is an aspect of the target word. You can use any one word description. However, the cue word cannot be the same as the target word."

Allocated Cue: "For each target word, you will be shown a cue word that is related to the target word. You will later be tested on your memory for the target words. At the time of test, you will be given the cue word you saw and will be asked to remember the specific target word for each cue."

Words were presented one at a time on the screen, and the task was self-paced. After completing the study portion of the task, participants were given the WM tasks described below. These tasks were partially self-paced, though they had a typical total duration of 15-20 minutes. Following this, participants were given a cued recall test. In the cued recall test, participants were shown the cue that they saw or generated at study and were asked to recall the corresponding target word. Recall was self-paced. Word order was randomized, and words from all conditions were randomly intermixed.

Figure 1.

Experiment 1 Generation Task Procedure.

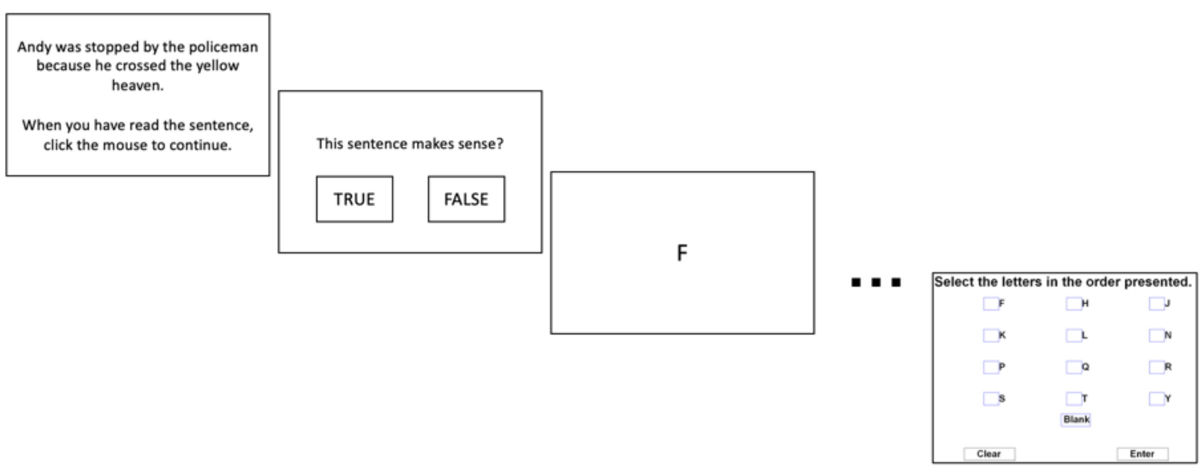


Reading Span

To assess WM, a reading span task was administered (Figure 2). This task required participants to alternate between judging if a sentence made sense in context and presentation of a letter taken from the set F, H, J, K, L, N, P, Q, R, S, T, Y. In the judgment portion of the task, participants were shown a sentence that either made sense or had one word replaced with a word that did not make sense in context. Following each sentence, a letter was presented for 1 s. This alternating sequence occurred three to seven times. Two trials of each list length were presented, for a total of 50 letters. The presentation order of different list lengths was randomized. At the end of each trial, participants were asked to recall the presented letters in order by clicking on the appropriate letters (Redick et al., 2012).

Figure 2.

Reading Span Procedure.

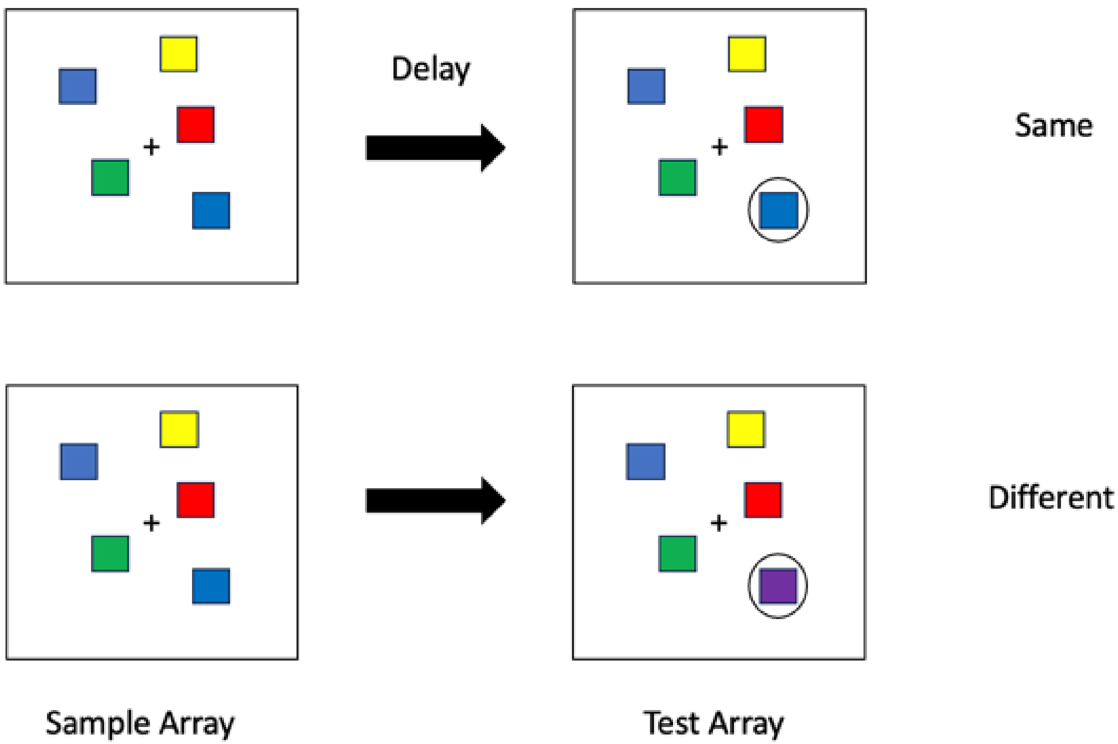


Change Detection

Additionally, to assess WM, a change detection task was administered (Figure 3). Participants were presented with multiple colored squares simultaneously on the screen in various locations. This display was shown for 100 ms followed by a blank screen for 900 ms. Next, a second display was presented that was either the same as the first, or with one square in a different color. Participants were asked to indicate if the second display was the same or different than the first. There were 16 trials of each display size, with display sizes of 4, 6, and 8 squares.

Figure 3.

Change Detection Procedure.



Analyses

Proportion correct, mean recall latency, and the proportion of recall errors in the cued recall test were analyzed. Recall latency was measured by the time between the cue onset and the last keypress.[1] Recall errors consisted of intra-list intrusions, extra-list intrusions, and omission errors. Intra-list intrusions occurred when a participant responded incorrectly with a studied target item. Extra-list intrusions occurred when a participant responded incorrectly with an item that was not presented in the experiment. Omission errors occurred when a participant did not produce a response for the target item. In the reading span task, the proportion of correct items recalled in the correct serial position was calculated. Standard capacity limit, K, was calculated for the change detection task using the equation, $K = N * (H - FA)$, where N is the relevant set size, H is the hit rate and FA is the false alarm rate (Cowan et al., 2005). Additionally, a WM z-score composite was calculated from reading span and change detection scores.

Various properties of generated cues in the cued recall task were also examined. Cue-to-target forward semantic associative strength, the number of targets associated to the cue, cumulative semantic associative strength of the cue to other targets, and the proportion of cues in the associative norm database were calculated in a similar manner as Tullis and Benjamin (2015). Cue-to-target forward associative strength was determined using the cue-to-target forward associative strength present in the normative data from the University of South Florida Free Association Norms (Nelson et al., 1998). The number of associated targets was calculated by counting the number of target words in the current experiment that had a given cue listed as an association in the University of South Florida Free Association Norms. Cumulative associative strength was calculated as the summation of forward associative strengths from a cue to all associated target words in the current experiment, subtracting out the forward associative

[1] To properly measure recall latency for use as an index of search set size, recall latency should be defined as the time between the cue onset and the first keypress. However, due to a programming error, data measuring the time between cue onset and the last keypress was collected. The results using this metric are reported; however, the results should be interpreted with caution given the measure is not as precise as the measures used in previous studies to measure recall latency.

strength between a cue and its corresponding target word that was studied by a participant. Proportion of cues in the database was calculated as the proportion of generated cues from a participant that were present as cues in the University of South Florida Free Association Norms.

A cue was considered to be more distinctive if it was associated with fewer targets and had a smaller cumulative semantic associative strength. Additionally, a cue was considered to be more unique if a lower proportion of generated cues were present in the database and cues showed lower intersubject overlap. For each participant and cue, intersubject overlap was measured by calculating the number of participants that generated the same cue in response to a given target word divided by the total number of participants that saw that given target word in each condition.

A repeated-measures ANOVA with condition (generated cue vs. generated descriptive cue vs. allocated cue) as a within-subjects factor, and an ANCOVA with condition (generated cue vs. generated descriptive cue vs. allocated cue) as a within-subjects factor and WM score as a continuous covariate, was conducted for each of the dependent measures except for the cue property metrics (forward associative strength, number of associated targets, cumulative associative strength, proportion of cues in the database, and intersubject overlap). For these measures, a paired-samples *t*-test comparing performance in the generated cue and generated descriptive cue conditions was conducted as well as an ANCOVA with condition (generated cue vs. generated descriptive cue) as a within-subjects factor and WM score as a continuous covariate. Additionally, a correlation matrix (Table 4) is provided for interpreting the direction of significant WM effects. These analyses provide insight on how different encoding tasks and working memory ability impact the types of cues generated, as well as final cued recall performance. Additionally, the interaction between condition and WM addresses if WM-related differences in long-term memory performance can be moderated by encoding task.

Results

Descriptive statistics for each condition (generated cue, generated descriptive cue, and allocated cue) are shown in Tables 2 and 3. The correlations between dependent variables are shown in Table 4.

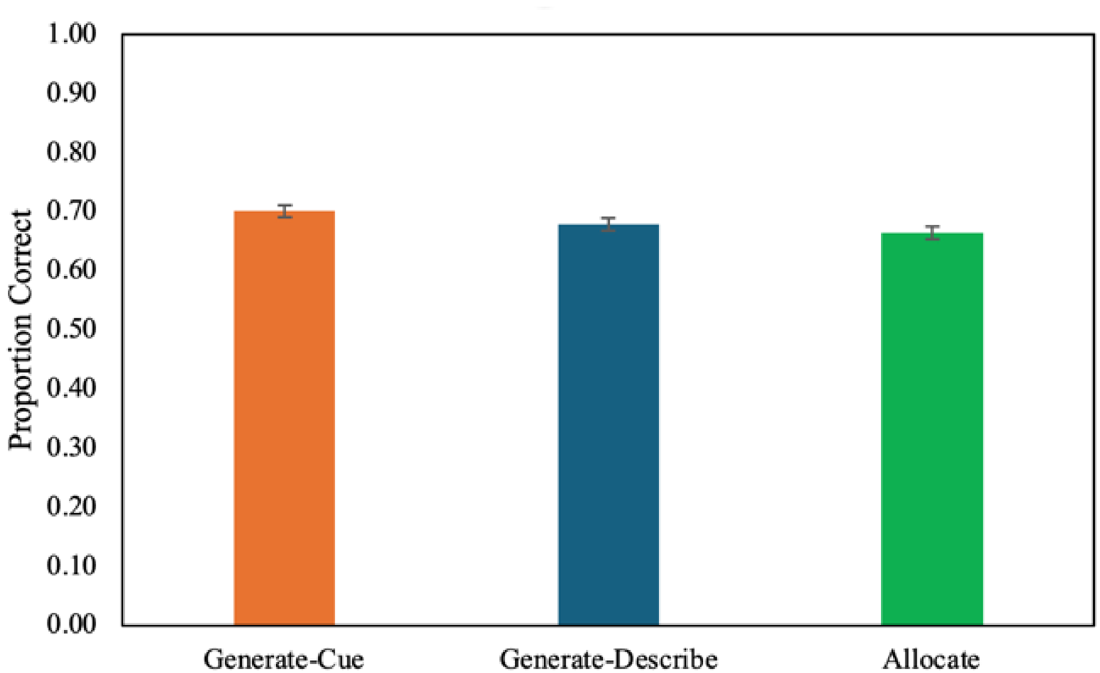
Proportion Correct

A significant difference in the proportion of words correctly recalled was observed between conditions (Figure 4). Performance was best in the generated cue condition and equivalent in the generated descriptive cue condition and allocated cue condition. This was confirmed with a repeated-measures ANOVA, $F(2, 564) = 8.22$, $MSE = .01$, $p < .001$, $\eta_p^2 = .03$. Performance was significantly higher for the generated cue condition relative to the allocated cue condition, $t(281) = 4.00$, $p < .001$, and relative to the generated descriptive cue condition, $t(281) = 2.88$, $p = .004$. However, there were no significant differences in performance between the generated descriptive cue condition and the allocated cue condition, $t(281) = 1.40$, $p = .171$.

Next to examine the effects of WM, an ANCOVA with condition (generated cue, generated descriptive cue, and allocated cue) as a within-subjects variable and WM composite *z*-score as a continuous covariate was conducted on proportion correct. See Table 4 for additional information on the correlation between WM and proportion correct, as well as other dependent variables. Mirroring the ANOVA results above, there was a main effect of condition, $F(2, 562) = 8.24$, $MSE = .01$, $p < .001$, $\eta_p^2 = .03$. The main effect of WM was also significant, $F(1, 281) = 52.56$, $MSE = .06$, $p < .001$, $\eta_p^2 = .16$, indicating that individuals with higher WM scores performed better across conditions. However, the critical interaction between condition and WM was not significant, $F(2, 562) = 1.96$, $MSE = .01$, $p = .143$, $\eta_p^2 < .01$.

Figure 4.

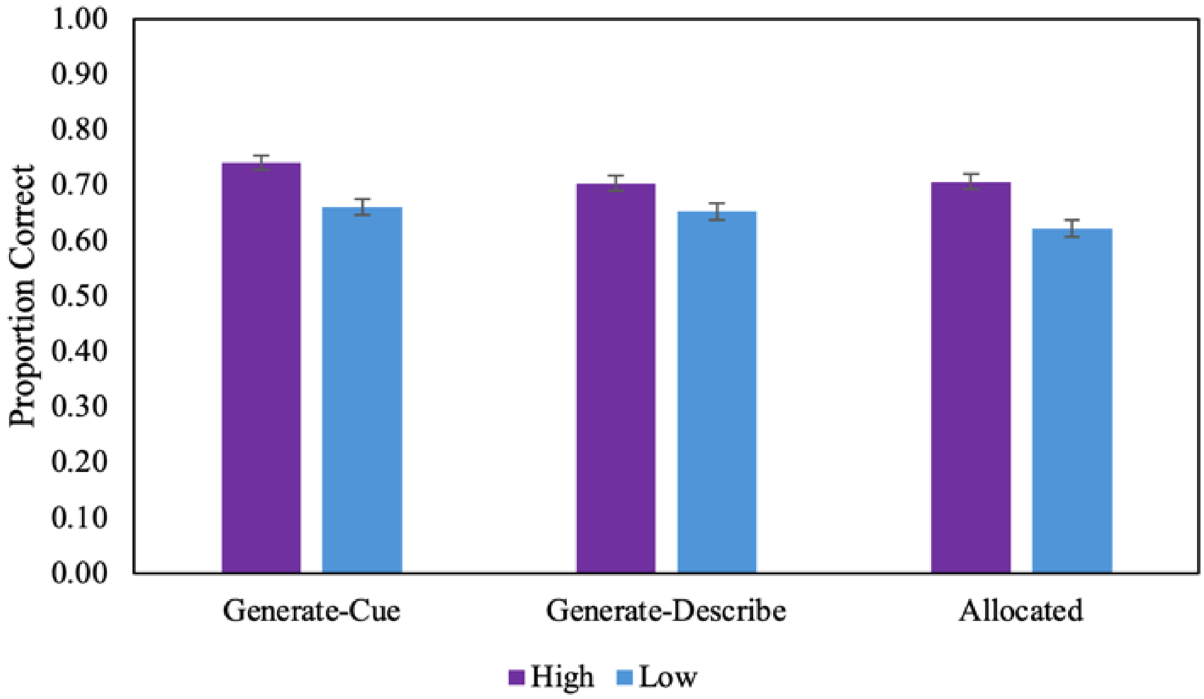
Experiment 1 Proportion Correct by Condition.



NOTE: Errors bars represent ± 1 standard error of the mean.

Figure 5.

Experiment 1 Proportion Correct by Condition and WM.



NOTE: Errors bars represent ± 1 standard error of the mean. WM is dichotomized into high and low groups for visualization. The high WM group is composed of participants with WM composite z-scores above the 50th percentile. The low WM group is composed of participants with WM composite z-scores equal to or below the 50th percentile.

Recall Latency

A significant difference in recall latency was observed between conditions. Recall latencies were lowest in the allocated cue condition and equivalent in the generated cue conditions. This was confirmed with a repeated-measures ANOVA, $F(2, 564) = 16.22$, $MSE = 334750.56$, $p < .001$, $\eta_p^2 = .05$. Recall latency was significantly lower for the allocated cue condition relative to the generated cue condition, $t(281) = -3.84$, $p < .001$, and the generated descriptive cue condition, $t(281) = 5.47$, $p < .001$. There was not a significant difference in recall latency between the generated cue condition and the generated descriptive cue condition, $t(281) = -1.96$, $p = .051$.

Next, to examine the effect of WM on recall latency, an ANCOVA with condition (generated cue, generated descriptive cue, and allocated cue) as a within-subjects variable and WM composite *z*-score as a continuous covariate was conducted. There was a main effect of condition, $F(2, 562) = 16.26$, $MSE = 333824.91$, $p < .001$, $\eta_p^2 = .06$, consistent with the ANOVA results above. The main effect of WM was also significant, $F(1, 281) = 13.74$, $MSE = 2674584.60$, $p < .001$, $\eta_p^2 = .05$, indicating that individuals with higher WM scores recalled words faster across conditions. However, the critical interaction between condition and WM was not significant, $F(2, 562) = 1.78$, $MSE = 333824.91$, $p = .169$, $\eta_p^2 < .01$.

Recall Errors

A significant difference between conditions was observed in the types of errors made by participants. Intra-list intrusions were lowest in the allocated cue condition and equivalent in the generated cue conditions. A repeated-measures ANOVA showed a significant difference in the proportion of intra-list intrusions across conditions, $F(2, 564) = 34.64$, $MSE = .001$, $p < .001$, $\eta_p^2 = .11$. A significantly lower proportion of intra-list intrusions occurred in the allocated cue condition relative to the generated cue condition, $t(281) = -6.67$, $p < .001$, and the generated descriptive cue condition, $t(281) = -7.33$, $p < .001$. There was not a significant difference in intra-list intrusions between the generated cue condition and the generated descriptive cue condition, $t(281) = -0.67$, $p = .550$.

An ANCOVA with condition (generated cue, generated descriptive cue, and allocated cue) as a within-subjects variable and WM composite *z*-score as a continuous covariate was also conducted on the proportion of intra-list intrusions. Consistent with the ANOVA results above, there was a main effect of condition, $F(2, 562) = 34.61$, $MSE = .001$, $p < .001$, $\eta_p^2 = .11$. The main effect of WM was also significant, $F(1, 281) = 15.57$, $MSE = .002$, $p < .001$, $\eta_p^2 = .05$, indicating that individuals with higher WM scores made fewer intra-list intrusion errors across conditions. However, the critical interaction between condition and WM was not significant, $F(2, 562) = 0.72$, $MSE = .001$, $p = .489$, $\eta_p^2 < .01$.

A significant difference in the proportion of extra-list intrusion errors was also observed between conditions. Extra-list intrusions were equivalent in the generated conditions and lower in the generated conditions relative to the allocated condition. This was confirmed with a repeated-measures ANOVA, $F(2, 564) = 61.98$, $MSE = .008$, $p < .001$, $\eta_p^2 = .18$. A significantly lower proportion of extra-list intrusions occurred in the generated cue condition relative to the allocated cue condition, $t(281) = -9.50$, $p < .001$. Additionally, a significantly lower proportion of extra-list intrusions occurred in the generated descriptive cue condition relative to the allocated cue condition, $t(281) = -7.89$, $p < .001$. There was not a significant difference in performance between the generated cue condition and the generated descriptive cue condition, $t(281) = -0.83$, $p = .452$.

Next to examine the effects of WM, an ANCOVA with condition (generated cue, generated descriptive cue, and allocated cue) as a within-subjects variable and WM composite *z*-score as a continuous covariate was conducted. Proportion of extra-list intrusions was the dependent variable. Mirroring the ANOVA results above, there was a main effect of condition, $F(2, 562) = 62.19$, $MSE = .008$, $p < .001$, $\eta_p^2 = .18$. The main effect of WM was also significant, $F(1, 281) = 39.08$, $MSE = .05$, $p < .001$, $\eta_p^2 = .12$, indicating that individuals with higher WM scores made fewer extra-list intrusion errors across conditions. However, the critical interaction between condition and WM was not significant, $F(2, 562) = 1.97$, $MSE = .008$, $p = .141$, $\eta_p^2 < .01$.

Finally, a significant difference in the proportion of omission errors was observed between conditions. Omission errors were lowest in the generated cue condition and equivalent in the generated descriptive cue condition and the allocated cue condition. This was confirmed with a repeated-measures ANOVA, $F(2, 564) = 6.87$, $MSE = .001$, $p < .001$, $\eta_p^2 = .02$. A significantly lower proportion of omissions occurred in the generated cue condition relative to the allocated cue condition, $t(281) = -3.67$, $p < .001$, and the generated descriptive cue condition, $t(281) = -3.00$, $p = .004$. There was not a significant difference in performance between the allocated cue condition and the generated descriptive cue condition, $t(281) = 0.33$, $p = .647$.

An ANCOVA with condition (generated cue, generated descriptive cue, and allocated cue) as a within-subjects variable and WM composite *z*-score as a continuous covariate also showed a main effect of condition on proportion of omission errors, $F(2, 562) = 6.92$, $MSE = .001$, $p < .001$, $\eta_p^2 = .02$, consistent with the ANOVA results above. The main effect of WM was not significant, $F(1, 281) = 1.67$, $MSE = .008$, $p = .197$, $\eta_p^2 < .01$. However, the interaction between condition and WM was significant, $F(2, 562) = 3.02$, $MSE = .001$, $p = .049$, $\eta_p^2 = .01$. Although none of the correlations were statistically significant, this interaction was likely driven by a numerically smaller correlation between WM and omission errors in the generated descriptive cue condition ($r = -.01$, $p = .935$) relative to the generated cue condition ($r = -.09$, $p = .152$) and the allocated cue condition ($r = -.11$, $p = .055$).

Cue Properties

A significant difference in the forward associative strength between a cue and the target was observed between conditions. This was confirmed with a paired-samples *t*-test, $t(282) = 10.91$, $p < .001$, $d = 0.65$ [.52, .78]. Higher forward associative strength was observed in the generated cue condition (M = .06, SD = .02) relative to the generated descriptive cue condition (M = .04, SD = .02).

To examine the effects of WM, an ANCOVA with condition (generated cue and generated descriptive cue) as a within-subjects variable and WM composite *z*-score as a continuous covariate was conducted. Cue-to-target forward associative strength was the dependent variable. Mirroring the *t*-test results above, there was a main effect of condition, $F(2, 281) = 120.11$, $MSE < .001$, $p < .001$, $\eta_p^2 = .30$. The main effect of WM was not significant, $F(1, 281) < 0.001$, $MSE = .001$, $p = .994$, $\eta_p^2 < .01$. Additionally, the interaction between condition and WM was not significant, $F(2, 281) = 3.44$, $MSE < .001$, $p = .065$, $\eta_p^2 = .01$.

Next, to measure cue distinctiveness, the number of associated targets to a cue and the cumulative forward associative strength of a cue to other targets were examined. A significant difference in the number of associated targets was observed between conditions. This was confirmed with a paired-samples *t*-test, $t(282) = 8.18$, $p < .001$, $d = 0.49$ [.36, .61]. Cues were associated with fewer targets in the generated descriptive cue condition (M = .70, SD = .21) relative to the generated cue condition (M = .82, SD = .19).

An ANCOVA with condition (generated cue and generated descriptive cue) as a within-subjects variable and WM composite *z*-score as a continuous covariate also showed a main effect of condition on the number of associated targets, $F(2, 281) = 66.80$, $MSE = .030$, $p < .001$, $\eta_p^2 = .19$, consistent with the *t*-test results above. The main effect of WM was also significant, F(1,

281) = 9.67, *MSE* = .052, *p* = .002, η_p^2= .03, indicating that participants with higher WM generated cues that were related to fewer targets. Additionally, the interaction between condition and WM was not significant, *F*(2, 281) = 0.54, *MSE* = .030, *p* = .463, η_p^2< .01.

A significant difference in the cumulative associative strength between a cue and other targets was observed between conditions. This was confirmed with a paired-samples *t*-test, *t*(282) = 2.00, *p* = .047, *d* = 0.12 [.002, .24]. Lower cumulative associative strength was observed in the generated descriptive cue condition (M = .02, SD = .01) relative to the generated cue condition (M = .03, SD = .01).

To examine the effects of WM, an ANCOVA with condition (generated cue and generated descriptive cue) as a within-subjects variable and WM composite *z*-score as a continuous covariate was conducted on the cumulative forward associative strength of a cue to other targets. Consistent with the *t*-test results above, there was a main effect of condition, *F*(2, 281) = 3.98, *MSE* < .001, *p* = .047, η_p^2= .01. However, the main effect of WM was not significant, *F*(1, 281) = 1.59, *MSE* < .001, *p* = .208, η_p^2< .01, nor was the interaction between condition and WM, *F*(2, 281) = 1.33, *MSE* < .001, *p* = .250, η_p^2< .01.

Finally, to measure cue uniqueness, the proportion of outputted cues in the database and intersubject overlap were examined. A significant difference in the proportion of outputted cues in the database was observed between conditions. This was confirmed with a paired-samples *t*-test, *t*(282) = 3.78, *p* < .001, *d* = .225 [.11, .34]. Fewer cues outputted in the generated descriptive cue condition were present in the database (M = .73, SD = .14) relative to the generated cue condition (M = .76, SD = .14).

Next, I conducted an ANCOVA with condition (generated cue and generated descriptive cue) as a within-subjects variable, WM composite *z*-score as a continuous covariate, and the proportion of cues in the database as the dependent variable. Mirroring the *t*-test results above, there was a main effect of condition, $F(2, 281) = 14.33$, $MSE = .01$, $p < .001$, $\eta_p^2 = .05$. The main effect of WM was also significant, $F(1, 281) = 18.47$, $MSE = .03$, $p < .001$, $\eta_p^2 = .06$, indicating that participants with higher WM produced fewer cues that were present in the database. However, the interaction between condition and WM was not significant, $F(2, 281) = 1.64$, $MSE = .01$, $p = .201$, $\eta_p^2 < .01$.

A significant difference in the proportion of overlapping cues across subjects was also observed between conditions. This was confirmed with a paired-samples *t*-test, $t(282) = 20.36$, $p < .001$, $d = 1.21$ [1.06, 1.36]. Lower intersubject overlap was observed in the generated descriptive cue condition (M = .11, SD = .03) relative to the generated cue condition (M = .17, SD = .04).

Finally, I conducted an ANCOVA with condition (generated cue and generated descriptive cue) as a within-subjects variable and WM composite *z*-score as a continuous covariate. The intersubject overlap was the dependent variable. There was a main effect of condition, $F(2, 281) = 417.74$, $MSE = .001$, $p < .001$, $\eta_p^2 = .60$, consistent with the *t*-test results above. However, the main effect of WM was not significant, $F(1, 281) = 2.94$, $MSE = .002$, $p = .088$, $\eta_p^2 = .01$, nor was the interaction between condition and WM, $F(2, 281) = 3.18$, $MSE = .001$, $p = .076$, $\eta_p^2 = .01$.

Table 2.
Experiment 1 Descriptive Statistics.

Variables	Mean	SD	Min	Max	Skew	Kurtosis
Working Memory						
Change Detection k	4.04	1.27	-0.42	6.00	-1.25	1.80
Reading span	38.52	7.42	15	50	-0.52	-0.27
Proportion Correct						
Generate-Cue	.70	.17	.23	1.00	-0.51	-0.37
Generate-Describe	.68	.18	.13	1.00	-0.49	-0.23
Allocate	.67	.18	.18	1.00	-0.21	-0.73
Recall Latency						
Generate-Cue	4845.95	1129.14	2281.58	9037.00	0.52	0.44
Generate-Describe	4944.38	1035.82	2590.53	8240.58	0.58	0.39
Allocate	4670.94	1056.87	2285.10	8626.28	0.51	0.56
Intralist Intrusions						
Generate-Cue	.05	.04	.00	.20	0.93	0.38
Generate-Describe	.05	.05	.00	.23	1.36	2.19
Allocate	.03	.03	.00	.15	1.23	1.31
Extralist Intrusions						
Generate-Cue	.20	.14	.00	.70	0.97	0.50
Generate-Describe	.20	.15	.00	.80	1.01	0.66
Allocate	.27	.17	.00	.70	0.32	-0.68
Omissions						
Generate-Cue	.03	.05	.00	.28	2.29	5.76
Generate-Describe	.04	.06	.00	.48	2.91	11.24
Allocate	.04	.07	.00	.58	3.24	15.32

NOTE: Generate-Cue = Generated Cue Condition; Generate-Describe = Generated Descriptive Cue Condition; Allocate = Allocated Cue Condition.

Table 3.

Experiment 1 Cue Properties Descriptive Statistics.

Variables	Mean	SD	Min	Max	Skew	Kurtosis
Forward Associative Strength						
Generate-Cue	.06	.02	.01	.12	0.15	-0.45
Generate-Describe	.04	.02	.00	.12	0.62	0.08
Allocate	.25	.001	.24	.25	-0.41	-1.58
Number of Associated Targets						
Generate-Cue	0.82	0.19	0.25	1.30	-0.10	-0.01
Generate-Describe	0.70	0.21	0.13	1.15	-0.31	-0.40
Allocate	1.39	0.05	1.33	1.45	0.19	-1.56
Cumulative Strength						
Generate-Cue	0.03	0.01	0.00	0.06	0.56	-0.06
Generate-Describe	0.02	0.01	0.00	0.07	0.61	0.30
Allocate	0.02	0.002	0.01	0.02	0.18	-1.48
Proportion of Cues in Database						
Generate-Cue	.76	.14	.18	1.00	-0.93	1.16
Generate-Describe	.73	.14	.25	.98	-0.93	0.86
Intersubject Overlap						
Generate-Cue	.17	.04	.06	.28	0.07	-0.16
Generate-Describe	.11	.03	.02	.20	-0.35	-0.06

NOTE: Generate-Cue = Generated Cue Condition; Generate-Describe = Generated Descriptive Cue Condition; Allocate = Allocated Cue Condition; Cumulative Strength = Cumulative Forward Associative Strength.

Table 4.

Experiment 1 Correlation Matrix.

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
1. WM Z Composite	-																												
2. Cue PC	**.37**	-																											
3. Allocate PC	**.37**	**.60**	-																										
4. Describe PC	**.28**	**.70**	**.56**	-																									
5. Cue RL	**-.21**	**-.38**	**-.22**	**-.12**	-																								
6. Allocate RL	**-.22**	**-.21**	**-.32**	-.05	**.76**	-																							
7. Describe RL	**-.15**	**-.22**	**-.17**	**-.28**	**.70**	**.68**	-																						
8. Cue Intra-list	**-.19**	**-.47**	**-.27**	**-.29**	**.24**	**.15**	**.14**	-																					
9. Allocate Intra-list	**-.15**	**-.25**	**-.44**	**-.25**	**.14**	**.24**	**.18**	**.15**	-																				
10. Describe Intra-list	**-.14**	**-.36**	**-.27**	**-.54**	.02	.04	**.19**	**.31**	**.19**	-																			
11. Cue Extra-list	**-.33**	**-.89**	**-.61**	**-.66**	**.29**	**.17**	**.18**	**.31**	**.28**	**.36**	-																		
12. Allocate Extra-list	**-.33**	**-.57**	**-.91**	**-.50**	**.16**	**.27**	.11	**.30**	**.29**	**.27**	**.65**	-																	
13. Describe Extra-list	**-.26**	**-.65**	**-.55**	**-.85**	.10	.03	**.21**	**.32**	**.31**	**.40**	**.75**	**.60**	-																
14. Cue Omission	-.09	**-.14**	-.04	-.03	**.19**	.10	**.12**	-.07	-.03	-.06	**-.15**	**-.23**	**-.20**	-															
15. Allocate Omission	-.11	-.09	**-.22**	**-.13**	.11	.10	.10	-.08	-.01	-.03	-.10	**-.17**	**-.14**	**.68**	-														
16. Describe Omission	-.01	.01	-.00	**-.14**	.07	.04	**.18**	-.09	**-.12**	-.04	**-.18**	**-.26**	**-.21**	**.58**	**.70**	-													
17. Cue FAS	.07	**.27**	.08	**.16**	**-.20**	-.10	**-.15**	-.06	-.10	-.08	**-.16**	-.04	-.07	-.12	-.08	-.07	-												
18. Allocate FAS	.03	.02	.02	-.01	-.11	**-.19**	-.11	.03	**-.28**	.10	-.06	.03	-.04	-.02	-.01	-.02	**.20**	-											
19. Describe FAS	-.07	-.05	-.08	**.26**	.03	.07	**-.23**	.02	-.07	**-.14**	.09	**.15**	**-.15**	-.06	**-.13**	**-.25**	**.16**	**.22**	-										
20. Cue Num Targets	**-.13**	-.09	-.10	-.08	.06	-.00	-.02	**.33**	.01	**.20**	**.16**	**.13**	**.19**	-.05	-.08	-.10	**.43**	**.17**	**.14**	-									
21. Allo Num Targets	-.03	.03	.03	.02	.05	.11	.08	-.04	**-.19**	-.11	.03	-.09	.04	.02	.07	.01	-.08	**-.78**	**-.21**	**-.19**	-								

Table 4 (*continued*)

22. Desc Num Targets	**-.16**	**-.19**	**-.14**	.03	**.14**	.11	-.05	**.24**	.07	**.17**	**.20**	**.18**	.07	-.05	-.09	**-.20**	.09	**.17**	**.63**	**.29**	**-.20**	-							
23. Cue C-FAS	-.01	**-.17**	**-.15**	**-.14**	**.20**	**.14**	.11	**.32**	.12	**.13**	**.20**	**.17**	**.15**	-.04	-.05	.04	.02	-.08	.00	**.52**	-.06	.11	-						
24. Allocate C-FAS	.01	-.07	-.08	-.02	.08	.09	.04	.03	.18	.04	.04	.11	-.01	-.01	-.09	.01	**-.16**	**-.17**	.03	.06	**-.48**	.07	**.21**	-					
25. Describe C-FAS	-.10	**-.14**	.04	-.07	.09	.04	.03	**.14**	.14	**.27**	**.13**	.04	.10	-.01	-.03	**-.12**	.03	.07	**.17**	**.16**	-.11	**.62**	.02	.08	-				
26. Cue Database	**-.19**	**-.27**	**-.16**	**-.18**	.09	.03	.03	**.39**	.06	**.27**	**.34**	**.17**	**.31**	.03	-.05	-.09	**.34**	.11	**.17**	**.78**	-.08	**.39**	**.34**	-.03	**.29**	-			
27. Describe Database	**-.25**	**-.27**	**-.17**	**-.23**	**.17**	.08	**.13**	**.34**	.08	**.34**	**.32**	**.18**	**.36**	-.00	-.04	-.00	.08	.05	**.28**	**.46**	-.04	**.67**	**.24**	-.01	**.45**	**.61**	-		
28. Cue Overlap	-.02	.11	.03	.03	**-.14**	-.08	-.08	.09	-.02	.04	.02	-.00	.08	-.09	-.07	-.10	**.50**	-.07	**.16**	**.43**	**.18**	**.27**	.05	**-.19**	**.18**	**.55**	**.33**	-	
29. Describe Overlap	**-.17**	**-.16**	**-.12**	.02	.05	.02	**-.13**	**.26**	.03	.10	**.19**	**.18**	.10	-.10	**-.12**	**-.22**	**.12**	**.15**	**.58**	**.35**	**-.14**	**.75**	**.12**	.01	**.41**	**.40**	**.62**	**.28**	-

NOTE: Bolded values represent significant correlations at p < .05. WM = Working Memory; Cue = Generated Cue Condition; Allocate = Allocated Cue Condition; Describe = Generated Descriptive Cue Condition; PC = Proportion Correct; RL = Recall Latency; Intra-list = Intra-list Intrusion; Extra-list = Extra-list Intrusion; FAS = Forward Associative Strength; Allo = Allocated Cue Condition; Num Targets = Number of Associated Targets; Desc = Generated Descriptive Cue Condition; C-FAS = Cumulative Forward Associative Strength; Database: Proportion of Cues in the Database; Overlap = Intersubject Overlap.

Discussion

The results of Experiment 1 were partially consistent with my predictions and previous results (e.g., Tullis & Benjamin, 2015, Experiment 1). As predicted, proportion correct was higher in the generated cue condition relative to the descriptive cue condition and the allocated cue condition. However, contrary to predictions, there was no difference in proportion correct between the generated descriptive cue condition and the allocated cue condition. To my knowledge, previous research has not directly compared performance in a generated descriptive cue condition and an allocated cue condition when the cue is allocated during encoding. For example, Tullis and Benjamin compared performance when participants were given their own descriptive cue or someone else's descriptive cue at retrieval. In this situation, cues at encoding are different from those at retrieval. Given this, performance in the analogous allocated condition of Tullis and Benjamin (2015) is much lower ($M \cong .3$) than performance in the allocated condition of the current study ($M = .67$, $SD = .18$).

Error analyses were also partially consistent with my predictions. The pattern of extra-list intrusions was as expected, with the lowest proportion of errors occurring in the generated cue condition, followed by the generated descriptive cue condition, and finally the allocated cue condition. Omissions showed a similar pattern, though there was no difference between the generated descriptive cue condition and the allocated cue condition. However, intra-list intrusions showed a contradicting pattern with the lowest proportion of errors occurring in the allocated cue condition, followed by the generated cue condition, and then the generated descriptive cue condition. This may have been influenced by the finding that cue-to-target forward associative strength was strongest in the allocated cue condition (see Appendix A for cue property statistics in the allocated cue condition), followed by the generated cue condition, and then the generated descriptive cue condition. However, cues in the allocated condition were also associated with more targets and had a higher cumulative associative strength relative to the generate conditions. Another possibility is that the nature of the generate task required participants to think of prior target words in order to not use those target words as cues for the current word and/or to not repeat cues generated for other targets. It may be the case that reinstatement of the other target words during encoding of the current target word leads to formation of associations between the items and potential intrusion during retrieval.

The result pattern for the properties of generated cues was not consistent with predictions. Analyses of the properties of cues indicated that more distinctive and unique cues were generated in the generated descriptive cue condition relative to the generated cue condition. Cues in the descriptive cue condition were associated with fewer targets, had a smaller cumulative forward associative strength, were less likely to be present in the database, and had lower intersubject overlap. Based on the results of Tullis and Benjamin (2015; Experiment 1), I expected the opposite pattern. This may have been influenced by the finding that fewer cues from my sample produced in the generated descriptive cue condition were present in the database relative to the generated cue condition. If a cue was not present in the database, it would likewise not be associated with other targets nor have a nonzero value for cumulative forward associative strength. Intersubject overlap, however, would not be influenced by this limitation. Another factor that may have influenced the discrepancy between the current experiment and Tullis and Benjamin's experiment 1 is the design. In the current study, encoding condition was manipulated within-subjects rather than between-subjects as in Tullis and Benjamin. Though blocking of conditions in the current study was intended to minimize influence of a specific condition on other conditions, it is possible that completing the generated cue condition prior to the generated descriptive cue condition influenced the types of cues that participants outputted in the generated descriptive cue condition in a way that would not be present in a between-subjects design. Finally, the calculation method for the number of targets associated with a cue and the calculation method for cumulative associative strength was different in Tullis and Benjamin (2015) and the current study. Tullis and Benjamin calculated these measures using the entire dataset available in the University of Florida free association norms, while the current study calculated these measures using only the studied target items of the current experiment. This change was made to better reflect the distinctiveness of a cue relative to studied items which would also potentially be in the memory search set, rather than all possible words, both experimental and non-experimental. However, using the data from the current experiment, analyses were also performed on these measures using the same calculation method as Tullis and Benjamin (2015). Consistent with the results of the current study, cues were associated with fewer targets in the generated descriptive cue condition relative to the generated cue condition. Additionally, there was not a significant difference in the cumulative associative strength between a cue and other target observed between conditions (see Appendix A for detailed statistics). Given this, the calculation method does not seem to be a primary factor for the differing results.

The pattern of WM results was consistent with predictions and the findings of Unsworth (2009b). Individuals with high WM correctly recalled a higher proportion of words, recalled words faster, and made fewer intra-list and extra-list intrusions relative to individuals with low WM. However, there were no WM-related differences in omission errors. This suggests that high WM individuals are better at focusing the search set and engaging in effective monitoring processes during recall.

Novel to the current study, WM-related differences in the properties of generated cues were examined. There was partial evidence that individuals with high WM generated more distinctive and unique cues (cues were associated with fewer targets and were less likely to be in the database). However, there were no WM-related differences in cumulative semantic associative strength, intersubject overlap, or cue-to-target forward associative strength. This finding partially supports the idea that high WM individuals use more effective cues to guide recall.

Across all the dependent variables, only the interaction between condition and WM on omission errors was significant. This was partially in the predicted direction. WM was more strongly related with omissions in the allocated cue condition and the generated cue condition relative to the generated descriptive cue condition. However, the correlations between WM and omission errors in each condition were non-significant. I predicted that there would be other significant interactions between condition and WM for proportion correction, recall latency, errors, and cue properties. Although encoding condition did influence performance on these dependent variables, it did not appear to facilitate more effective encoding for low WM individuals as predicted. This may suggest that WM-related differences in recall performance are driven primarily by generation of different types of cues during encoding (e.g., inter-item cues) or by retrieval factors as opposed to encoding factors.

EXPERIMENT 2

This experiment was pre-registered (https://aspredicted.org/qi2vt5.pdf). Experiment 2 aimed to replicate the results of Experiment 1 as well as to examine the role of retrieval factors by varying the level of cue support at retrieval. Participants either generated a cue at encoding or were allocated a cue at encoding. At retrieval, participants were either given a cued recall test or a free recall test. Identical to Experiment 1, WM was measured to examine WM-related differences in recall performance, and to determine if self-generating cues during encoding could reduce these differences in performance. Additionally, prior research has suggested that WM-related differences in performance are larger in free recall relative to cued recall. One proposed explanation for this effect is that individuals with high WM use more effective cues at retrieval. This could, however, be due to initially generating effective cues at encoding. Given this, another aim of Experiment 2 was to examine if the effect size for the interaction between WM and retrieval condition would be reduced when cues were generated at encoding. Finally, the effect of WM on the effectiveness of generated cues was examined.

Method

Participants

Three hundred and one participants were recruited for Experiment 2. A power analysis was conducted on G*Power 3.1 to detect a small effect size of η_p^2 = .04 for the interaction between encoding condition and WM with 80% power in an analysis of covariance (ANCOVA). With 277 participants, a small effect size of η_p^2 = .04 would be significant at $p < .05$ with 80%

power. Additional participants were added to this estimate to account for potential exclusions, outliers, or issues with data collection.

Participants were undergraduate students recruited through Sona and compensated with course credit. Nineteen participants were excluded from the final sample due to program crashes, reporting of colorblindness, or generating more than 50% of invalid cues (e.g., outputting multiple words or the same word as the target word) in one or more conditions. This produced a final sample of N = 282 (47.9% women, 50.4% male, 1.4% nonbinary, M_{age}= 19.27, SD_{age}= 1.20).

Design and Materials

One hundred and twenty concrete nouns from the University of South Florida Free Association Norms (Nelson et al., 1998) were used for the experiment. The independent variables, encoding condition (allocated cue vs. generated cue) and retrieval condition (cued recall vs. free recall), were manipulated between-subjects for a total of 4 groups. Participants were randomly assigned to a condition. Word order was randomized within a condition. For the allocated cue conditions, cue words were selected randomly from the responses of Experiment 1 participants in the cue generation condition. These cues were screened prior to Experiment 2 to correct for any spelling mistakes, incomplete responses, and responses that did not comply with the instructions (i.e., multi-cue responses, repeats or anagrams of the target word, etc.).

Procedure

Generation Task

The procedure for the generation task is depicted in Figure 5. All instructions and stimuli were presented on the computer. Participants were first instructed that they would study multiple target words and be tested on their memory for those words. Additionally, participants received instructions regarding the encoding task (i.e., allocated or generated) and the format of the final test (i.e., cued recall or free recall). The instructions are provided below:

Allocated/Cued Recall: "For each target word, you will be shown a cue word to help you remember the target word at a later test. At the time of the test, you will be given the cue word you saw and will be asked to remember the specific target word for each cue. In the upcoming task, please study both the target and cue words."

Allocated/Free Recall: "For each target word, you will be shown a cue word to help you remember the target word at a later test. However, at the time of the test, you will not be given the cue word or target word. You will be asked to recall the studied target words. In the upcoming task, please study both the target and cue words."

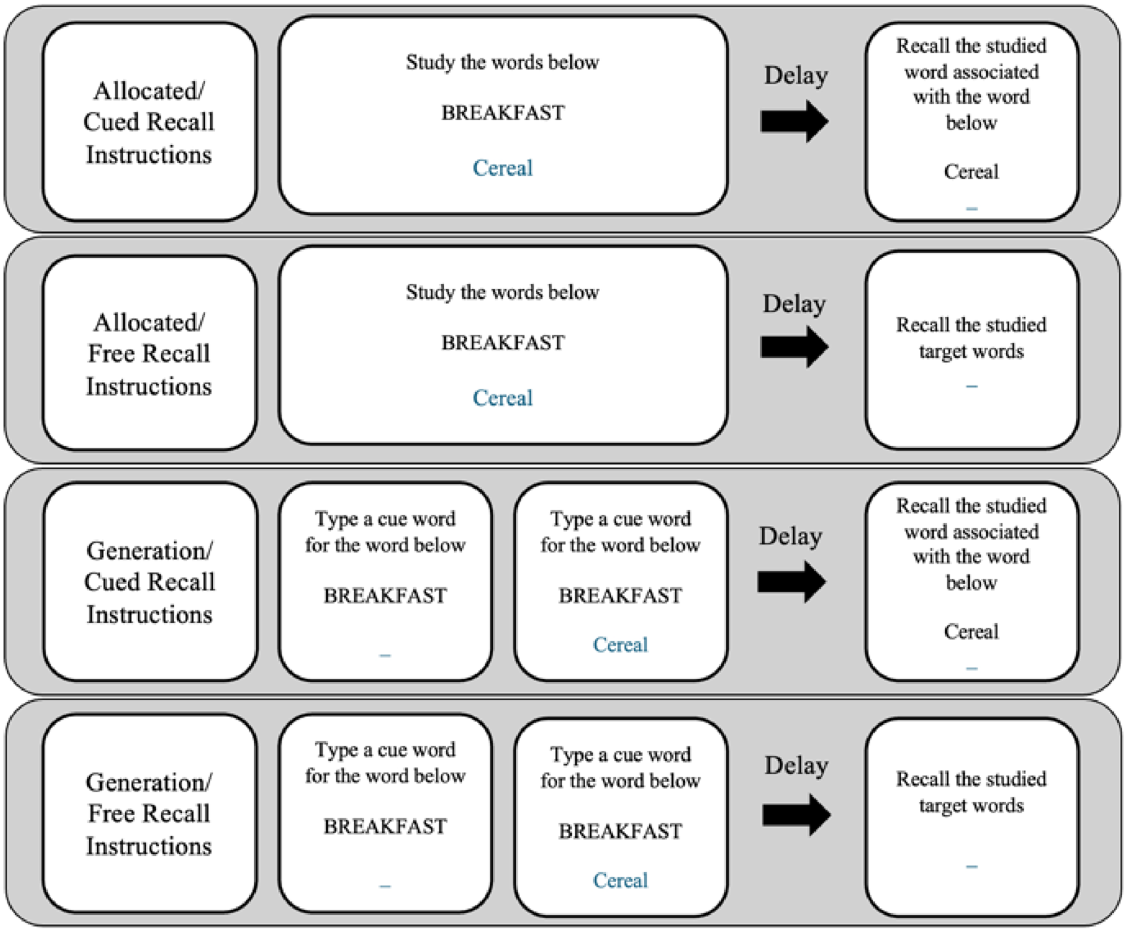
Generated/Cued Recall: "For each target word, you will generate some type of cue to help you remember the target word at a later test. At the time of the test, you will be given the cue word you generated and will be asked to remember the specific target word for each cue. Please generate the cues that will be most useful to you in remembering the target words. You can use any one word cue. However, the cue word cannot be the same as the target word."

Generated/Free Recall: "For each target word, you will generate some type of cue to help you remember the target word at a later test. However, at the time of the test, you will not be given the cue word or target word. You will be asked to recall the studied target words. Please generate the cues that will be most useful to you in remembering the target words. You can use any one word cue. However, the cue word cannot be the same as the target word."

Participants studied and were tested on all 120 target words. Words were presented one at a time on the screen. The task was self-paced. Participants in the allocated cue conditions were shown the target word with a corresponding cue word presented below. Participants in the generated cue conditions were shown the target word and asked to generate a cue word to help them remember the target word. After completing the study portion of the task, participants were given the WM tasks described below. Following this, participants were either given a cued recall test or a free recall test. In the cued recall test, participants were shown the cue that they saw or generated at study and were asked to recall the corresponding target word. In the free recall test, participants were asked to recall as many target words as they could remember, but they were not given any cue words. Recall for both tests was self-paced.

Figure 6.

Experiment 2 Generation Task Procedure.



Reading Span

The same version of the reading span task as used in Experiment 1 was used in Experiment 2.

Change Detection

The same version of the change detection task as used in Experiment 1 was used in Experiment 2.

Analyses

Proportion correct, mean recall latency, and the proportion of recall errors in the final recall test was analyzed. For the cued recall test, recall latency was measured by the time between the cue onset and the first keypress. For the free recall test, recall latency was measured by the time between pressing the enter key to start the trial/submit a response and onset of the first letter keypress. Recall errors consisted of intra-list intrusions, extra-list intrusions, and omission errors, calculated in the same manner as Experiment 1. Intra-list intrusions and omission errors were only analyzed in the cued recall test while extra-list intrusions were analyzed in both cued and free recall tests.

Various properties of generated cues in the recall task were also examined. Cue-to-target forward semantic associative strength, the number of targets associated to the cue, cumulative semantic associative strength of the cue to other targets, the proportion of cues in the associative norm database, and intersubject overlap were calculated in the same manner as Experiment 1. A cue was considered to be more distinctive if it was associated with fewer targets and had a smaller cumulative semantic associative strength. Additionally, a cue was considered to be more unique if a lower proportion of generated cues were present in the database and cues showed lower intersubject overlap.

Finally, reading span score and standard capacity, K, in the color change detection task were calculated in the same manner as Experiment 1. Additionally, a WM *z*-score composite was calculated from the reading span and change detection scores.

An univariate ANOVA with encoding condition (allocated cue vs. generated cue) and retrieval condition (cued recall vs. free recall) as between-subjects factors, as well as an ANCOVA with encoding condition (allocated cue vs. generated cue) and retrieval condition (cued recall vs. free recall) as between-subjects factors and WM score as a continuous covariate, was conducted for proportion correct, recall latency, and extra-list intrusions. An univariate ANOVA with encoding condition (allocated cue vs. generated cue) as a between-subjects factor, as well as an ANCOVA with encoding condition (allocated cue vs. generated cue) as a between-subjects factor and WM score as a continuous covariate, was conducted for intra-list and omission errors. Finally, correlations between WM and cue properties were examined. For analyses using only a subset of participants (i.e., only cued recall conditions or only generate conditions), a WM *z*-score was calculated using only the data from those participants rather than the full sample.

Results

Descriptive statistics for each condition (generated-cued, generated-free, allocated-cued, and allocated-free) are shown in Tables 5 and 6. The correlations between dependent variables are shown in Tables 7, 8, 9, and 10.

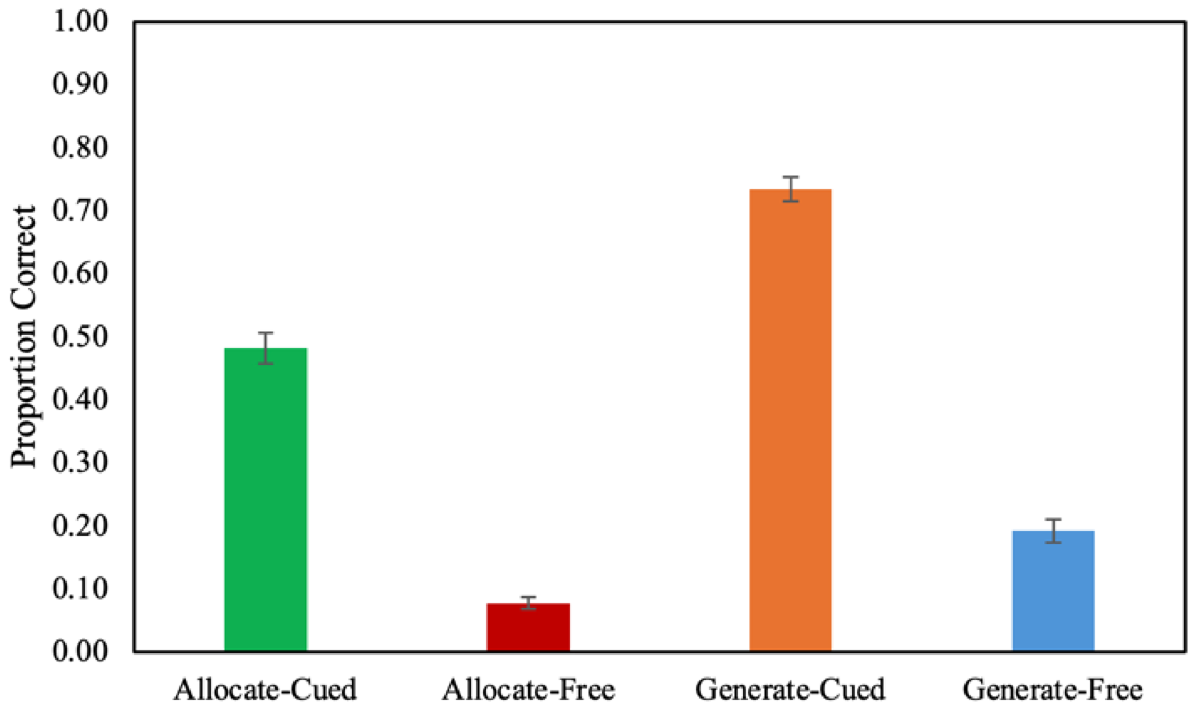
Proportion Correct

A significant difference in the proportion of words correctly recalled was observed between conditions (Figure 6). Performance was best in the generated-cued condition, followed by the allocated-cued condition, the generated-free condition, and finally the allocated-free condition. This was confirmed with a 2-way ANOVA with encoding condition and retrieval condition as between-subject factors. There was a significant main effect of encoding condition, $F(1, 278) = 93.10$, $MSE = .03$, $p < .001$, $\eta_p^2 = .25$. Performance was significantly higher for the generated cue conditions relative to the allocated cue conditions. Additionally, there was a significant main effect of retrieval condition, $F(1, 278) = 625.25$, $MSE = .03$, $p < .001$, $\eta_p^2 = .69$. Performance was significantly higher for the cued recall conditions relative to the free recall conditions. The interaction between encoding and retrieval conditions was also significant, $F(1, 278) = 13.05$, $MSE = .03$, $p < .001$, $\eta_p^2 = .05$. The difference between cued and free recall was larger in the generate conditions, $t(132) = 19.36$, $p < .001$, $d = 3.55$ [3.01, 4.1], relative to the allocated conditions, $t(147) = 15.58$, $p < .001$, $d = 2.48$ [2.05, 2.90].

To examine the effects of WM on proportion correct, I conducted an ANCOVA with encoding condition and retrieval condition as between-subjects factors and WM composite *z*-score as a continuous covariate. Mirroring the ANOVA results above, a significant main effect of encoding condition was observed, $F(1, 274) = 93.97$, $MSE = .03$, $p < .001$, $\eta_p^2 = .26$, as well as a significant main effect of retrieval condition, $F(1, 274) = 615.73$, $MSE = .03$, $p < .001$, $\eta_p^2 = .69$. The interaction between encoding and retrieval conditions was also significant, $F(1, 274) = 12.34$, $MSE = .03$, $p < .001$, $\eta_p^2 = .04$. There was a significant main effect of WM, $F(1, 274) = 9.75$, $MSE = .03$, $p = .002$, $\eta_p^2 = .03$, indicating that individuals with high WM recalled a higher proportion of items correctly relative to low WM individuals. However, the two way interaction between encoding condition and WM, $F(1, 274) = 0.02$, $MSE = .03$, $p = .884$, $\eta_p^2 < .01$, as well as the two way interaction between retrieval condition and WM, $F(1, 274) = 0.69$, $MSE = .03$, $p = .409$, $\eta_p^2 < .01$, were nonsignificant. Finally, the three way interaction between encoding condition, retrieval condition, and WM was not significant, $F(1, 274) = 0.17$, $MSE = .03$, $p = .685$, $\eta_p^2 < .01$.

Figure 7.

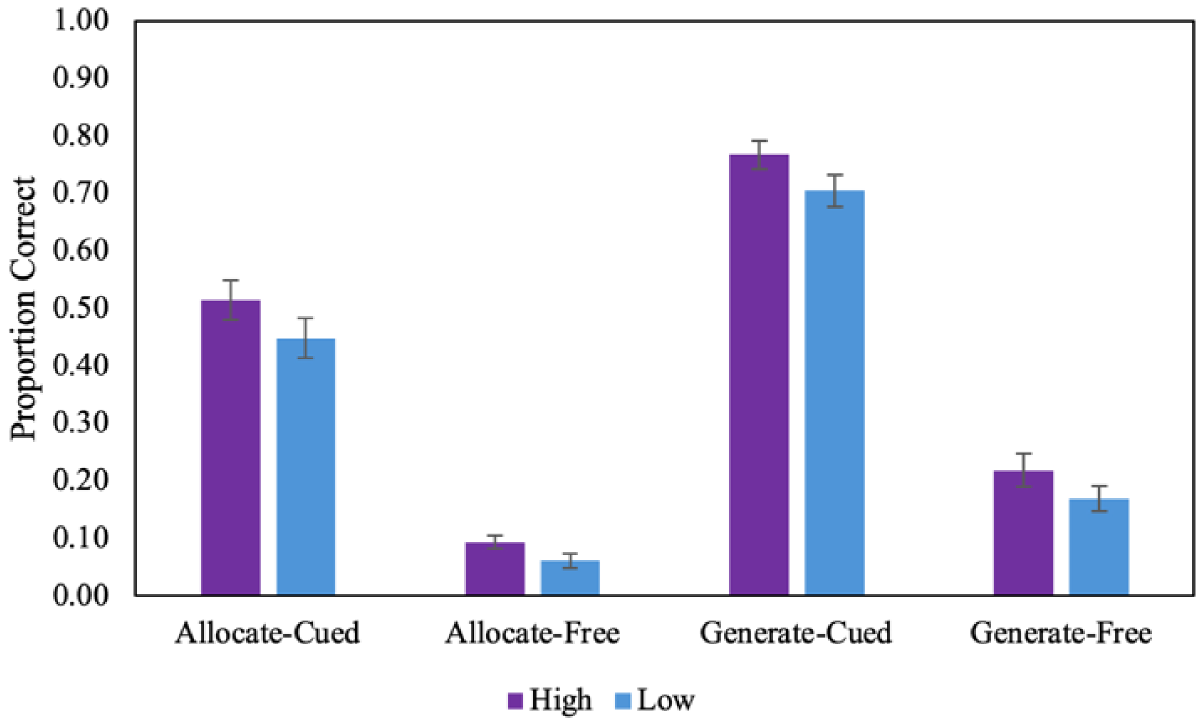
Experiment 2 Proportion Correct by Condition.



NOTE: Errors bars represent ± 1 standard error of the mean.

Figure 8.

Experiment 2 Proportion Correct by Condition and WM.



NOTE: Errors bars represent ± *1 standard error of the mean. WM is dichotomized into high and low groups for visualization. The high WM group is composed of participants with WM composite z-scores above the 50th percentile. The low WM group is composed of participants with WM composite z-scores equal to or below the 50th percentile.*

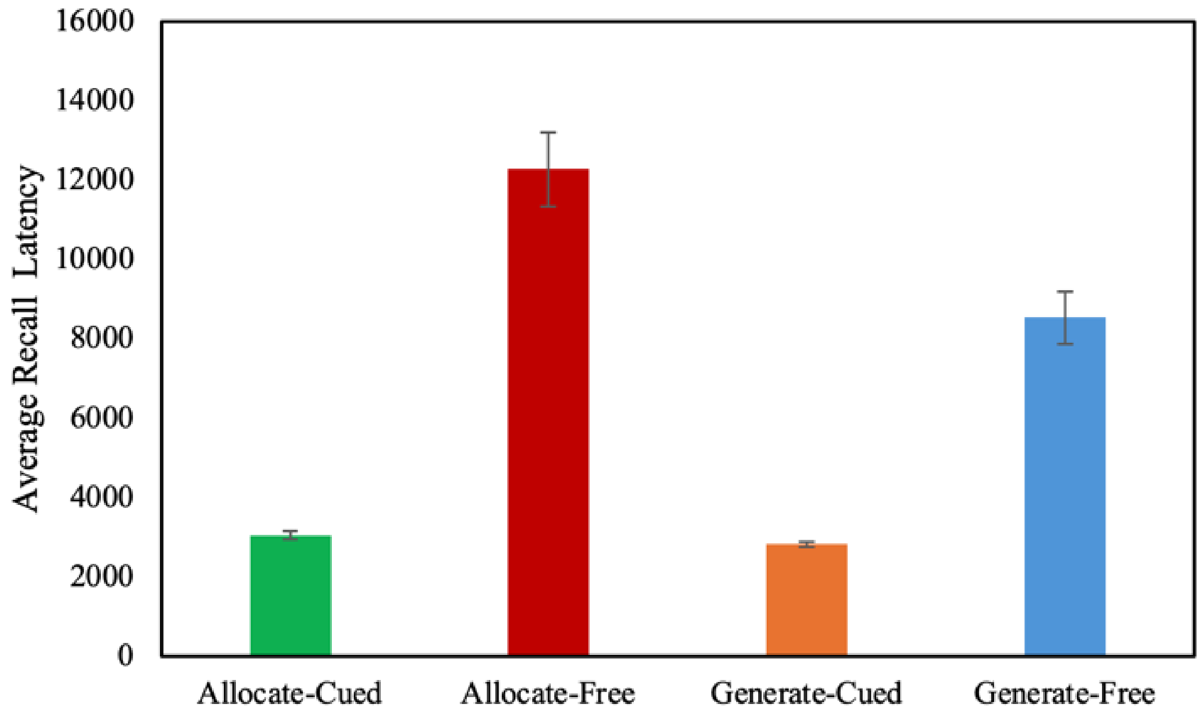
Recall Latency

A significant difference in recall latency was observed between conditions (Figure 7). Recall latencies were lowest in the cued recall conditions, followed by the generate-free condition, and the allocate-free condition. This was confirmed with a 2-way ANOVA with encoding condition and retrieval condition as between-subject factors. There was a significant main effect of encoding condition, $F(1, 278) = 12.59$, $MSE = 21870397.98$, $p < .001$, $\eta_p^2 = .04$. Recall latency was significantly lower for the generated cue conditions relative to the allocated cue conditions. Additionally, there was a significant main effect of retrieval condition, $F(1, 278) = 178.83$, $MSE = 21870397.98$, $p < .001$, $\eta_p^2 = .39$. Recall latency was significantly lower for the cued recall conditions relative to the free recall conditions. The interaction between encoding and retrieval conditions was also significant, $F(1, 278) = 9.84$, $MSE = 21870397.98$, $p < .001$, $\eta_p^2 = .03$. There was no significant effect of encoding condition on recall latency when participants were tested with a cued recall test, $t(147) = 0.30$, $p = .765$. However, recall latency was significantly lower for the generate condition relative to the allocate condition when tested with a free recall test, $t(131) = 4.60$, $p < .001$.

An ANCOVA with encoding condition and retrieval condition as between-subjects factors and WM composite *z*-score as a continuous covariate produced similar results for recall latency. Consistent with the ANOVA results above, a significant main effect of encoding condition was observed, $F(1, 274) = 12.14$, $MSE = 21915053.49$, $p < .001$, $\eta_p^2 = .04$, as well as a significant main effect of retrieval condition, $F(1, 274) = 179.93$, $MSE = 21915053.49$, $p < .001$, $\eta_p^2 = .40$. The interaction between encoding and retrieval conditions was also significant, $F(1, 274) = 9.39$, $MSE = 21915053.49$, $p = .002$, $\eta_p^2 = .03$. The main effect of WM was not significant, $F(1, 274) = 1.58$, $MSE = 21915053.49$, $p = .210$, $\eta_p^2 < .01$. Additionally, the two way interaction between encoding condition and WM, $F(1, 274) = 0.04$, $MSE = 21915053.49$, $p = .851$, $\eta_p^2 < .01$, as well as the two way interaction between retrieval condition and WM, $F(1, 274) = 1.49$, $MSE = 21915053.49$, $p = .223$, $\eta_p^2 < .01$, were not significant. Finally, the three way interaction between encoding condition, retrieval condition, and WM was not significant, $F(1, 274) = 0.01$, $MSE = 21915053.49$, $p = .938$, $\eta_p^2 < .01$.

Figure 9.

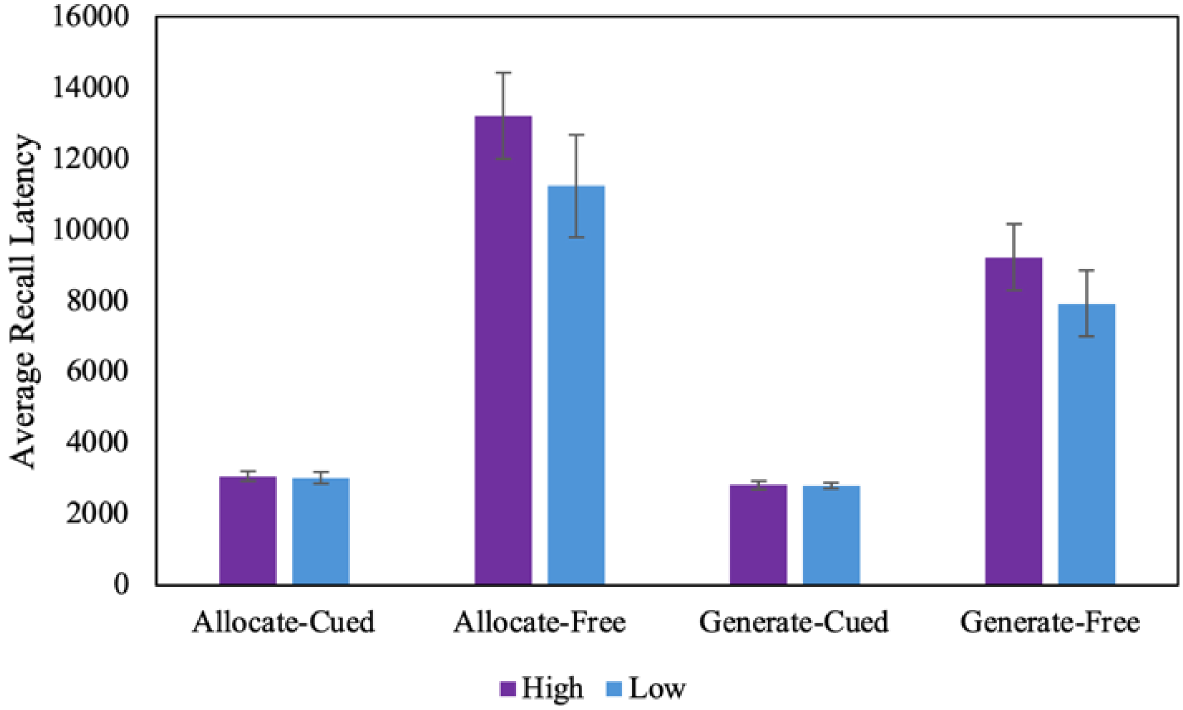
Experiment 2 Mean Recall Latency by Condition.



NOTE: Errors bars represent ± 1 standard error of the mean.

Figure 10.

Experiment 2 Mean Recall Latency by Condition and WM.



NOTE: Errors bars represent ± 1 standard error of the mean. WM is dichotomized into high and low groups for visualization. The high WM group is composed of participants with WM composite z-scores above the 50th percentile. The low WM group is composed of participants with WM composite z-scores equal to or below the 50th percentile.

Recall Errors

To analyze recall errors, intra-list intrusions and omissions were examined in cued recall. Intra-list intrusions were lower in the allocate-cued condition than the generate-cued condition. However, omission errors were lower in the generate-cued condition relative to the allocate-cued condition. This was confirmed with independent-samples t-tests. There was a significant difference between conditions in intra-list intrusions, $t(147) = -3.54$, $p < .001$, $d = -0.58$ [-.91, -.25]. Intra-list intrusions were significantly lower for the allocate condition (M = .03, SD = .02) relative to the generate condition (M = .04, SD = .03). There was also a significant difference between conditions in omission errors, $t(147) = -6.18$, $p < .001$, $d = -1.01$ [-1.35, -.67]. Omission errors were significantly lower for the generate condition (M = .04, SD = .04) relative to the allocate condition (M = .18, SD = .19).

To examine the effects of WM on intra-list intrusions, an ANCOVA with encoding condition as a between-subjects factor and WM composite *z*-score as a continuous covariate was conducted. A significant main effect of encoding condition was observed, $F(1, 145) = 15.08$, $MSE = .001$, $p < .001$, $\eta_p^2 = .09$, consistent with the *t*-test results above. The main effect of WM was also significant, $F(1, 145) = 5.86$, $MSE = .001$, $p = .017$, $\eta_p^2 = .04$, indicating that individuals with high WM made fewer intra-list intrusion errors. However, the interaction between encoding condition and WM was not significant, $F(1, 145) = 3.45$, $MSE = .001$, $p = .065$, $\eta_p^2 = .02$.

Next, to examine the effects of WM on omission errors, an ANCOVA with encoding condition as a between-subjects factor and WM composite *z*-score as a continuous covariate was conducted. Mirroring the *t*-test results above, a significant main effect of encoding condition was observed, $F(1, 145) = 36.75$, $MSE = .02$, $p < .001$, $\eta_p^2 = .20$. The main effect of WM was not significant, $F(1, 145) = 0.54$, $MSE = .02$, $p = .465$, $\eta_p^2 < .01$. Additionally, the interaction between encoding condition and WM was not significant, $F(1, 145) = 0.64$, $MSE = .02$, $p = .427$, $\eta_p^2 < .01$.

Next, extra-list intrusions were examined in both cued and free recall conditions. A significant difference in extra-list intrusions was observed between conditions. Extra-list intrusions were lower in the generated cue conditions relative to the allocated cue conditions. This was confirmed with a 2-way ANOVA with encoding condition and retrieval condition as between-subject factors. There was a significant main effect of encoding condition, $F(1, 278) =$

15.74, *MSE* = .05, $p < .001$, $\eta_p^2 = .05$. Extra-list intrusions were significantly lower for the generated cue conditions relative to the allocated cue conditions. The main effect of retrieval condition was not significant, $F(1, 278) = 1.49$, *MSE* = .05, $p = .223$, $\eta_p^2 < .01$. The interaction between encoding and retrieval conditions was also not significant, $F(1, 278) = 2.34$, *MSE* = .05, $p = .127$, $\eta_p^2 < .01$.

Finally, an ANCOVA with encoding condition and retrieval condition as between-subjects factors and WM composite *z*-score as a continuous covariate was conducted to analyze extra-list intrusion errors. A significant main effect of encoding condition was observed, $F(1, 274) = 15.63$, *MSE* = .05, $p < .001$, $\eta_p^2 = .05$, consistent with the ANOVA results reported above. The main effect of retrieval condition was not significant, $F(1, 274) = 1.02$, *MSE* = .05, $p = .314$, $\eta_p^2 < .01$. The interaction between encoding and retrieval conditions was also not significant, $F(1, 274) = 2.41$, *MSE* = .05, $p = .121$, $\eta_p^2 < .01$. The main effect of WM was significant, $F(1, 274) = 5.53$, *MSE* = .05, $p = .019$, $\eta_p^2 = .02$, indicating that individuals with high WM made fewer extra-list intrusion errors. However, the two way interaction between encoding condition and WM, $F(1, 274) = 0.93$, *MSE* = .05, $p = .336$, $\eta_p^2 < .01$, as well as the two way interaction between retrieval condition and WM, $F(1, 274) = 0.39$, *MSE* = .05, $p = .534$, $\eta_p^2 < .01$, were not significant. Finally, the three way interaction between encoding condition, retrieval condition, and WM was not significant, $F(1, 274) = 0.07$, *MSE* = .05, $p = .797$, $\eta_p^2 < .01$.

Cue Properties

WM was not significantly correlated with forward associative strength (r = .15, $p = .079$), number of associated targets (r = .09, $p = .326$), cumulative associative strength (r = .04, *p* =

.686), proportion of cues in the database (r = .02, *p* = .796), or intersubject overlap (r = .16, *p* = .061).

Table 5.

Experiment 2 Descriptive Statistics.

Variables	Mean	SD	Min	Max	Skew	Kurtosis
Working Memory						
Change Detection k	4.12	1.17	-.83	6.00	-1.28	2.27
Reading Span	39.76	6.84	8	50	-1.11	2.43
Proportion Correct						
Generate CR	.74	.16	.19	.95	-1.01	0.62
Allocate CR	.48	.22	.07	.97	0.35	-0.75
Generate FR	.19	.14	.00	.73	1.22	2.46
Allocate FR	.08	.08	.00	.33	1.24	1.26
Recall Latency						
Generate CR	2815.21	560.81	1628.02	4647.26	0.30	0.44
Allocate CR	3045.10	934.80	1316.79	6046.18	0.66	0.29
Generate FR	8535.71	5199.58	1657.58	32228.56	2.19	7.04
Allocate FR	12270.94	7875.36	1453.83	40032.08	1.48	2.76
Intralist Intrusions						
Generate CR	.04	.03	.00	.13	0.74	-0.17
Allocate CR	.03	.02	.00	.09	1.15	1.22
Extralist Intrusions						
Generate CR	.17	.14	.02	.65	1.33	1.44
Allocate CR	.32	.22	.00	.82	0.56	-0.86
Generate FR	.24	.25	.00	.95	1.67	1.81
Allocate FR	.31	.28	.00	.93	0.86	-0.45
Omissions						
Generate CR	.04	.04	.00	.20	1.75	3.40
Allocate CR	.18	.19	.00	.73	1.12	0.43

NOTE: Generate CR = Generated Cued Recall Condition; Allocate CR = Allocated Cued Recall Condition; Generate FR = Generated Free Recall Condition; Allocate FR = Allocated Free Recall Condition.

Table 6.

Experiment 2 Cue Properties Descriptive Statistics.

Variables	Mean	SD	Min	Max	Skew	Kurtosis
Forward Associative Strength						
Generate CR	.06	.02	.03	.10	0.22	-0.54
Allocate CR	.02	.01	.01	.04	0.38	-0.39
Generate FR	.06	.02	.02	.11	0.16	-0.12
Allocate FR	.02	.01	.01	.03	-0.08	-0.89
Number of Associated Targets						
Generate CR	0.74	0.15	0.33	1.11	-0.18	-0.17
Allocate CR	0.34	0.06	0.22	0.53	0.27	0.51
Generate FR	0.79	0.14	0.49	1.08	-0.21	-0.60
Allocate FR	0.34	0.06	0.22	0.50	0.10	0.13
Cumulative Strength						
Generate CR	0.02	0.01	0.01	0.04	0.31	-0.44
Allocate CR	0.01	0.002	0.00	0.01	0.16	-0.59
Generate FR	0.02	0.01	0.01	0.05	0.58	1.68
Allocate FR	0.01	0.002	0.00	0.02	1.21	2.37
Proportion of Cues in Database						
Generate CR	.71	.14	.31	.93	-0.45	-0.52
Allocate CR	.45	.04	.32	.54	-0.46	0.53
Generate FR	.73	.12	.44	.92	-0.59	-0.36
Allocate FR	.44	.04	.33	.53	-0.30	-0.19
Intersubject Overlap						
Generate CR	.13	.03	.07	.20	0.06	-0.10
Allocate CR	.09	.002	.08	.09	0.81	1.57
Generate FR	.13	.03	.06	.18	-0.23	-0.72
Allocate FR	.09	.002	.08	.09	0.07	-0.47

NOTE: CR = Cued Recall; FR = Free Recall; Cumulative Strength = Cumulative Forward Associative Strength.

Table 7.

Experiment 2 Generate Cued Recall Condition Correlation Matrix.

Variables	1	2	3	4	5	6	7	8	9	10	11
1. WM Z Composite	-										
2. Generate CR PC	.20	-									
3. Generate CR RL	.06	**-.53**	-								
4. Generate CR Intra-list	**-.27**	**-.80**	**.39**	-							
5. Generate CR Extra-list	-.16	**-.94**	**.47**	**.79**	-						
6. Generate CR Omission	-.01	-.01	.17	-.14	**-.25**	-					
7. Generate CR FAS	.01	.19	-.17	-.12	-.11	-.13	-				
8. Generate CR Num Targets	-.11	-.20	-.05	**.43**	**.30**	-.03	**.49**	-			
9. Generate CR C-FAS	-.07	**-.43**	.12	**.49**	**.48**	.04	.11	**.73**	-		
10. Generate CR Database	-.04	**-.35**	.10	**.55**	**.44**	.01	**.25**	**.89**	**.73**	-	
11. Generate CR Overlap	-.03	.06	**-.24**	.15	.05	-.21	**.69**	**.72**	**.31**	**.56**	-

NOTE: Bolded values represent significant correlations at p < .05. WM = Working Memory; WM Z Composite score calculated from participants in the generate cued recall condition (n = 71); CR = Cued Recall; PC = Proportion Correct; RL = Recall Latency; Intra-list = Intra-list Intrusion; Extra-list = Extra-list Intrusion; FAS = Forward Associative Strength; Num Targets = Number of Associated Targets; Database = Proportion of Cues in the Database; Overlap = Intersubject Overlap.

Table 8.

Experiment 2 Allocate Cued Recall Condition Correlation Matrix.

Variables	1	2	3	4	5	6	7	8	9	10	11
1. WM Z Composite	-										
2. Allocate CR PC	.20	-									
3. Allocate CR RL	.02	-.01	-								
4. Allocate CR Intra-list	-.05	**-.31**	**.22**	-							
5. Allocate CR Extra-list	**-.26**	**-.66**	.11	**.49**	-						
6. Allocate CR Omission	.08	**-.34**	-.15	**-.33**	**-.48**	-					
7. Allocate CR FAS	-.01	**.24**	-.08	-.15	**-.24**	.02	-				
8. Allocate CR Num Targets	.13	-.02	-.22	-.10	-.07	.13	**.53**	-			
9. Allocate CR C-FAS	-.01	-.04	-.02	-.03	.07	-.04	.18	**.42**	-		
10. Allocate CR Database	-.13	-.05	.003	-.20	-.04	.12	**.51**	**.63**	**.28**	-	
11. Allocate CR Overlap	.01	-.08	-.01	.10	.07	-.01	-.13	.07	.09	-.11	-

NOTE: Bolded values represent significant correlations at p < .05. WM = Working Memory; WM Z Composite score calculated from participants in the allocate cued recall condition (n = 78); CR = Cued Recall; PC = Proportion Correct; RL = Recall Latency; Intra-list = Intra-list Intrusion; Extra-list = Extra-list Intrusion; FAS = Forward Associative Strength; Num Targets = Number of Associated Targets; Database = Proportion of Cues in the Database; Overlap = Intersubject Overlap.

Table 9.

Experiment 2 Generate Free Recall Condition Correlation Matrix.

Variables	1	2	3	4	5	6	7	8	9
1. WM Z Composite	-								
2. Generate FR PC	.21	-							
3. Generate FR RL	.16	**.32**	-						
4. Generate FR Extra-list	-.06	**-.60**	-.15	-					
5. Generate FR FAS	**.29**	.14	-.16	**-.26**	-				
6. Generate FR Num Targets	**.33**	-.13	-.23	.14	**.46**	-			
7. Generate FR C-FAS	.16	.06	-.03	.16	-.05	**.59**	-		
8. Generate FR Database	.12	-.23	-.16	.24	.22	**.82**	**.55**	-	
9. Generate FR Overlap	**.32**	-.05	**-.27**	-.04	**.67**	**.68**	.16	**.42**	-

NOTE: Bolded values represent significant correlations at p < .05. WM = Working Memory; WM Z Composite score calculated from participants in the generate free recall condition (n = 62); FR = Free Recall; PC = Proportion Correct; RL = Recall Latency; Extra-list = Extra-list Intrusion; FAS = Forward Associative Strength; Num Targets = Number of Associated Targets; Database = Proportion of Cues in the Database; Overlap = Intersubject Overlap.

Table 10.

Experiment 2 Allocate Free Recall Condition Correlation Matrix.

Variables	1	2	3	4	5	6	7	8	9
1. WM Z Composite	-								
2. Allocate FR PC	.22	-							
3. Allocate FR RL	.05	.16	-						
4. Allocate FR Extra-list	-.12	**-.53**	.06	-					
5. Allocate FR FAS	.07	.08	.11	.10	-				
6. Allocate FR Num Targets	-.20	.15	.16	-.06	**.32**	-			
7. Allocate FR C-FAS	**-.28**	.13	.01	-.14	.05	**.48**	-		
8. Allocate FR Database	-.17	.09	.10	-.09	**.45**	**.67**	.22	-	
9. Allocate FR Overlap	-.08	**-.30**	-.06	.16	**.15**	.09	.08	.11	-

NOTE: Bolded values represent significant correlations at p < .05. WM = Working Memory; WM Z Composite score calculated from participants in the allocate free recall condition (n = 71); FR = Free Recall; PC = Proportion Correct; RL = Recall Latency; Extra-list = Extra-list Intrusion; FAS = Forward Associative Strength; Num Targets = Number of Associated Targets; Database = Proportion of Cues in the Database; Overlap = Intersubject Overlap.

Discussion

Consistent with my predictions and Experiment 1, proportion correct was higher in the generated cue condition relative to the allocated cue condition. Additionally, proportion correct was higher for cued recall relative to free recall, and the difference between cued and free recall was larger when cues were generated at encoding relative to allocated at encoding. While generating cues during encoding benefitted both cued and free recall performance, this benefit was larger when those cues were provided back to participants relative to when they were not presented to participants at retrieval. Generating a cue during encoding may strengthen the association in memory between a cue and target word, making it easier to recover the target when presented with a cue at test. However, in free recall, an additional recovery process is required. A participant would need to be able to reinstate and recover a cue word for it to be beneficial in aiding recovery of the target word. Increased attention at encoding involved with active processing (generating rather than reading) may aid in recovery of that cue (consistent with the finding of higher proportion correct in the generate-free condition relative to the allocate-free condition). However, a participant would need to use additional encoded information such as temporal context or related words to reinstate that cue. This process is not explicitly required with cued recall.

Also consistent with predictions, recall latency was lower in the generated cue condition relative to the allocated cue condition. Lower recall latency would suggest that generating a cue at encoding helps participants narrow their search set at retrieval. However, this effect was only observed under free recall conditions, not cued recall conditions. Providing a cue at test seems to narrow the memory search set to a similar degree independent of if the cue was generated or allocated. However, when a participant needs to reinstate the cue themselves, generation seems to facilitate recovery of a more narrow cue/cue set to focus the search set relative to allocation. As mentioned above, increased attention at encoding may facilitate encoding of additional information which would aid in narrowing the search set.

Consistent with Experiment 1, intra-list intrusion errors were lower in the allocated cue condition relative to the generated cue condition. The opposite, but predicted effect, was observed for extra-list intrusion errors and omission errors. These types of errors were lower in the generated cue condition relative to the allocated cue condition. These results would also align with the processes proposed above and in the discussion of Experiment 1. When generating a cue, a participant may reinstate previously studied items in order to not repeat a cue or target word. This additional information may help in reinstating a cue and narrowing the search set at a later test; but, it would result in increased intrusions of incorrect target items. However, as generation strengthens the association between a cue and target (correct and/or incorrect), participants would be less likely to recall non-studied items or no items at all.

WM results were largely consistent with predictions and prior results (see Unsworth, 2019, for review). There is some evidence that suggests enhanced recovery ability (i.e., higher proportion correct); however, no WM effects were observed on omission errors (current study; Unsworth, 2009b). Additionally, high WM individuals made fewer intrusion errors, which suggests enhanced monitoring ability (current study; Unsworth, 2019). However, contrary to predictions and prior results, there was no effect of WM on recall latency. Given this, the current experiment does not support the idea that individuals with high WM are better at focusing their search set during retrieval.

Additionally, contrary to predictions and prior results, across all analyses, there were no interactions between WM and encoding condition, retrieval condition, or both. Neither self-generation of cues at encoding, nor provision of cues during retrieval reduced WM-related differences in performance. Interestingly, WM-related differences were not larger in free recall relative to cued recall as observed in previous studies (e.g., Unsworth, 2009a). It may have been the case that performance for high WM individuals was impaired in free recall due to a mismatch between encoding and retrieval conditions. During encoding, cues are encoded with targets. However, during retrieval, those cues are no longer present. Prior work has suggested that individuals with high WM may be more susceptible to situations involving an encoding/retrieval mismatch such as this (e.g., Unsworth, 2011).

Finally, there were no significant correlations between WM and the number of associated targets, cumulative associative strength, the proportion of cues in the database, or intersubject overlap. While Experiment 1 provided partial evidence that individuals with high WM generate more distinctive and unique cues, this result was not replicated in Experiment 2. In Experiment 1, condition was manipulated within-subject while in Experiment 2, condition was manipulated between-subjects. Within-subject designs minimize inter-subject variability which typically results in higher correlations relative to between-subject designs. Additionally, within-subject designs are better powered to detect a given effect relative to between-subject designs. Given this, the difference in design may have influenced the discrepancy in results.

GENERAL DISCUSSION

The current study replicated the finding that self-generation of cues during encoding benefits long-term memory performance. Additionally, this effect was shown relative to allocation of cues during encoding, rather than during retrieval as the majority of prior studies has done. Manipulating generation versus allocation during encoding allows for the isolation of the self-generation effect, separate from an encoding/retrieval mismatch effect. This self-generation effect was observed in both cued and free recall. The size of this effect differed between Experiments 1 and 2, with a larger difference between the generate-cued and allocate-cued conditions being observed in Experiment 2 relative to the analogous comparison of the generated cue and allocated cue conditions in Experiment 1. This effect size difference could have been driven by multiple factors. For example, in Experiment 1 condition was manipulated within-subject while, in Experiment 2, condition was manipulated between-subjects. Additionally, participants studied fewer words in each condition of Experiment 1 (i.e., 40 words) relative to Experiment 2 (i.e., 120 words). Finally, the cues used in the allocated condition of Experiments 1 and 2 differed. In Experiment 1, the highest semantic associate to a target word (according to the University of South Florida Free Association Norms) was selected as the allocated cue word. In Experiment 2, the allocated cue words were selected from cues generated in the generated cue condition of Experiment 1. This resulted in a large difference between experiments in the average cue-to-target forward associative strength for the allocated cue condition. In Experiment 1, cue-to-target forward associative strength was much higher in the allocated cue condition ($M = .25$, $SD = .001$) relative to the generated cue condition ($M = .06$, $SD = .02$). Whereas in Experiment 2, cue-to-target forward associative strength was lower in the allocated cue condition ($M = .02$, $SD = .01$) relative to the generated cue condition ($M = .06$, SD

= .02), but averages were better equated. Given this, Experiment 2 is likely a better representation of the true effect size.

Analyses of errors and recall latencies also provided novel insights on how self-generation affects memory search processes. In cued recall, cue generation improved proportion correct and reduced omission and extra-list intrusion errors. However, intra-list intrusions errors were higher relative to the allocation condition. For free recall, cue generation improved proportion correct, reduced extra-list intrusion errors, and reduced recall latencies. However, again, intra-list intrusions errors were higher relative to the allocation condition. In relation to the SAM model (Raaijmakers & Shiffrin, 1980) and its proposed retrieval processes, this suggests that cue generation helps narrow the search set in free recall (shorter recall latencies) and improves recovery ability (higher proportion correct, fewer omission errors). However, cue generation had mixed effects on monitoring ability. Participants were better able to evaluate the accuracy of non-studied items that may have come to mind during recall (fewer extra-list intrusions); however, they were worse at evaluating the accuracy of studied items that may have come to mind in response to the incorrect cue during recall (more intra-list intrusions).

In addition to contributions to the self-generation literature, the current study added to the WM/long-term memory literature. Replicating prior work (e.g., Unsworth, 2007, 2009b), individuals with high WM showed higher proportion correct and fewer intra-list and extra-list intrusion errors relative to individuals with low WM in both cued and free recall. Also replicating prior results (Unsworth, 2009b), there was no effect of WM on omission errors in cued recall. However, contrary to prior work (e.g., Unsworth, 2007), there was also no effect of WM on recall latency in Experiment 2. These effects were observed when analyzing WM as a continuous variable rather than using an extreme-groups approach to examine performance differences between high and low WM individuals as much of the literature has done. Extreme group analyses suffer from issues related to reduced power and arbitrary/sample-specific cut-offs between groups. Again, relating these results to the retrieval processes discussed previously, this suggests that individuals with high WM are better at monitoring processes (fewer extra-list and intra-list intrusions). However, there were mixed results on recovery ability (higher proportion correct, no difference in omission errors). Additionally, there was no significant difference in recall latencies between individuals with high and low WM, suggesting no differences in ability to focus the search set.

The current study introduced a novel manipulation of self-generation in conjunction with WM. This allowed for examination of WM-related differences in the effectiveness of generated cues during encoding. There was some evidence that individuals with high WM generated more effective cues in Experiment 1. High WM individuals generated cues that were associated with fewer targets (index of distinctiveness) and that were less likely to be in the database (index of uniqueness). However, there were no WM-related differences in cumulative semantic associative strength, intersubject overlap, or cue-to-target forward associative strength. In Experiment 2, there were no significant effects of WM on cue properties. Given this, the current study did not provide strong evidence that differences in generating effective item-level cues at encoding are driving the WM/long-term memory relationship.

Overall, there were no significant interactions with WM, other than omissions in Experiment 1. WM was more strongly related with omissions in the allocated cue condition and the generated cue condition relative to the generated descriptive cue condition, which was partially aligned with predictions (i.e., reduced WM-related differences in the generated cue conditions). However, none of the correlations between WM and omission errors were significant in any of the conditions. Overall, neither self-generation of cues at encoding (Experiments 1 and 2) nor provision of cues during retrieval (Experiment 2) helped reduce WM-related differences in long-term memory performance.

Limitations and Future Directions

One limitation of the current study was that only cue generation at the item level was examined. There may also be WM-related differences in the generation and utilization of inter-item cues. For example, Spillers & Unsworth (2011) observed greater temporal contiguity for high WM individuals relative to low WM individuals. In other words, they were able to use the products of retrieval to cue additional items. These inter-item cues that form connections between multiple studied items may be particularly important for tasks such as free recall, where organizational cues that can cue multiple items are important in conjunction with distinctive cues for a specific item. Additionally, in the current study, participants were only able to generate a single word cue. However, they may have been generating additional cues such as inter-item cues which they did not directly output. Future research allowing participants to generate multiple cues may be able to better address this concern. Finally, it is possible that encoding of certain cues (e.g., temporal-contextual cues) occurs automatically and outside of conscious awareness. The current study was not designed to examine generation of these cues at encoding. Future work manipulating encoding strategies to focus on certain types of cues may be able to provide a better understanding of the impact of these cues on retrieval.

Another limitation of the current study was the database used for examination of cue properties. The University of South Florida Free Association Norms (Nelson et al., 1998) was selected due to its use in previous research (Tullis & Benjamin, 2015; Tullis & Finley, 2021) to examine the cue properties analyzed in the current study. However, these norms were established from data collected over 50 years ago (Nelson & Brooks, 1974). Over time connotation, frequency, or associations between words may change. Future research examining properties of cues may benefit from using a more recent dataset. In Experiment 1 of the current study, despite evidence that cues in the generated descriptive cue condition were more distinctive and unique relative to cues in the generated cue condition, performance was typically better in the generated cue condition (higher proportion correct, fewer intra-list intrusions, and fewer omissions). Future research is needed to determine if this effect is replicable. The finding that cues were more distinctive and unique could have been influenced by the fact that fewer words in the generated descriptive cue condition were present in the association norm database. Additionally, while checks were performed to determine if generated cues were valid (i.e., a single word that was different than the target word), exclusions were not made based on if the generated cue was indeed a descriptor for the target word.

Regarding the WM tasks, although the score distribution and correlation between reading span and color change detection was largely consistent with prior work (e.g., see Robison et al., 2024 for a recent large-scale, multi-university study), the correlation was still relatively low, particularly for Experiment 2 (Experiment 1: $r = .29$; Experiment 2: $r = .18$). This is not ideal for using these measures as a composite score. To address this limitation, results using only reading span score as the covariate in the various ANCOVA analyses reported above are included in Appendices A and B, as complex span tasks have been more often used in prior literature. Although both reading span and color change detection tasks have been used extensively as measures of WM, they may tap into different components of WM. For example, color change detection has been proposed to measure primary memory ability, or the ability to hold information in an active state (Robison et al., 2024). This task relies less on an explicit retrieval component, as recall occurs immediately after a brief presentation of stimuli, and information is assumed to never leave an active state. It is therefore notable that correlations between color change detection and recall performance were significant despite color change detection and cued/free recall featuring very different task demands (i.e., visual vs. verbal stimuli, different presentation duration, immediate vs. delayed recall). Unsworth (2016) provided evidence that, in order to fully statistically account for the relationship between WM and long-term memory, both encoding processes (i.e., strategy use) and retrieval processes (i.e., search efficiency and monitoring ability) needed to be accounted for. Future work investigating the relationship between WM and long-term memory may benefit from using multiple different working memory tasks to minimize task-specific variance. Additionally, multifaceted accounts of WM propose that WM may involve many different components including attention control, primary memory, and secondary memory (Robison et al., 2024; Unsworth et al., 2014). Certain tasks may reflect specific components to differing degrees. Factor analysis approaches to research may be beneficial in accounting for these different components while reducing some of the issues with *z*-score composites between tasks with small correlations.

Finally, another limitation of the current study is power. Given that prior research had not examined the particular interactions that the current study was primarily interested in (i.e., the interaction between condition and WM), there was not a relevant effect size to base the reported power analyses on. Instead, I opted to use a small-to-medium effect size estimate for the power analysis (i.e., η_p^2 = .04). Additionally, this power analysis was not adjusted for the multiple dependent variables being examined in the current study. For Experiment 2 in particular, condition was manipulated between-subjects resulting in ANCOVAs and correlations based on subsamples ranging from n = 62 to n = 78, which is not ideal for individual differences research.

However, the overall sample size was still much larger than prior research on the topic and allowed for a feasible, large-scale examination of individual differences in WM and long-term memory. Additionally, there was sufficient power in the current study to reliably detect significant effects of $\eta_p^2 = 0.01$.

Future research may also be interested in investigating the effects of retention interval. In the current study, performance was only assessed after a short delay (~15-20 minutes). It is possible that performance differences between the generated cue condition and the allocated cue condition may change with a longer delay. For example, Mäntylä (1986) observed a sharper decrease in performance with retention interval when participants generated cues at encoding and were given back those cues at retrieval, relative to when they were provided with another participants' cues at retrieval. Mäntylä proposed that, in addition to distinctiveness, the effectiveness of a generated cue depends on context. With larger changes in context, self-generation may not be as beneficial. It is possible that WM may interact with this effect as well. Previous research has suggested that high WM individuals are better at reinstating and using effective temporal contextual cues during recall relative to low WM individuals (e.g., Spillers & Unsworth, 2011; Unsworth & Engle, 2006). There may be individual differences in how performance is affected by larger changes in context.

Conclusion

The current study presented novel examinations of the self-generation effect and suggested that self-generation may improve recovery of a target word in cued and free recall, as well as narrowing of the search set in free recall. However, its effect on monitoring ability is mixed. Monitoring for incorrect, non-studied items was improved. However monitoring for incorrect, studied items was worse relative to an allocated cue condition

To my knowledge, the current study is the first to examine WM-related differences in long-term memory performance following self-generation of cues at encoding. This research has practical implications. WM ability has been shown to be predictive of various measures of academic success, such as reading comprehension, SAT scores, and fluid intelligence (for review, see Unsworth & Redick, 2017). Given this, it is important to be able to identify strategies that help reduce the gap in learning for students at varying levels of WM ability. However, in the current study, self-generation of cues during encoding failed to reduce WM-related differences. Additional research is needed to better understand the processes that underlie the relationship between WM and long-term memory. Previous research has proposed a number of mechanisms that may drive this relationship: encoding strategy, intensity of attention at encoding, encoding of relevant/irrelevant information, ability to narrow the search set, recovery ability, and monitoring ability. The current study provides the strongest evidence for a difference in monitoring ability given that participants with high WM made fewer intrusion errors relative to participants with low WM. Additionally, there was partial evidence for a difference in recovery ability (higher proportion correct). However, the current study did not seem to support the mechanisms of intensity of attention at encoding or narrowing of the search set. When attention to each item was facilitated with the cue generation task, WM-related differences were not reduced. Additionally, there were no WM-related differences in recall latency. Encoding strategy may still be a potential mechanism driving the WM/long-term memory relationship; however, it does not seem that generating item-level cues is the underlying strategy. Future work examining the creation and use of inter-item level cues in addition to effective item-level cues may help improve our theoretical understanding of individual differences in long-term memory.

APPENDIX A. EXPERIMENT 1 SUPPLEMENTAL

Table A1

Experiment 1 Cue Intrusions

Variables	Mean	SD	Min	Max	Skew	Kurtosis
Cue Intrusions						
Generate-Cue	.05	.05	.00	.28	1.60	3.22
Generate-Describe	.05	.05	.00	.30	1.63	3.63
Allocate	.04	.04	.00	.23	1.33	1.98
Intra-experimental Intrusions						
Generate-Cue	.10	.08	.00	.38	0.92	0.52
Generate-Describe	.10	.08	.00	.43	1.17	1.29
Allocate	.07	.06	.00	.30	1.11	1.06
Extra-experimental Intrusions						
Generate-Cue	.15	.11	.00	.53	0.99	0.61
Generate-Describe	.15	.12	.00	.55	1.08	0.75
Allocate	.23	.15	.00	.65	0.36	-0.48

NOTE: Intra-experimental Intrusions = Targets and Cues; Extra-experimental Intrusions = Non-targets and Non-cues.

Table A2

Experiment 1 Cue Properties - Tullis & Benjamin Calculation

Variables	Mean	SD	Min	Max	Skew	Kurtosis
Number of Associated Targets						
Generate-Cue	10.56	2.19	2.45	15.55	-0.56	0.36
Generate-Describe	10.27	2.26	2.75	15.40	-0.63	0.48
Allocate	11.70	0.36	11.23	12.13	-0.17	-1.46
Cumulative Strength						
Generate-Cue	0.56	0.10	0.14	0.77	-0.76	0.69
Generate-Describe	0.55	0.11	0.18	0.76	-0.87	0.77
Allocate	0.55	0.02	0.53	0.57	0.25	-1.52

NOTE: Cumulative Strength = Cumulative Forward Associative Strength.

Cue Intrusions

In addition to intrusions of target items, intrusions of studied or generated cues were examined due to their presence as studied experimental stimuli. These intrusions are labeled below as cue intrusions. Intra-list and extra-list intrusions were also examined with a modified calculation. Intra-list intrusions (or intra-experimental intrusions as labeled in this section) were defined as all studied items, including both cue and target words. Extra-list intrusions (or extra-experimental intrusions as labeled in this section) were defined as outputted responses that were not studied targets or cues.

A significant difference in the proportion of cue intrusions was observed between conditions. This was confirmed with a repeated-measures ANOVA, $F(2, 564) = 5.90$, $MSE = .001$, $p = .003$, $\eta_p^2 = .02$. A significantly lower proportion of cue intrusions occurred in the allocated cue condition relative to the generated cue condition, $t(281) = -3.00$, $p = .002$, and the generated descriptive cue condition, $t(281) = -3.00$, $p = .003$. There was not a significant difference in cue intrusions between the generated cue condition and the generated descriptive cue condition, $t(281) = 0.03$, $p = .976$.

A significant difference in the proportion of intra-experimental intrusions was also observed between conditions. This was confirmed with a repeated-measures ANOVA, $F(2, 564) = 32.10$, $MSE = .003$, $p < .001$, $\eta_p^2 = .10$. A significantly lower proportion of intra-experimental intrusions occurred in the allocated cue condition relative to the generated cue condition, $t(281) = -7.25$, $p < .001$, and the generated descriptive cue condition, $t(281) = -7.75$, $p < .001$. There was not a significant difference in intra-experimental intrusions between the generated cue condition and the generated descriptive cue condition, $t(281) = -0.50$, $p = .685$.

A significant difference in the proportion of extra-experimental intrusions was also observed between conditions. This was confirmed with a repeated-measures ANOVA, $F(2, 564) = 97.00$, $MSE = .007$, $p < .001$, $\eta_p^2 = .26$. A significantly lower proportion of extra-experimental intrusions occurred in the generated cue condition relative to the allocated cue condition, $t(281) = 12.00$, $p < .001$. Additionally, a significantly lower proportion of extra-experimental intrusions occurred in the generated descriptive cue condition relative to the allocated cue condition, $t(281) = -10.00$, $p < .001$. There was not a significant difference in extra-experimental intrusions between the generated cue condition and the generated descriptive cue condition, $t(281) = -1.00$, $p = .366$.

Cue Intrusions and Working Memory

An ANCOVA with condition (generated cue, generated descriptive cue, and allocated cue) as a within-subjects variable and WM composite *z*-score as a continuous covariate was conducted. Proportion of cue intrusions was the dependent variable. Mirroring the ANOVA results above, there was a main effect of condition, $F(2, 562) = 5.94$, $MSE = .001$, $p = .003$,

η_p^2= .02. The main effect of WM was also significant, $F(1, 281) = 37.54$, $MSE = .004$, $p < .001$, η_p^2= .12, indicating that individuals with higher WM scores made fewer cue intrusion errors across conditions. However, the critical interaction between condition and WM was not significant, $F(2, 562) = 2.79$, $MSE = .001$, $p = .062$, $\eta_p^2 < .01$.

An ANCOVA with condition (generated cue, generated descriptive cue, and allocated cue) as a within-subjects variable and WM composite z-score as a continuous covariate was conducted to analyze the proportion of intra-experimental intrusions. Mirroring the ANOVA results above, there was a main effect of condition, $F(2, 562) = 32.25$, $MSE = .003$, $p < .001$, η_p^2= .10. The main effect of WM was also significant, $F(1, 281) = 35.87$, $MSE = .009$, $p < .001$, η_p^2= .12, indicating that individuals with higher WM scores made fewer intra-experimental intrusion errors across conditions. However, the critical interaction between condition and WM was not significant, $F(2, 562) = 2.35$, $MSE = .003$, $p = .096$, $\eta_p^2 < .01$.

An ANCOVA with condition (generated cue, generated descriptive cue, and allocated cue) as a within-subjects variable and WM composite z-score as a continuous covariate was conducted. Proportion of extra-experimental intrusions was the dependent variable. Mirroring the ANOVA results above, there was a main effect of condition, $F(2, 562) = 97.40$, $MSE = .007$, $p < .001$, η_p^2= .26. The main effect of WM was also significant, $F(1, 281) = 30.89$, $MSE = .032$, $p < .001$, η_p^2= .10, indicating that individuals with higher WM scores made fewer extra-experimental intrusion errors across conditions. However, the critical interaction between condition and WM was not significant, $F(2, 562) = 2.17$, $MSE = .007$, $p = .116$, $\eta_p^2 < .01$.

Cue Properties Tullis and Benjamin (2015) Calculation

A significant difference in the number of associated targets (Tullis & Benjamin calculation) was observed between conditions. This was confirmed with a paired-samples t-test, *t*(282) = 2.45, *p* = .015, *d* = 0.15 [.03, .26]. Cues were associated with fewer targets in the generated descriptive cue condition (M = 10.27, SD = 2.26) relative to the generated cue condition (M = 10.56, SD = 2.19).

There was not a significant difference observed between conditions in the cumulative associative strength (Tullis & Benjamin calculation) between a cue and other targets. This was confirmed with a paired-samples *t*-test, *t*(282) = 1.66, *p* = .099, *d* = 0.10 [-.02, .22].

Cue Properties & Working Memory Tullis & Benjamin (2015) Calculation

I conducted an ANCOVA with condition (generated cue and generated descriptive cue) as a within-subjects variable and WM composite *z*-score as a continuous covariate. The number of associated targets was the dependent variable. Mirroring the *t*-test results above, there was a main effect of condition, *F*(2, 281) = 5.96, *MSE* = 1.99, *p* = .015, η_p^2= .02. The main effect of WM was also significant, *F*(1, 281) = 22.51, *MSE* = 7.39, *p* < .001, η_p^2= .07, indicating that participants with higher WM generated cues that were related to fewer targets. The interaction between condition and WM was not significant, *F*(2, 281) = 0.06, *MSE* = 1.99, *p* = .804, η_p^2< .01.

An ANCOVA with condition (generated cue and generated descriptive cue) as a within-subjects variable and WM composite *z*-score as a continuous covariate was conducted to examine the cumulative forward associative strength of a cue to other targets. Mirroring the *t*-test results above, there was not a significant main effect of condition, *F*(2, 281) = 2.74, *MSE* < .01, *p* = .099, η_p^2= .01. However, the main effect of WM was significant, *F*(1, 281) = 19.91, *MSE* = .018, *p* < .001, η_p^2= .07, indicating that individuals with high WM generated cues with lower cumulative associative strength. Finally, the the interaction between condition and WM was not significant, *F*(2, 281) = 0.78, *MSE* < .01, *p* = .378, η_p^2< .01.

Cue Properties with Allocate Condition

A significant difference in the cue-to-target forward associative strength was observed between conditions. This was confirmed with a repeated-measures ANOVA, $F(2, 564)$ = 10466.51, $MSE < .001$, $p < .001$, $\eta_p^2 = .97$. A significantly higher associative strength occurred in the allocated cue condition relative to the generated cue condition, $t(281) = 186.00$, $p < .001$, and the generated descriptive cue condition, $t(281) = 206.00$, $p < .001$. Additionally, a significantly higher associative strength occurred in the generated cue condition relative to the generated descriptive cue condition, $t(281) = 10.00$, $p < .001$.

A significant difference in the number of associated targets was observed between conditions. This was confirmed with a repeated-measures ANOVA, $F(2, 564) = 1473.13$, $MSE = .026$, $p < .001$, $\eta_p^2 = .84$. A significantly lower number of associated targets was observed in the generated cue condition relative to the allocated cue condition, $t(281) = -47.50$, $p < .001$. A significantly lower number of associated targets was also observed in the generated descriptive cue condition relative to the generated cue condition, $t(281) = -8.43$, $p < .001$, and the allocated cue condition, $t(281) = -49.14$, $p < .001$.

A significant difference in the cumulative forward associative strength of a cue to other targets was observed between conditions. This was confirmed with a repeated-measures ANOVA, $F(2, 564) = 86.08$, $MSE < .001$, $p < .001$, $\eta_p^2 = .23$. A significantly lower cumulative associative strength was observed in the allocated cue condition relative to the generated cue condition, $t(281) = 10.00$, $p < .001$, and the generated descriptive cue condition, $t(281) = 8.00$, $p < .001$. A significantly lower cumulative associative strength was also observed in the generated descriptive cue condition relative to the generated cue condition, $t(281) = 2.00$, $p = .047$.

Cue Properties with Allocate Condition (Tullis & Benjamin (2015) Calculation)

A significant difference in the number of associated targets was observed between conditions. This was confirmed with a repeated-measures ANOVA, $F(2, 564) = 68.71$, $MSE = 2.36$, $p < .001$, $\eta_p^2 = .20$. A significantly lower number of associated targets was observed in the generated cue condition relative to the allocated cue condition, $t(281) = -8.64$, $p < .001$. A significantly lower number of associated targets was also observed in the generated descriptive cue condition relative to the generated cue condition, $t(281) = -2.45$, $p = .015$, and the allocated cue condition, $t(281) = -10.51$, $p < .001$.

There were no significant differences observed between conditions in the cumulative forward associative strength of a cue to other targets. This was confirmed with a repeated-measures ANOVA, $F(2, 564) = 1.41$, $MSE = .01$, $p = .245$, $\eta_p^2 < .01$.

Cue Properties and Working Memory with Allocate Condition

Next I analyzed WM-related differences in the properties of cues for all conditions. I conducted an ANCOVA with condition (generated cue, generated descriptive cue, and allocated cue) as a within-subjects variable and WM composite *z*-score as a continuous covariate. Cue-to-target forward associative strength was the dependent variable. Mirroring the ANOVA results above, there was a main effect of condition, $F(2, 562) = 10517.43$, $MSE < .001$, $p < .001$, $\eta_p^2 = .97$. The main effect of WM was not significant, $F(1, 281) < 0.001$, $MSE < .001$, $p = .986$, $\eta_p^2 < .01$. Additionally, the critical interaction between condition and WM was not significant, $F(2, 562) = 2.37$, $MSE < .001$, $p = .094$, $\eta_p^2 < .01$.

I conducted an ANCOVA with condition (generated cue, generated descriptive cue, and allocated cue) as a within-subjects variable and WM composite *z*-score as a continuous covariate. The number of associated targets was the dependent variable. Mirroring the ANOVA results above, there was a main effect of condition, $F(2, 562) = 1484.80$, $MSE = .026$, $p < .001$, $\eta_p^2 = .84$.

The main effect of WM was also significant, $F(1, 281) = 10.76$, $MSE = .033$, $p = .001$, $\eta_p^2 = .04$, indicating that participants with higher WM generated cues that were related to fewer targets. Additionally, the critical interaction between condition and WM was significant, $F(2, 562) = 3.23$, $MSE = .026$, $p = .040$, $\eta_p^2 = .01$. WM-related differences in the number of associated targets were observed in the generated cue and generated descriptive cue conditions, but not the allocated cue condition.

I conducted an ANCOVA with condition (generated cue, generated descriptive cue, and allocated cue) as a within-subjects variable and WM composite *z*-score as a continuous covariate. The cumulative forward associative strength of a cue to other targets was the dependent variable. Mirroring the ANOVA results above, there was a main effect of condition, $F(2, 562) = 86.21$, $MSE < .001$, $p < .001$, $\eta_p^2 = .24$. However, the main effect of WM was not significant, $F(1, 281) = 1.48$, $MSE < .001$, $p = .226$, $\eta_p^2 < .01$, nor was the interaction between condition and WM, $F(2, 562) = 1.42$, $MSE < .001$, $p = .242$, $\eta_p^2 < .01$.

Cue Properties and Working Memory with Allocate Condition (Tullis and Benjamin (2015) Calculation)

An ANCOVA with condition (generated cue, generated descriptive cue, and allocated cue) as a within-subjects variable and WM composite *z*-score as a continuous covariate was conducted to analyze the number of associated targets to a cue. Mirroring the ANOVA results above, there was a main effect of condition, $F(2, 562) = 71.45$, $MSE = 2.26$, $p < .001$, $\eta_p^2 = .20$. The main effect of WM was also significant, $F(1, 281) = 22.26$, $MSE = 4.98$, $p < .001$, $\eta_p^2 = .07$, indicating that participants with higher WM generated cues that were related to fewer targets. Additionally, the critical interaction between condition and WM was significant, $F(2, 562) = 12.27$, $MSE = 2.26$, $p < .001$, $\eta_p^2 = .04$. WM-related differences in the number of associated targets were observed in the generated cue and generated descriptive cue conditions, but not the allocated cue condition.

An ANCOVA with condition (generated cue, generated descriptive cue, and allocated cue) as a within-subjects variable and WM composite *z*-score as a continuous covariate was conducted. The cumulative forward associative strength of a cue to other targets was the dependent variable. Mirroring the ANOVA results above, there was not a significant main effect of condition, $F(2, 562) = 1.47$, $MSE = .01$, $p = .232$, $\eta_p^2 < .01$. However, the main effect of WM was significant, $F(1, 281) = 19.30$, $MSE = .01$, $p < .001$, $\eta_p^2 = .06$, indicating that participants with higher WM generated cues with lower cumulative associative strength. Additionally, the interaction between condition and WM was significant, $F(2, 562) = 12.01$, $MSE = .01$, $p < .001$, $\eta_p^2 = .04$. WM-related differences in the number of associated targets were observed in the generated cue and generated descriptive cue conditions, but not the allocated cue condition.

Analyses with only Reading Span

Proportion Correct

An ANCOVA with condition (generated cue, generated descriptive cue, and allocated cue) as a within-subjects variable and Reading Span score as a continuous covariate was conducted with proportion correct as the dependent variable. The main effect of condition was not significant, $F(2, 562) = 1.93$, $MSE = .01$, $p = .146$, $\eta_p^2 < .01$. However, the main effect of WM was significant, $F(1, 281) = 26.63$, $MSE = .06$, $p < .001$, $\eta_p^2 = .09$, indicating that individuals with higher WM scores performed better across conditions. The critical interaction between condition and WM was not significant, $F(2, 562) = 1.97$, $MSE = .01$, $p = .141$, $\eta_p^2 < .01$.

Recall Latency

An ANCOVA with condition (generated cue, generated descriptive cue, and allocated cue) as a within-subjects variable and Reading Span score as a continuous covariate was conducted. The main effect of condition was not significant, $F(2, 562) = 2.37$, $MSE = 333802.44$, $p = .094$, $\eta_p^2 < .01$. However, the main effect of WM was significant, $F(1, 281) = 28.50$, $MSE = 2546973.88$, $p < .001$, $\eta_p^2 = .09$, indicating that individuals with higher WM scores recalled words faster across conditions. The critical interaction between condition and WM was not significant, $F(2, 562) = 1.80$, $MSE = 333802.44$, $p = .166$, $\eta_p^2 < .01$.

Recall Errors

Next, I analyzed the types of errors made by participants. I conducted an ANCOVA with condition (generated cue, generated descriptive cue, and allocated cue) as a within-subjects variable and Reading Span score as a continuous covariate. Proportion of intra-list intrusions was the dependent variable. There was a significant main effect of condition, $F(2, 562) = 3.79$, $MSE = .001$, $p = .023$, $\eta_p^2 = .01$. The main effect of WM was also significant, $F(1, 281) = 11.44$, $MSE = .002$, $p < .001$, $\eta_p^2 = .04$, indicating that individuals with higher WM scores made fewer intra-list intrusion errors across conditions. However, the critical interaction between condition and WM was not significant, $F(2, 562) = 0.77$, $MSE = .001$, $p = .464$, $\eta_p^2 < .01$.

An ANCOVA with condition (generated cue, generated descriptive cue, and allocated cue) as a within-subjects variable and Reading Span score as a continuous covariate was conducted to analyze the proportion of extra-list intrusions. There was a significant main effect of condition, $F(2, 562) = 4.14$, $MSE = .01$, $p = .016$, $\eta_p^2 = .02$. The main effect of WM was also significant, $F(1, 281) = 19.72$, $MSE = .05$, $p < .001$, $\eta_p^2 = .07$, indicating that individuals with higher WM scores made fewer extra-list intrusion errors across conditions. However, the critical interaction between condition and WM was not significant, $F(2, 562) = 0.77$, $MSE = .008$, $p = .461$, $\eta_p^2 < .01$.

Finally, I conducted an ANCOVA with condition (generated cue, generated descriptive cue, and allocated cue) as a within-subjects variable and Reading Span score as a continuous covariate. Proportion of omission errors was the dependent variable. The main effect of condition was not significant, $F(2, 562) = 2.23$, $MSE = .001$, $p = .108$, $\eta_p^2 < .01$. Additionally, the main effect of WM was not significant, $F(1, 281) = 2.59$, $MSE = .008$, $p = .109$, $\eta_p^2 < .01$. The interaction between condition and WM was also not significant, $F(2, 562) = 1.84$, $MSE = .001$, $p = .159$, $\eta_p^2 < .01$.

Cue Properties

An ANCOVA with condition (generated cue and generated descriptive cue) as a within-subjects variable and Reading Span score as a continuous covariate was conducted to analyze cue-to-target forward associative strength. The main effect of condition was not significant, $F(2, 281) = 0.01$, $MSE < .001$, $p = .934$, $\eta_p^2 < .01$, nor was the main effect of WM, $F(1, 281) = 0.14$, $MSE = .001$, $p = .708$, $\eta_p^2 < .01$. The interaction between condition and WM was significant, $F(2, 281) = 4.82$, $MSE < .001$, $p = .029$, $\eta_p^2 = .02$. The interaction was likely driven by nonsignificant correlations in the opposite direction from each other. Although neither of the correlations were significant, individuals with higher WM generated cues with slightly lower forward associative strength in the generated descriptive cue condition relative to individuals with low WM (r = -.07, $p = .266$). However, individuals with higher WM generated cues with slightly higher forward associative strength in the generated cue condition relative to individuals with low WM (r = .10, $p = .083$).

An ANCOVA with condition (generated cue and generated descriptive cue) as a within-subjects variable and Reading Span score as a continuous covariate was conducted. The number of associated targets was the dependent variable. The main effect of condition was not significant, $F(2, 281) = 0.19$, $MSE = .03$, $p = .663$, $\eta_p^2 < .01$. However, the main effect of WM was significant, $F(1, 281) = 8.64$, $MSE = .052$, $p = .004$, $\eta_p^2 = .03$, indicating that participants with higher WM generated cues that were related to fewer targets. The interaction between condition and WM was not significant, $F(2, 281) = 1.27$, $MSE = .030$, $p = .260$, $\eta_p^2 < .01$.

An ANCOVA with condition (generated cue and generated descriptive cue) as a within-subjects variable and WM composite *z*-score as a continuous covariate was conducted to analyze the cumulative forward associative strength of a cue to other targets. The main effect of condition was not significant, $F(2, 281) = 0.18$, $MSE < .001$, $p = .673$, $\eta_p^2 < .01$. However, the main effect of WM was significant, $F(1, 281) = 4.24$, $MSE < .001$, $p = .040$, $\eta_p^2 = .02$, indicating that participants with higher WM generated cues with lower cumulative forward associative strength. The interaction between condition and WM was not significant, $F(2, 281) = 0.66$, $MSE < .001$, $p = .417$, $\eta_p^2 < .01$.

Next, to measure cue uniqueness, I analyzed the proportion of generated cues in the database and the intersubject overlap. I conducted an ANCOVA with condition (generated cue and generated descriptive cue) as a within-subjects variable and Reading Span score as a continuous covariate. The proportion of cues in the database was the dependent variable. The main effect of condition was not significant, $F(2, 281) = 2.31$, $MSE = .01$, $p = .130$, $\eta_p^2 < .01$. However, the main effect of WM was significant, $F(1, 281) = 17.99$, $MSE = .03$, $p < .001$, $\eta_p^2 = .06$, indicating that participants with higher WM produced fewer cues that were present in the database. The interaction between condition and WM was also significant, $F(2, 281) = 5.20$, *MSE* = .01, *p* = .023, η_p^2= .02. The correlation between WM and proportion of cues in the database was numerically higher in the generated descriptive cue condition (r = -.27, *p* < .001) relative to the generated cue condition (r = -.17, *p* = .005).

Finally, I conducted an ANCOVA with condition (generated cue and generated descriptive cue) as a within-subjects variable and Reading Span score as a continuous covariate to analyze the intersubject overlap. There was a main effect of condition, *F*(2, 281) = 4.45, *MSE* = .001, *p* = .036, η_p^2= .02. However, the main effect of WM was not significant, *F*(1, 281) = 1.75, *MSE* = .002, *p* = .187, η_p^2< .01, nor was the interaction between condition and WM, *F*(2, 281) = 3.18, *MSE* = .001, *p* = .076, η_p^2= .01.

Figures

In all appendix figures error bars represent ± 1 standard error of the mean. In figures A8-A14, WM is dichotomized into high and low groups for visualization. The high WM group is composed of participants with WM composite *z*-scores above the 50^{th} percentile. The low WM group is composed of participants with WM composite *z*-scores equal to or below the 50^{th} percentile.

Figure A1.

Experiment 1 Proportion of Intra-list Intrusions by Condition.

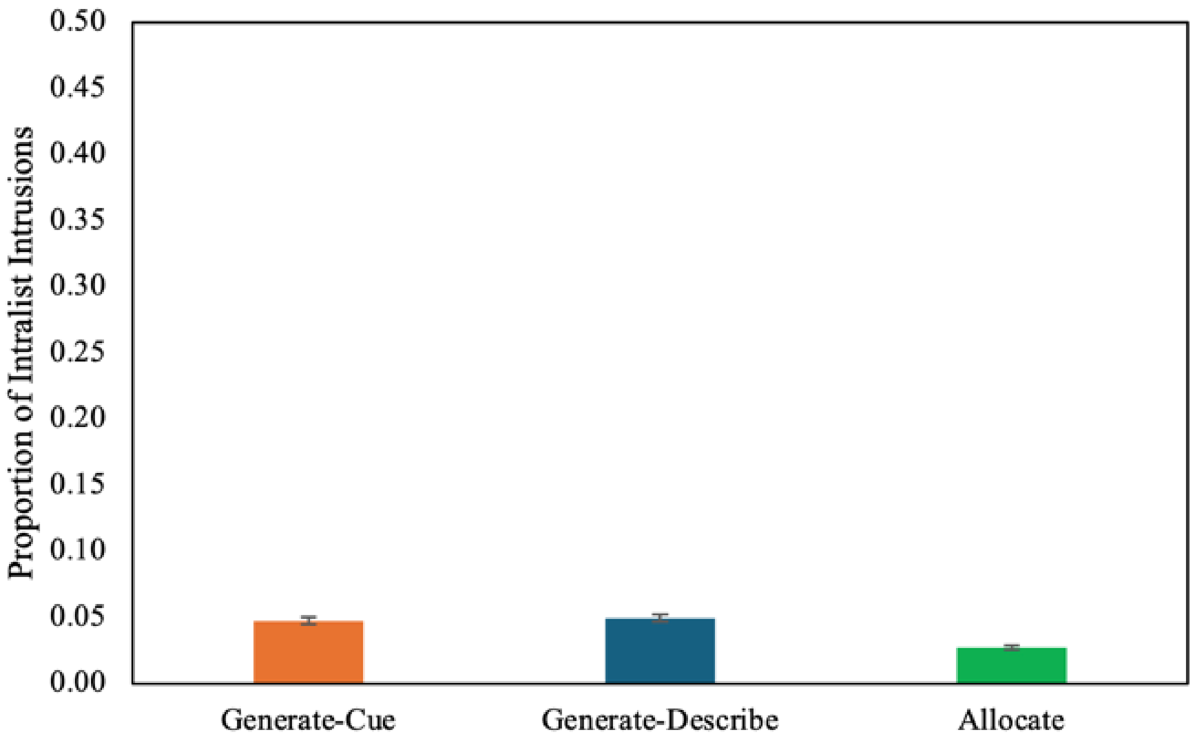


Figure A2.

Experiment 1 Proportion of Extra-list Intrusions by Condition.

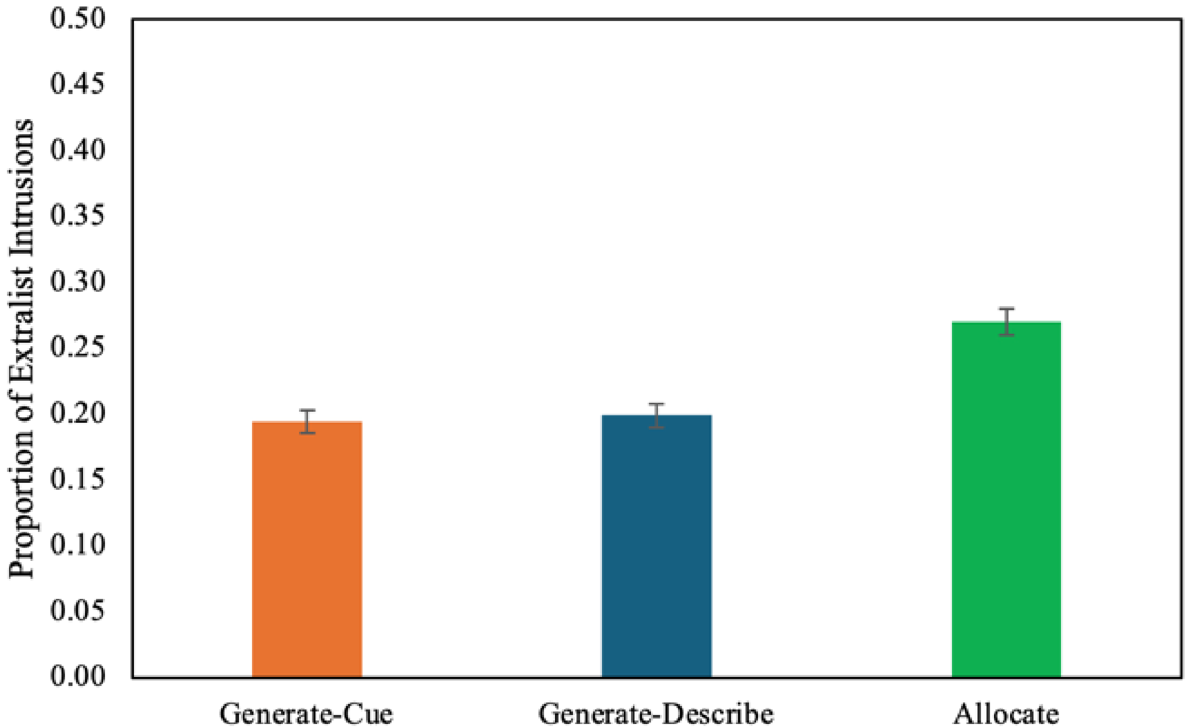


Figure A3.

Experiment 1 Proportion of Omission Errors by Condition.

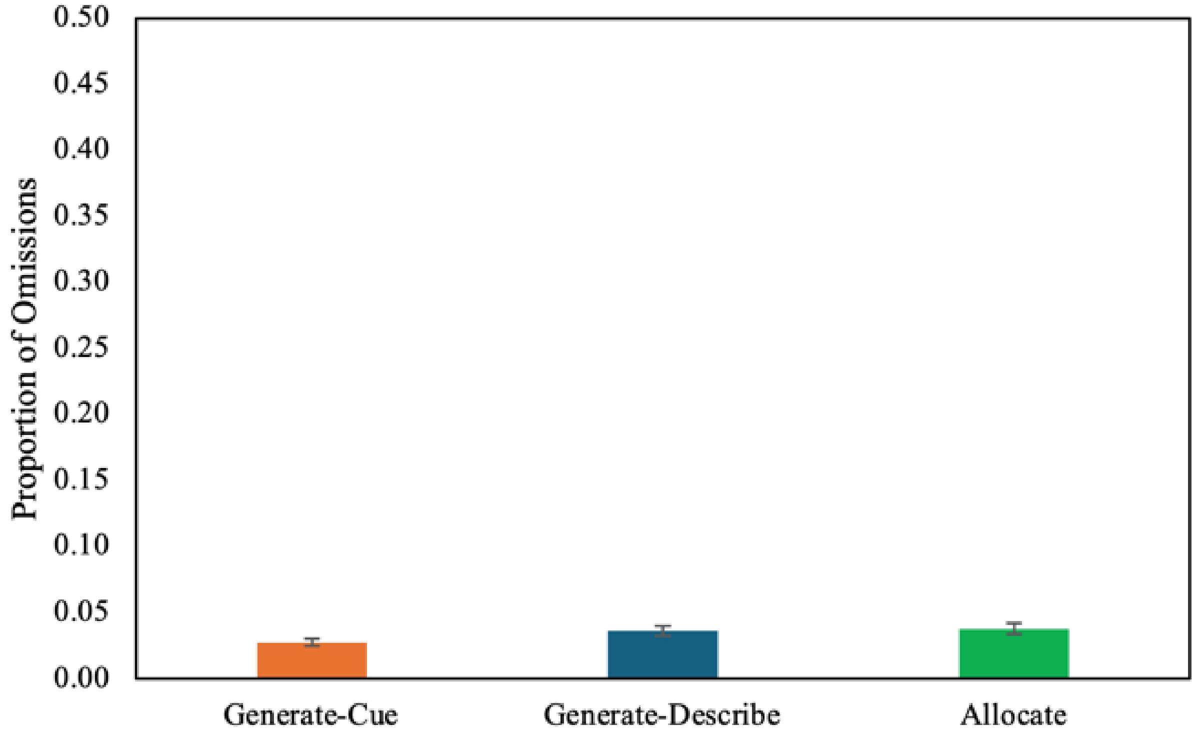


Figure A4.

Experiment 1 Mean Number of Associated Targets by Condition.

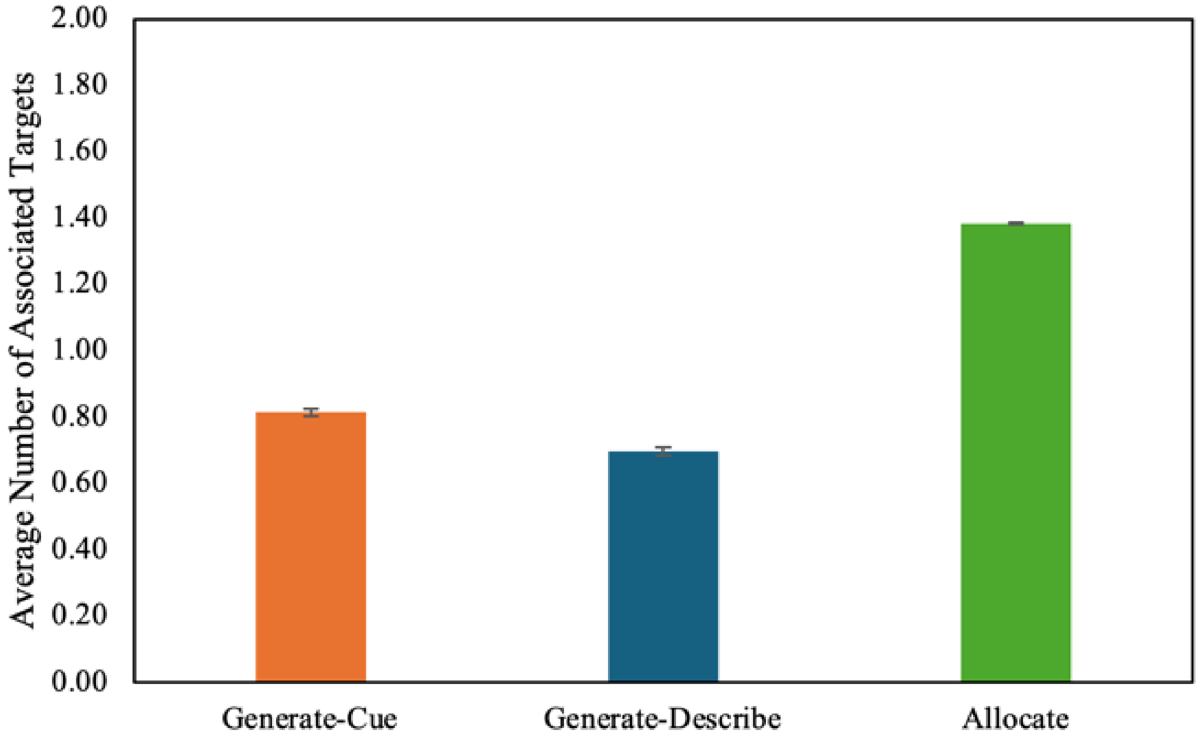


Figure A5.

Experiment 1 Mean Cumulative Associative Strength by Condition.

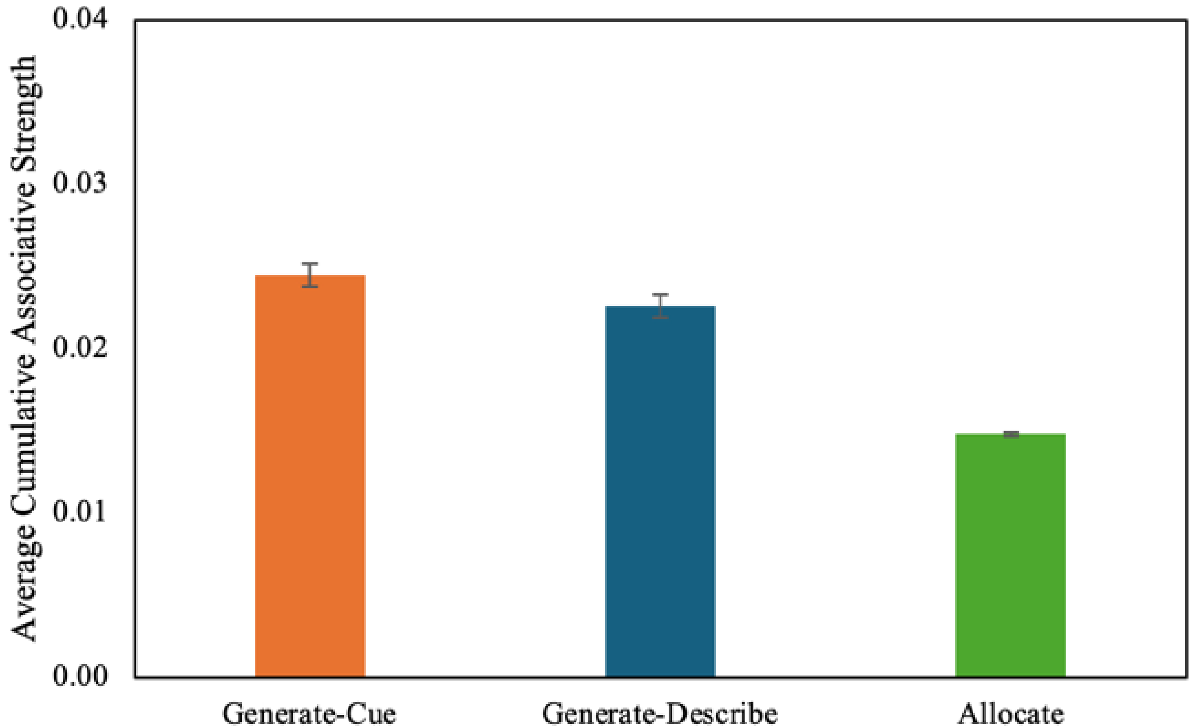


Figure A6.

Experiment 1 Proportion of Cues in the Database by Condition.

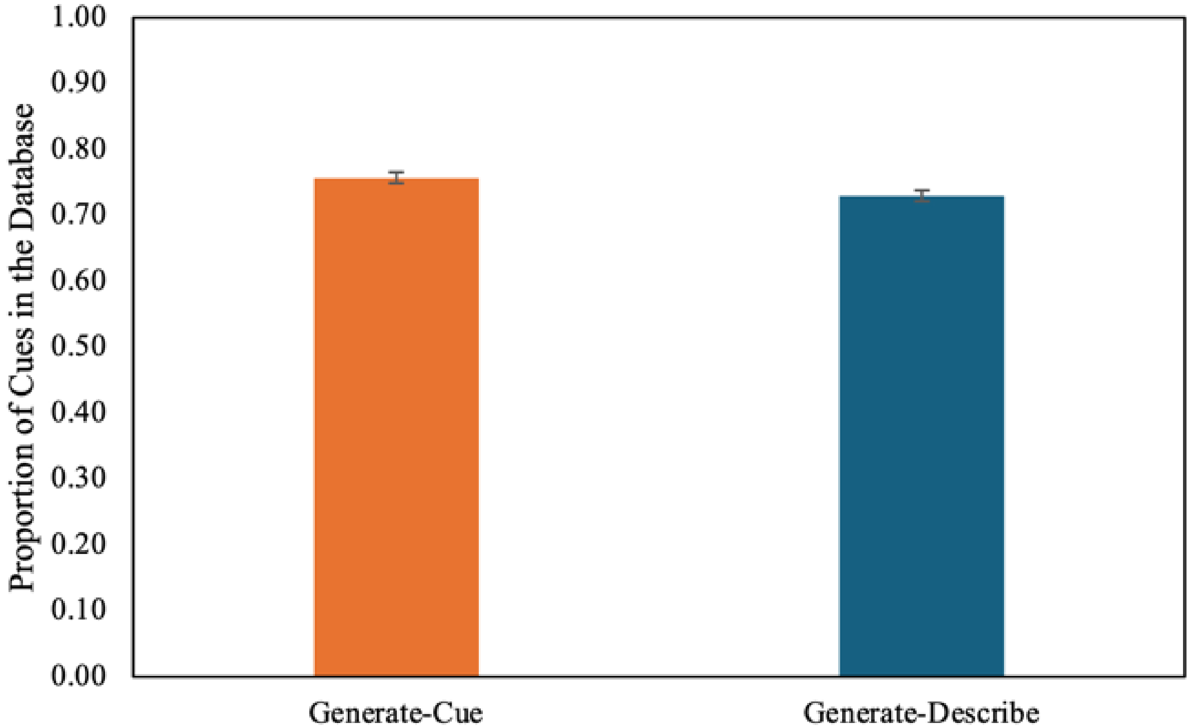


Figure A7.

Experiment 1 Intersubject Overlap by Condition.

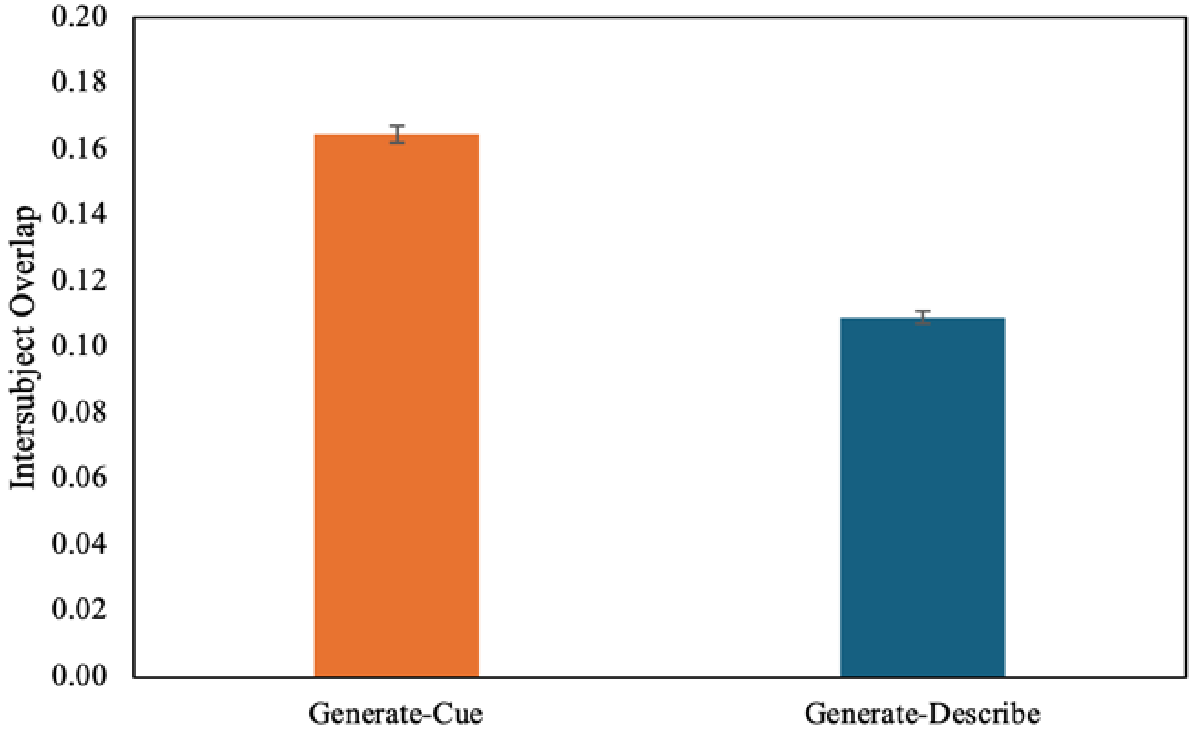


Figure A8.

Experiment 1 Proportion of Intra-list Intrusions by Condition and WM.

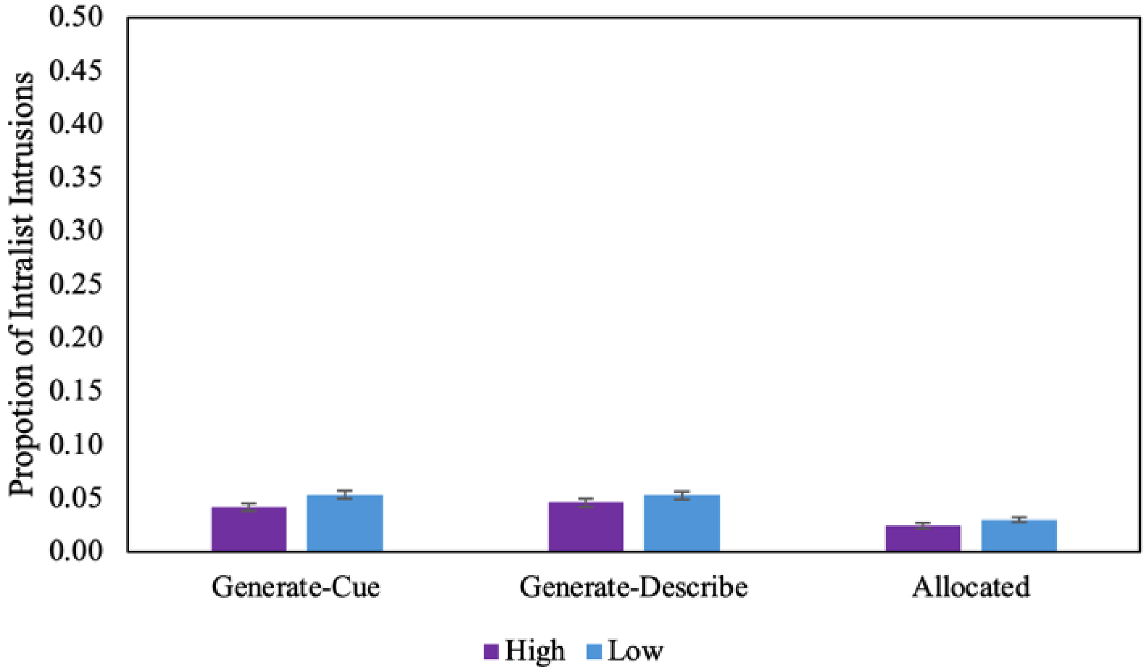


Figure A9.

Experiment 1 Proportion of Extra-list Intrusions by Condition and WM.

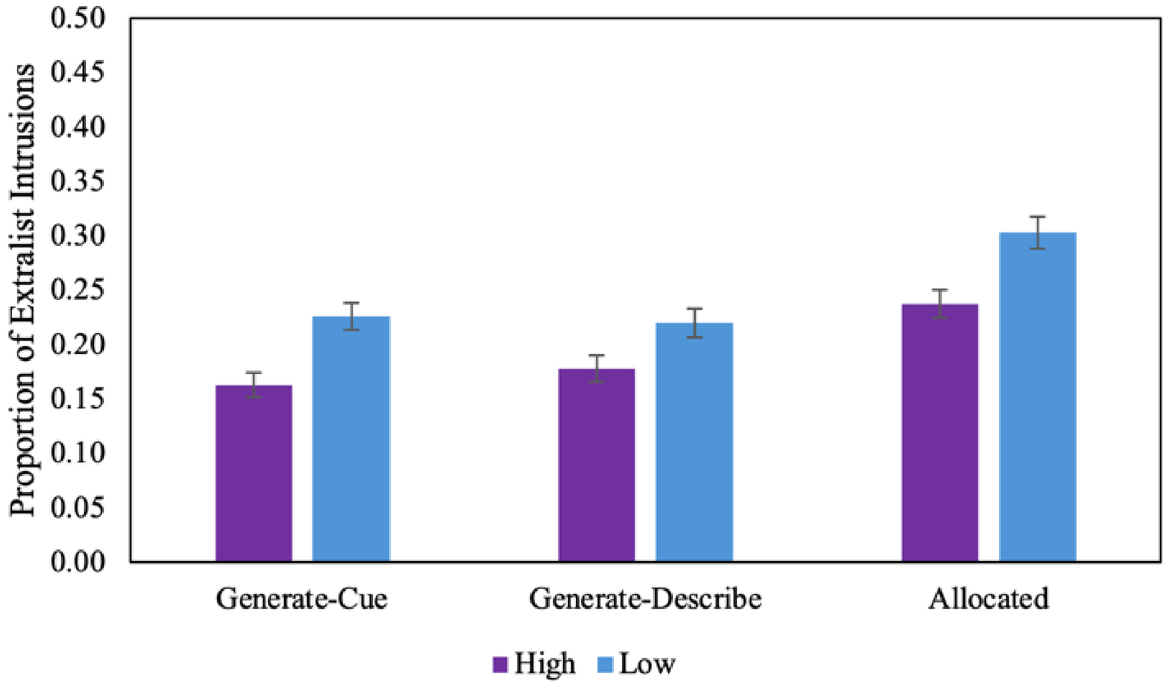


Figure A10.

Experiment 1 Proportion of Omission Errors by Condition and WM.

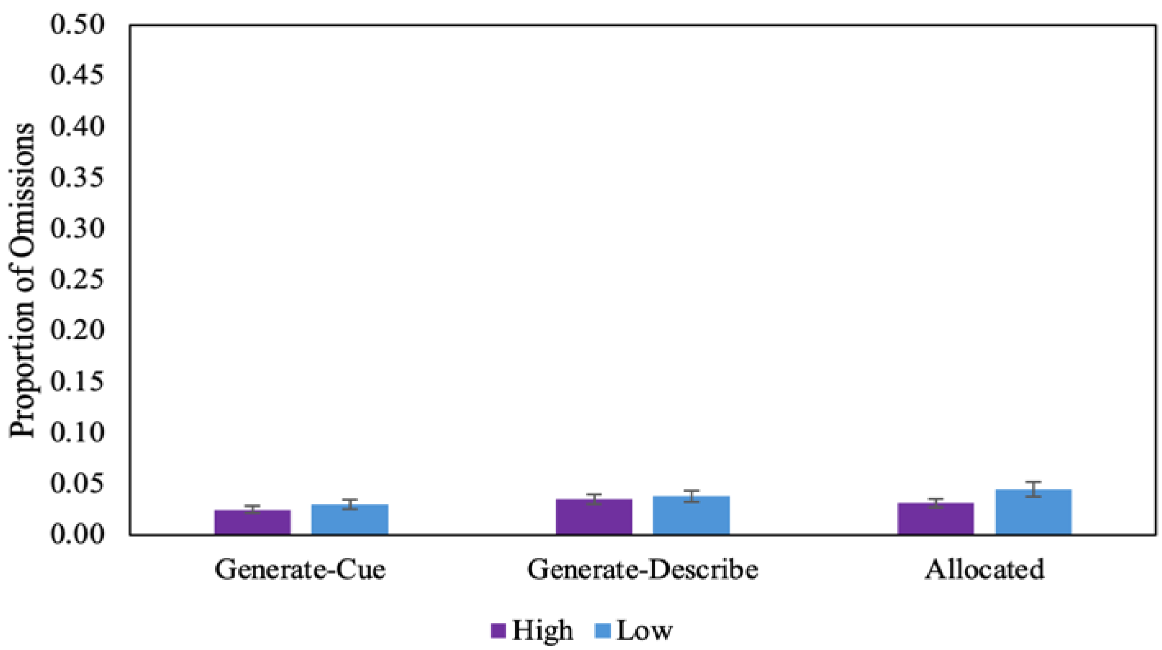


Figure A11.

Experiment 1 Mean Number of Associated Targets by Condition and WM.

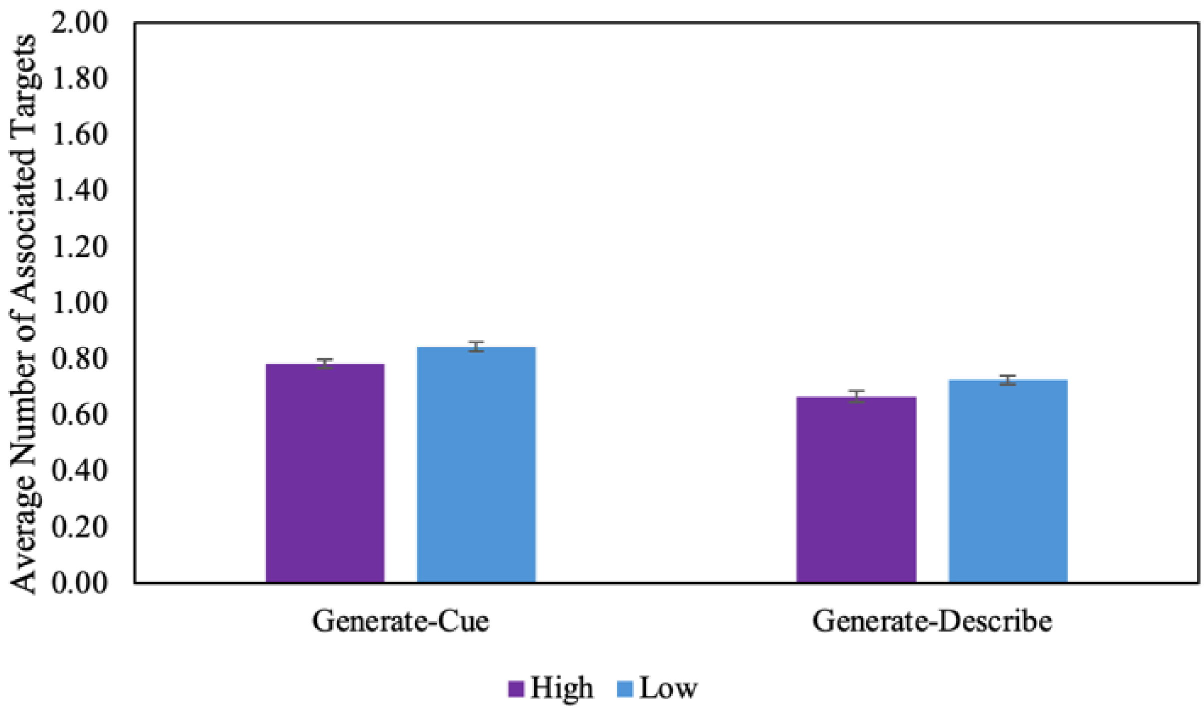


Figure A12.

Experiment 1 Mean Cumulative Associative Strength by Condition and WM.

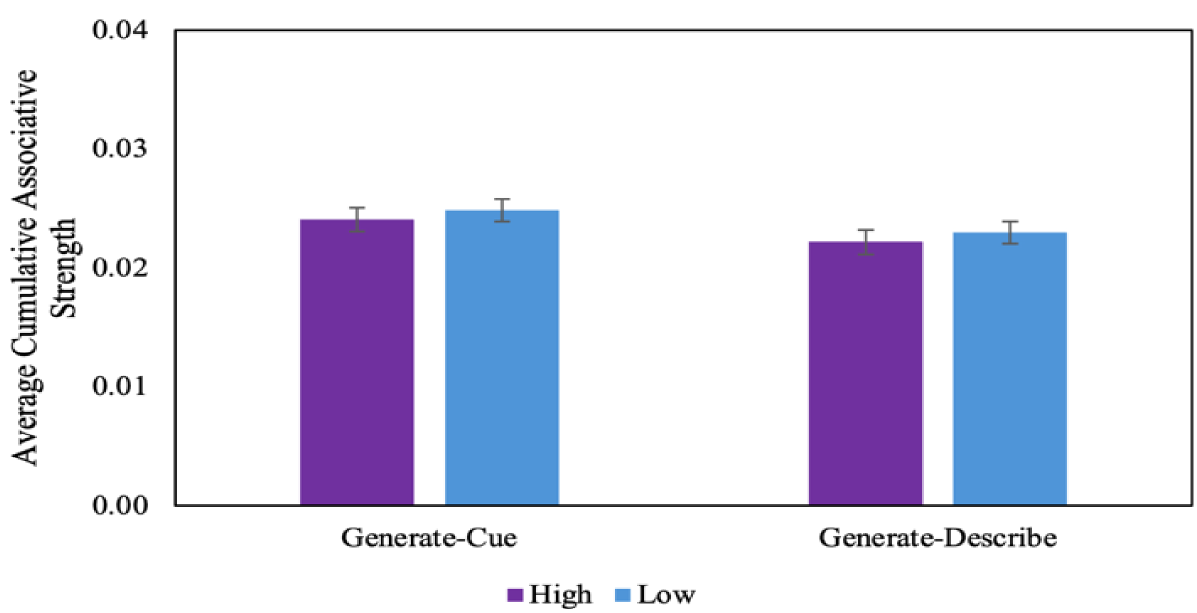


Figure A13.

Experiment 1 Proportion of Cues in the Database by Condition and WM.

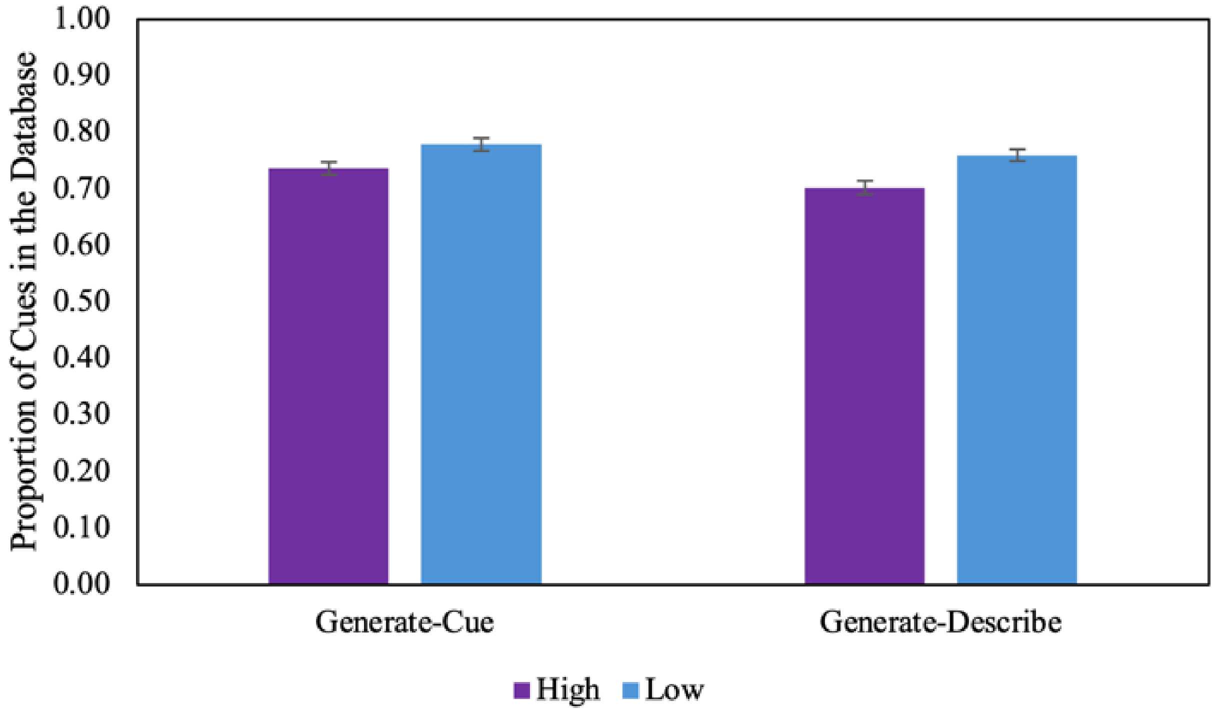
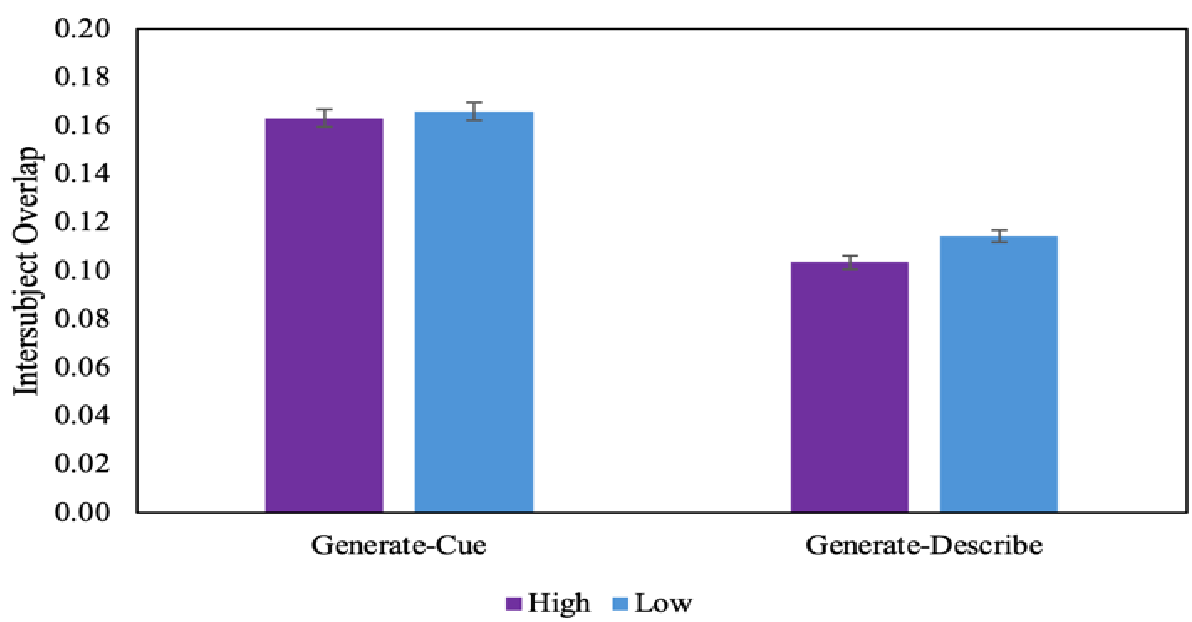


Figure A14.

Experiment 1 Intersubject Overlap by Condition and WM.



APPENDIX B. EXPERIMENT 2 SUPPLEMENTAL

Table B1

Experiment 2 Cue Intrusions

Variables	Mean	SD	Min	Max	Skew	Kurtosis
Cue Intrusions						
Generate CR	.04	.05	.00	.21	1.72	2.61
Allocate CR	.01	.02	.00	.18	5.87	43.19
Generate FR	.13	.25	.00	.88	2.17	3.23
Allocate FR	.12	.20	.00	.90	2.33	5.48
Intra-experimental Intrusions						
Generate CR	.08	.07	.00	.32	1.23	1.03
Allocate CR	.04	.03	.00	.21	2.44	10.27
Extra-experimental Intrusions						
Generate CR	.13	.10	.01	.44	1.33	1.17
Allocate CR	.30	.21	.00	.81	0.56	-0.81
Generate FR	.11	.10	.00	.50	1.54	3.57
Allocate FR	.19	.20	.00	.78	1.23	0.71

NOTE: CR = Cued Recall; FR = Free Recall; Intra-experimental Intrusions = Targets and Cues; Extra-experimental Intrusions = Non-targets and Non-cues.

Table B2

Experiment 2 Cue Properties - Tullis & Benjamin Calculation

Variables	Mean	SD	Min	Max	Skew	Kurtosis
Number of Associated Targets						
Generate CR	9.66	2.12	3.96	13.65	-0.27	-0.66
Allocate CR	5.80	0.63	4.06	7.09	-0.53	0.25
Generate FR	10.21	1.90	6.26	14.00	-0.19	-0.64
Allocate FR	5.80	0.63	4.38	7.46	0.13	-0.09
Cumulative Strength						
Generate CR	0.52	0.11	0.22	0.70	-0.34	-0.73
Allocate CR	0.33	0.03	0.24	0.40	-0.29	0.14
Generate FR	0.53	0.10	0.22	0.70	-0.38	-0.63
Allocate FR	0.33	0.03	0.25	0.39	-0.29	-0.11

NOTE: CR = Cued Recall; FR = Free Recall; Cumulative Strength = Cumulative Forward Associative Strength.

Recall Latency - Separate Cued and Free Recall Analyses

An independent-samples *t*-test did not show a significant difference between conditions in recall latency for cued recall, $t(147) = 1.80$, $p = .074$, $d = .30$ [-.03, .62]. However, an independent-samples *t*-test did show a significant difference between conditions in recall latency for free recall, $t(131) = 3.18$, $p = .002$, $d = .55$ [.20, .90]. Recall latency was lower for the generated cue condition (M = 8535.71, SD = 5199.58) relative to the allocated cue condition (M = 12270.94, SD = 7875.36).

Cue Intrusions

Cue intrusions were examined in both cued and free recall conditions. A significant difference in cue intrusions was observed between conditions. This was confirmed with a 2-way ANOVA with encoding condition and retrieval condition as between-subject factors. There was not a significant main effect of encoding condition, $F(1, 278) = 1.50$, $MSE = .02$, $p = .222$, η_p^2

$< .01$. However, the main effect of retrieval condition was significant, $F(1, 278) = 27.76$, $MSE = .02$, $p < .001$, $\eta_p^2 = .09$. Cue intrusions were significantly lower for the cued recall conditions relative to the free recall conditions, $t(280) = 5.16$, $p < .001$. The interaction between encoding and retrieval conditions was not significant, $F(1, 278) = 0.10$, $MSE = .02$, $p = .754$, $\eta_p^2 < .01$.

An independent-samples t-test showed a significant difference between conditions in intra-experimental intrusions, $t(147) = -4.77$, $p < .001$, $d = -0.78$ [-1.11, -0.45]. Intra-experimental intrusions were significantly lower for the allocate condition (M = .04, SD = .03) relative to the generate condition (M = .08, SD = .07).

Next, extra-experimental intrusions were examined in both cued and free recall conditions. A significant difference in extra-experimental intrusions was observed between conditions. This was confirmed with a 2-way ANOVA with encoding condition and retrieval condition as between-subjects factors. There was a significant main effect of encoding condition, $F(1, 278) = 42.47$, $MSE = .03$, $p < .001$, $\eta_p^2 = .13$. Extra-experimental intrusions were significantly lower for the generated cue conditions relative to the allocated cue conditions, $t(280) = 6.55$, $p < .001$. The main effect of retrieval condition was also significant, $F(1, 278) = 10.49$, $MSE = .03$, $p = .001$, $\eta_p^2 = .04$. Extra-experimental intrusions were significantly lower for the free recall conditions relative to the cued recall conditions, $t(280) = -3.25$, $p = .001$. The interaction between encoding and retrieval conditions was also significant, $F(1, 278) = 5.61$, $MSE = .03$, $p = .019$, $\eta_p^2 = .02$. There was a significant difference in extra-experimental intrusions between cued and free recall in the allocated cue conditions, $t(280) = 4.00$, $p < .001$; however, there was not a significant difference in extra-experimental intrusions between cued and free recall in the generated cue conditions, $t(280) = 0.59$, $p = .550$.

Cue Intrusions and Working Memory

I conducted an ANCOVA with encoding condition (Generate vs. Allocate) and retrieval condition (Cued Recall vs. Free Recall) as between-subjects factors and WM composite *z*-score as a continuous covariate. Cue intrusion errors were the dependent variable. Mirroring the ANOVA results above, there was not a significant main effect of encoding condition, $F(1, 274) = 1.44$, $MSE = .03$, $p = .231$, $\eta_p^2 < .01$. The main effect of retrieval condition was significant, $F(1, 274) = 26.38$, $MSE = .03$, $p < .001$, $\eta_p^2 = .09$. The interaction between encoding and retrieval conditions was not significant, $F(1, 274) = 0.14$, $MSE = .03$, $p = .706$, $\eta_p^2 < .01$. The main effect of WM was not significant, $F(1, 274) = 0.28$, $MSE = .03$, $p = .597$, $\eta_p^2 < .01$. Additionally, the two way interaction between encoding condition and WM, $F(1, 274) = 0.22$, $MSE = .03$, $p = .641$, $\eta_p^2 < .01$, as well as the two way interaction between retrieval condition and WM, $F(1, 274) = 0.11$, $MSE = .03$, $p = .742$, $\eta_p^2 < .01$, were not significant. Finally, the three way interaction between encoding condition, retrieval condition, and WM was not significant, $F(1, 274) = 0.03$, $MSE = .03$, $p = .855$, $\eta_p^2 < .01$.

I conducted an ANCOVA with encoding condition as a between-subjects factor and WM composite *z*-score as a continuous covariate. Intra-experimental intrusion errors were the dependent variable. Mirroring, the *t*-test results above, a significant main effect of encoding condition was observed, $F(1, 145) = 25.28$, $MSE = .003$, $p < .001$, $\eta_p^2 = .15$. The main effect of WM was also significant, $F(1, 145) = 8.75$, $MSE = .003$, $p = .004$, $\eta_p^2 = .06$, indicating that individuals with high WM made fewer intra-experimental intrusion errors. However, the interaction between encoding condition and WM was not significant, $F(1, 145) = 2.41$, $MSE = .003$, $p = .122$, $\eta_p^2 = .02$.

I conducted an ANCOVA with encoding condition and retrieval condition as between-subjects factors and WM composite *z*-score as a continuous covariate. Extra-experimental intrusion errors were the dependent variable. Mirroring the ANOVA results above, a significant main effect of encoding condition was observed, $F(1, 274) = 42.83$, $MSE = .03$, $p < .001$, $\eta_p^2 = .14$. The main effect of retrieval condition was also significant, $F(1, 274) = 12.19$, $MSE = .03$, $p < .001$, $\eta_p^2 = .04$, as well as the interaction between encoding and retrieval conditions, $F(1, 274) = 6.17$, $MSE = .03$, $p = .014$, $\eta_p^2 = .02$. The main effect of WM was significant, $F(1, 274) = 7.35$, $MSE = .03$, $p = .007$, $\eta_p^2 = .03$, indicating that individuals with high WM made fewer extra-experimental intrusions. However, the two way interaction between encoding condition and WM, $F(1, 274) = 3.10$, $MSE = .03$, $p = .179$, $\eta_p^2 = .01$, as well as the two way interaction between retrieval condition and WM, $F(1, 274) = 0.29$, $MSE = .03$, $p = .590$, $\eta_p^2 < .01$, were not significant. Finally, the three way interaction between encoding condition, retrieval condition, and WM was not significant, $F(1, 274) = 0.03$, $MSE = .03$, $p = .859$, $\eta_p^2 < .01$.

Cue Properties and Working Memory (Tullis and Benjamin (2015) Calculation)

WM was not significantly correlated with the number of associated targets (r = -.01, $p = .881$) or cumulative associative strength (r = -.01, $p = .921$).

Cue Properties with Allocate Condition

An independent-samples *t*-test showed a significant difference between conditions in forward associative strength, $t(280) = -22.18$, $p < .001$, $d = -2.65$ [-2.97, -2.32]. Forward associative strength was lower in the allocated cue condition (M = .02, SD = .01) relative to the generated cue condition (M = .06, SD = .02).

An independent-samples *t*-test showed a significant difference between conditions in the number of associated targets, $t(280) = -31.96$, $p < .001$, $d = -3.81$ [-4.20, -3.42]. Cues were associated with fewer targets in the allocated cue condition (M = .34, SD = .06) relative to the generated cue condition (M = .76, SD = .15).

An independent-samples *t*-test showed a significant difference between conditions in cumulative associative strength, $t(280) = -24.72$, $p < .001$, $d = -2.95$ [-3.29, -2.61]. Cumulative associative strength was lower in the allocated cue condition (M = .01, SD = .002) relative to the generated cue condition (M = .02, SD = .01).

An independent-samples *t*-test showed a significant difference between conditions in the proportion of cues in the database, $t(280) = -24.61$, $p < .001$, $d = -2.94$ [-3.27, -2.60]. The proportion of cues in the database was lower in the allocated cue condition (M = .44, SD = .04) relative to the generated cue condition (M = .72, SD = .13).

An independent-samples *t*-test showed a significant difference between conditions in intersubject overlap, $t(280) = -18.55$, $p < .001$, $d = -2.21$ [-2.51, -1.91]. Intersubject overlap was lower in the allocated cue condition (M = .09, SD = .002) relative to the generated cue condition (M = .13, SD = .03).

Cue Properties with Allocate Condition (Tullis and Benjamin (2015) Calculation)

An independent-samples *t*-test showed a significant difference between conditions in the number of associated targets, $t(280) = -23.52$, $p < .001$, $d = -2.81$ [-3.13, -2.48]. Cues were associated with fewer targets in the allocated cue condition (M = 5.80, SD = .63) relative to the generated cue condition (M = 9.91, SD = 2.03).

An independent-samples *t*-test showed a significant difference between conditions in cumulative associative strength, $t(280) = -21.99$, $p < .001$, $d = -2.62$ [-2.94, -2.30]. Cumulative associative strength was lower in the allocated cue condition (M = .33, SD = .03) relative to the generated cue condition (M = .52, SD = .10).

Analyses with only Reading Span

Proportion Correct

I conducted an ANCOVA with encoding condition (Generate vs. Allocate) and retrieval condition (Cued Recall vs. Free Recall) as between-subjects factors and Reading span score as a continuous covariate to analyze proportion correct. There was not a significant main effect of encoding condition, $F(1, 274) = 3.49$, $MSE = .03$, $p = .063$, $\eta_p^2 = .01$. There was a significant main effect of retrieval condition, $F(1, 274) = 18.60$, $MSE = .03$, $p < .001$, $\eta_p^2 = .06$. The interaction between encoding and retrieval conditions was not significant, $F(1, 274) = 1.73$, $MSE = .03$, $p = .189$, $\eta_p^2 = .01$. There was a not a significant main effect of WM, $F(1, 274) = 1.61$, $MSE = .03$, $p = .206$, $\eta_p^2 = .01$. Additionally, the two way interaction between encoding condition and WM, $F(1, 274) = 0.08$, $MSE = .03$, $p = .782$, $\eta_p^2 < .01$, as well as the two way interaction between retrieval condition and WM, $F(1, 274) = 0.03$, $MSE = .03$, $p = .856$, $\eta_p^2 < .01$, were not significant. Finally, the three way interaction between encoding condition, retrieval condition, and WM was not significant, $F(1, 274) = 0.56$, $MSE = .03$, $p = .454$, $\eta_p^2 < .01$.

Recall Latency

I conducted an ANCOVA with encoding condition (Generate vs. Allocate) and retrieval condition (Cued Recall vs. Free Recall) as between-subjects factors and Reading span score as a continuous covariate to analyze recall latency. A significant main effect of encoding condition was observed, $F(1, 274) = 5.34$, $MSE = 21819523.98$, $p = .022$, $\eta_p^2 = .02$, as well as a significant main effect of retrieval condition, $F(1, 274) = 7.44$, $MSE = 21819523.98$, $p = .007$, $\eta_p^2 = .03$. The interaction between encoding and retrieval conditions was not significant, $F(1, 274) = 3.45$, $MSE = 21819523.98$, $p = .064$, $\eta_p^2 = .03$. The main effect of WM was not significant, $F(1, 274) = 0.40$, $MSE = 21819523.98$, $p = .529$, $\eta_p^2 < .01$. Additionally, the two way interaction between encoding condition and WM, $F(1, 274) = 2.97$, $MSE = 21819523.98$, $p = .086$, $\eta_p^2 = .01$, as well as the two way interaction between retrieval condition and WM, $F(1, 274) = 0.24$, $MSE = 21819523.98$, $p = .626$, $\eta_p^2 < .01$, were not significant. Finally, the three way interaction between encoding condition, retrieval condition, and WM was not significant, $F(1, 274) = 1.77$, $MSE = 21819523.98$, $p = .185$, $\eta_p^2 = .01$.

Recall Errors

I conducted an ANCOVA with encoding condition as a between-subjects factor and Reading Span score as a continuous covariate. Intra-list intrusion errors were the dependent variable. The main effect of encoding condition was not significant, $F(1, 145) = 1.06$, $MSE = .001$, $p = .305$, $\eta_p^2 = .01$. Additionally, the main effect of WM was not significant, $F(1, 145) = 2.02$, $MSE = .001$, $p = .157$, $\eta_p^2 = .01$. The interaction between encoding condition and WM was also not significant, $F(1, 145) = 0.19$, $MSE = .001$, $p = .665$, $\eta_p^2 < .01$.

I conducted an ANCOVA with encoding condition (Generate vs. Allocate) as a between-subjects factor and Reading Span score as a continuous covariate. Omission errors were the dependent variable. The main effect of encoding condition was not significant, $F(1, 145) = 0.23$, $MSE = .02$, $p = .632$, $\eta_p^2 < .01$. The main effect of WM was not significant, $F(1, 145) = 0.94$, $MSE = .02$, $p = .334$, $\eta_p^2 = .01$. Additionally, the interaction between encoding condition and WM was not significant, $F(1, 145) = 2.44$, $MSE = .02$, $p = .121$, $\eta_p^2 = .02$.

I conducted an ANCOVA with encoding condition and retrieval condition as between-subjects factors and Reading Span score as a continuous covariate. Extra-list intrusion errors were the dependent variable. There was not a significant main effect of encoding condition, $F(1, 274) = 2.62$, $MSE = .05$, $p = .107$, $\eta_p^2 = .01$. The main effect of retrieval condition was also not significant, $F(1, 274) = 0.03$, $MSE = .05$, $p = .856$, $\eta_p^2 < .01$. The interaction between encoding and retrieval conditions was not significant, $F(1, 274) = 1.53$, $MSE = .05$, $p = .218$, $\eta_p^2 = .01$. The main effect of WM was not significant, $F(1, 274) = 1.75$, $MSE = .05$, $p = .187$, $\eta_p^2 = .01$. The two way interaction between encoding condition and WM, $F(1, 274) = 0.96$, $MSE = .05$, $p = .329$, $\eta_p^2 < .01$, as well as the two way interaction between retrieval condition and WM, $F(1, 274) = 0.001$, $MSE = .05$, $p = .970$, $\eta_p^2 < .01$, were not significant. Finally, the three way interaction between encoding condition, retrieval condition, and WM was not significant, $F(1, 274) = 1.05$, $MSE = .05$, $p = .307$, $\eta_p^2 < .01$.

Cue Properties

Reading Span score was not significantly correlated with forward associative strength (r = .09, *p* = .310), number of associated targets (r = .03, *p* = .763), cumulative associative strength (r = .03, *p* = .756), proportion of cues in the database (r = .01, *p* = .922), or intersubject overlap (r = .09, *p* = .314).

Figures

In all appendix figures error bars represent ± 1 standard error of the mean. In figures B8-B14, WM is dichotomized into high and low groups for visualization. The high WM group is composed of participants with WM composite *z*-scores above the 50^{th} percentile. The low WM group is composed of participants with WM composite *z*-scores equal to or below the 50^{th} percentile.

Figure B1.

Experiment 2 Proportion of Intra-list Intrusions by Condition.

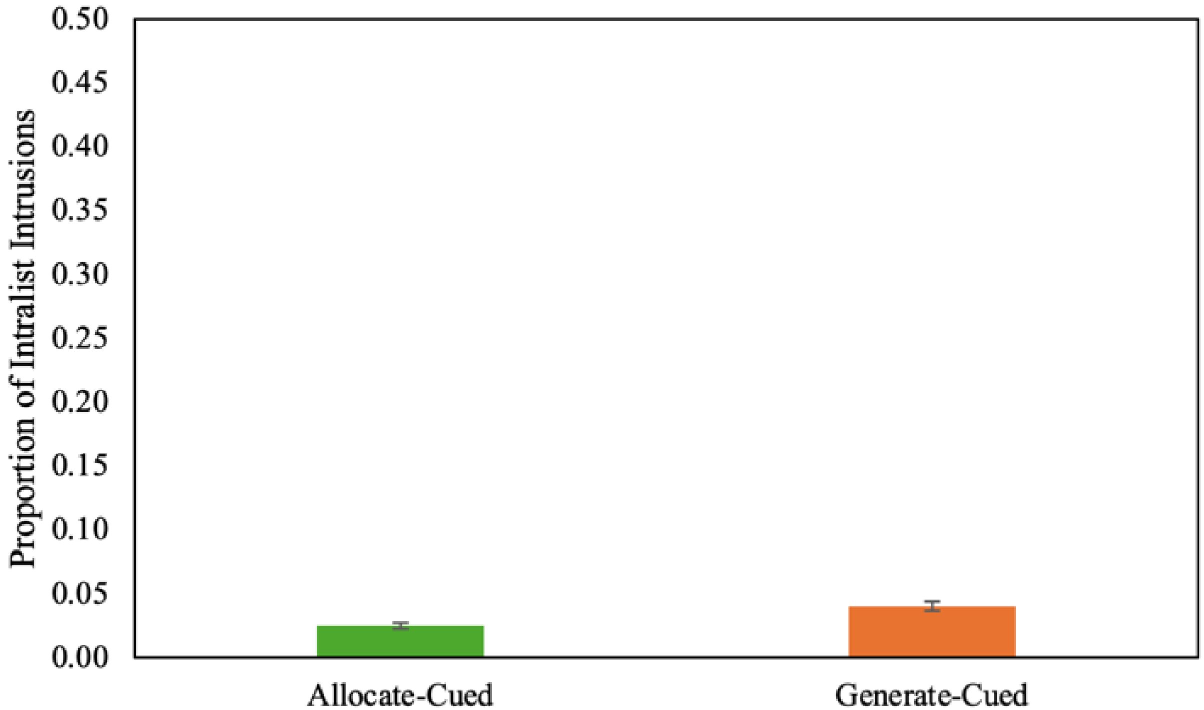


Figure B2.

Experiment 2 Proportion of Extra-list Intrusions by Condition.

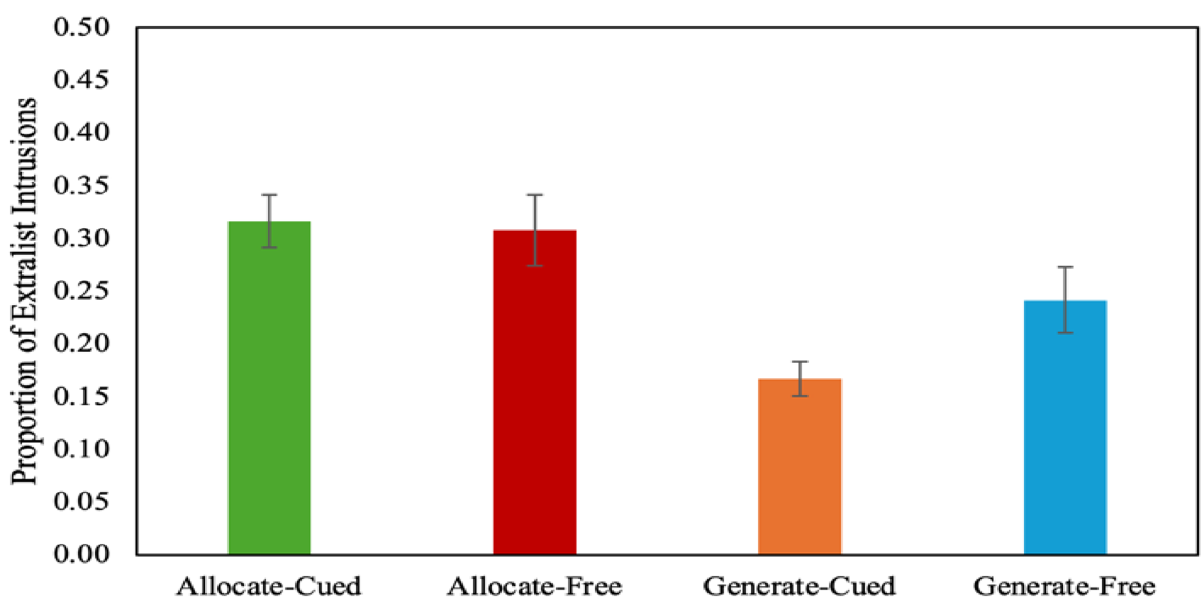


Figure B3.

Experiment 2 Proportion of Omission Errors by Condition.

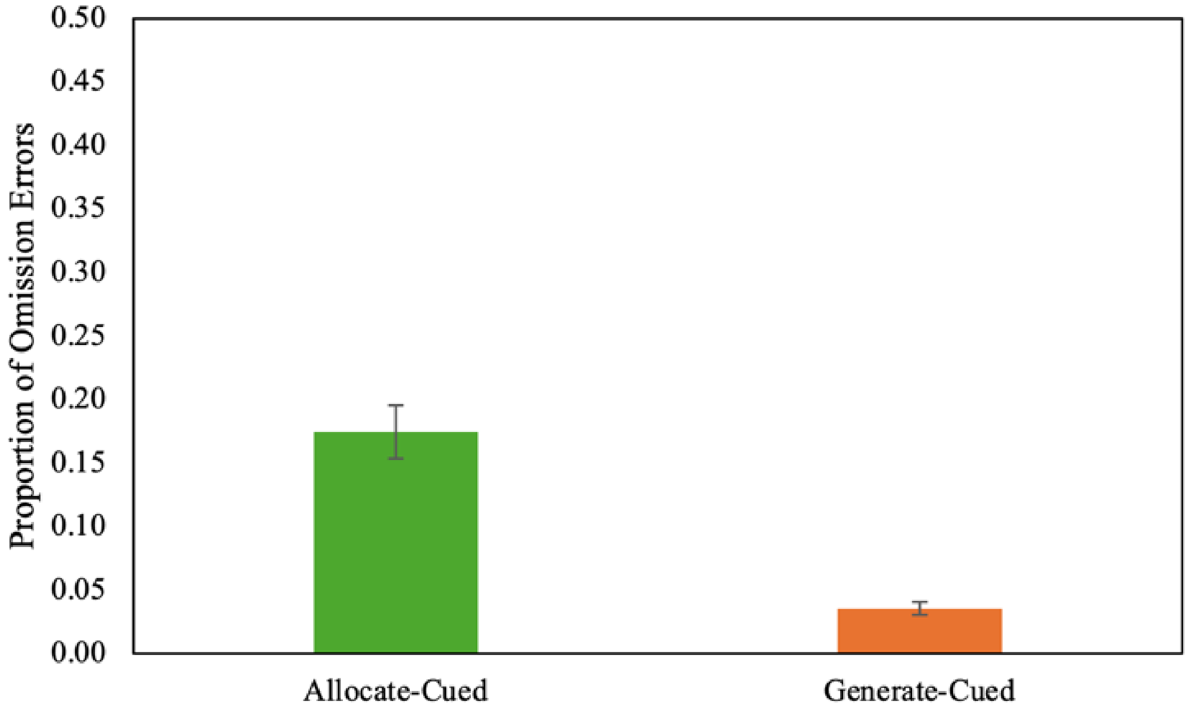


Figure B4.

Experiment 2 Mean Number of Associated Targets by Condition.

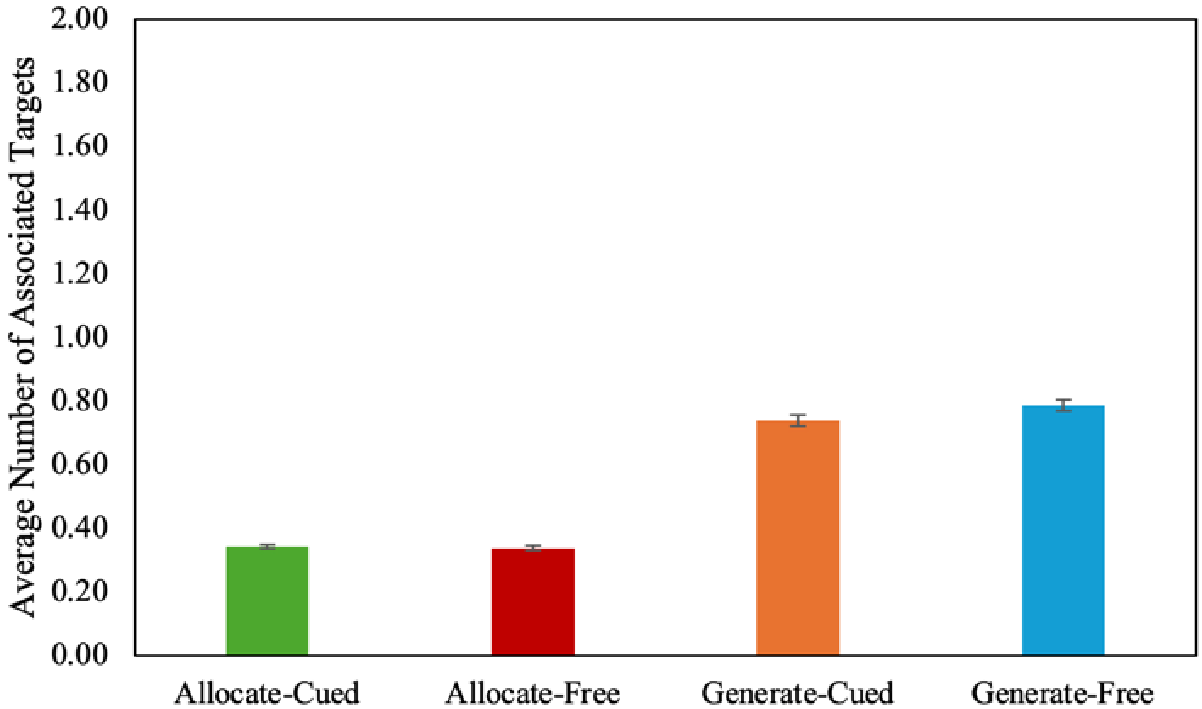


Figure B5.

Experiment 2 Mean Cumulative Associative Strength by Condition.

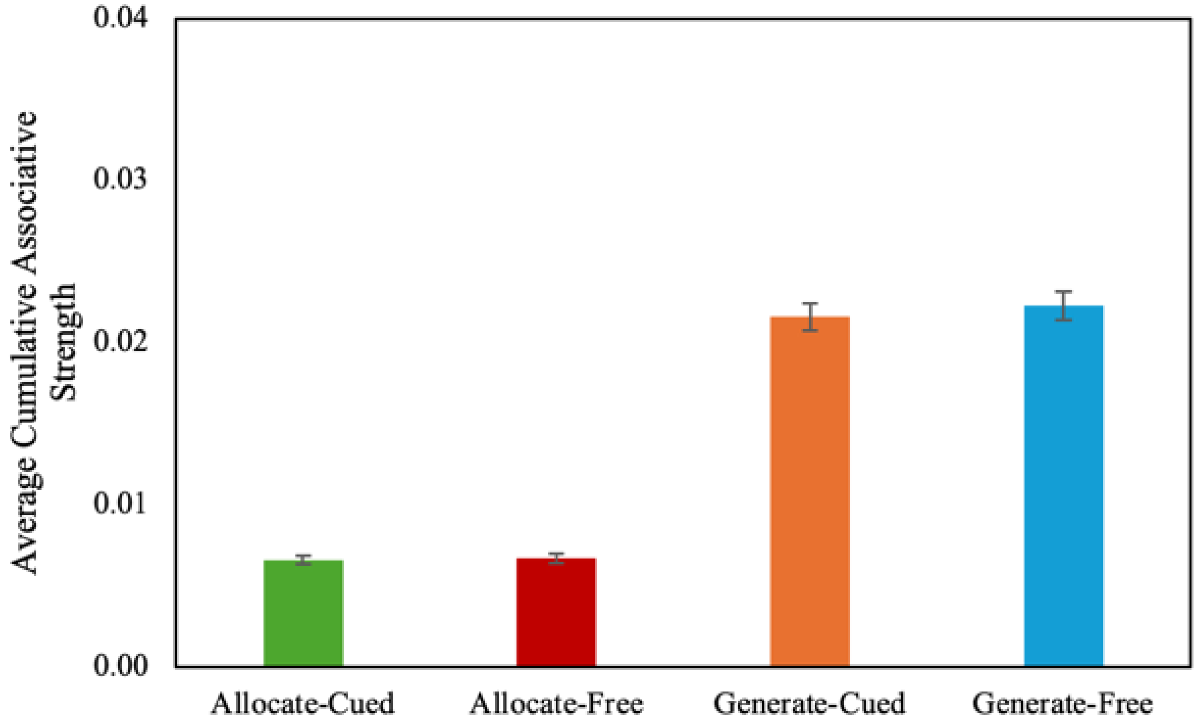


Figure B6.

Experiment 2 Proportion of Cues in the Database by Condition.

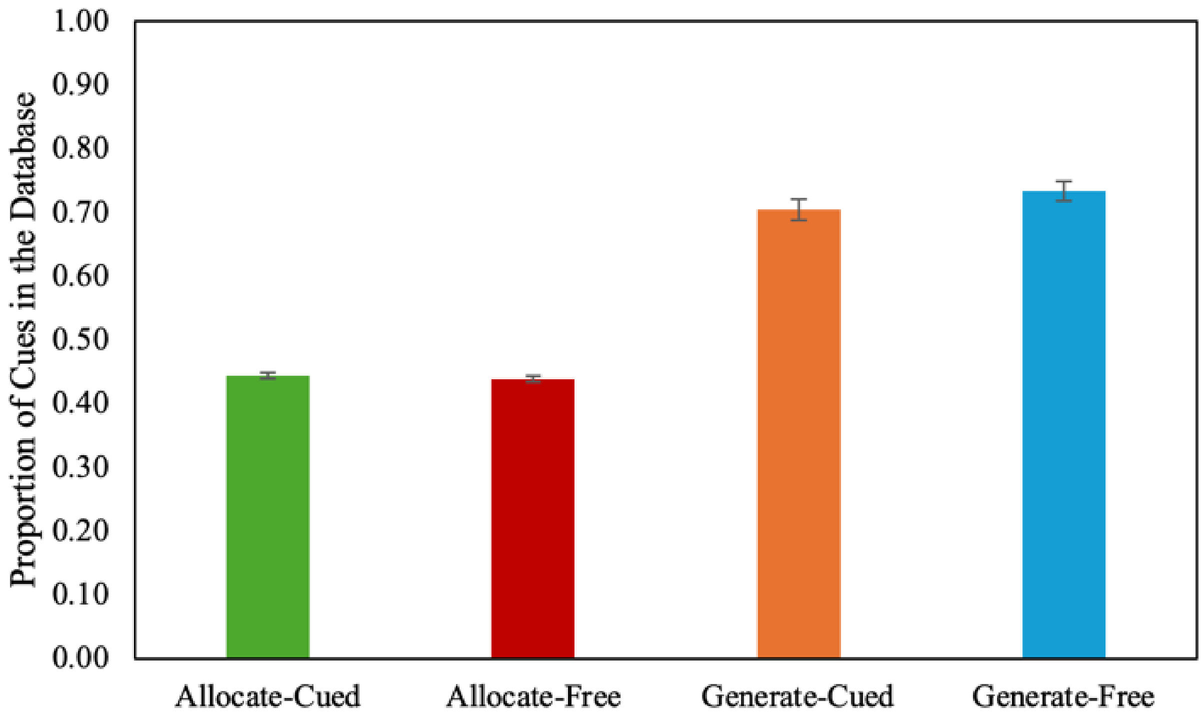


Figure B7.

Experiment 2 Intersubject Overlap by Condition.

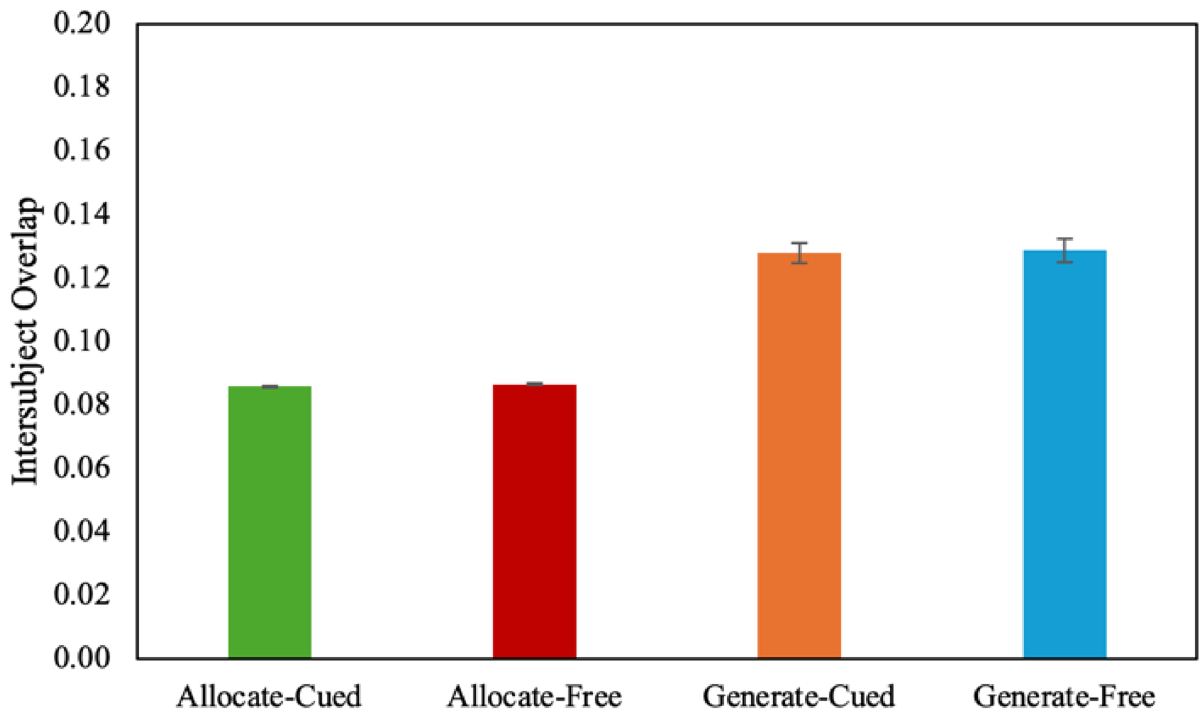


Figure B8.

Experiment 2 Proportion of Intra-list Intrusions by Condition and WM.

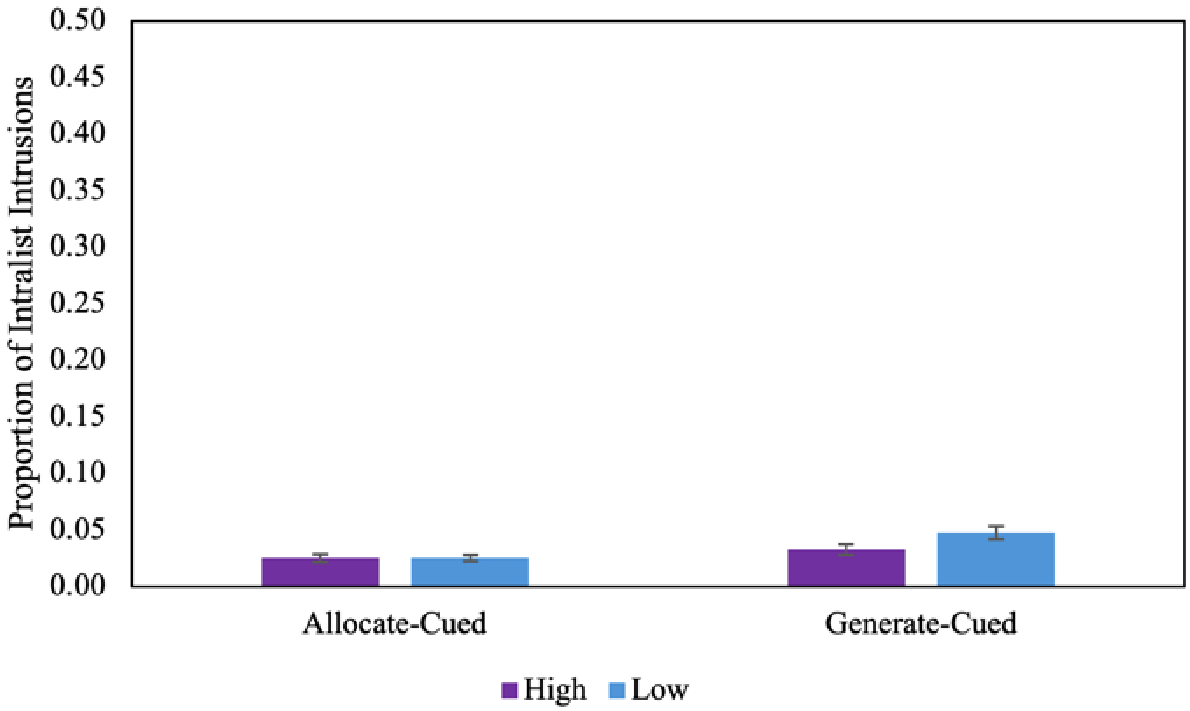


Figure B9.

Experiment 2 Proportion of Extra-list Intrusions by Condition and WM.

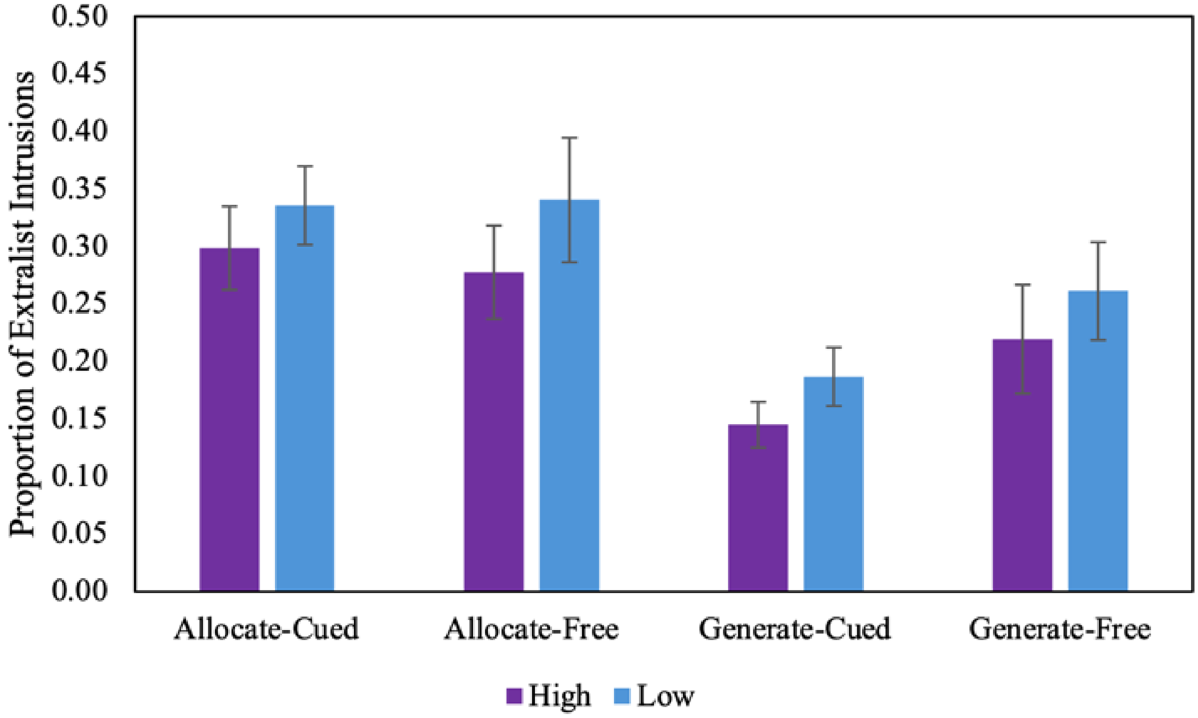


Figure B10.

Experiment 2 Proportion of Omission Errors by Condition and WM.

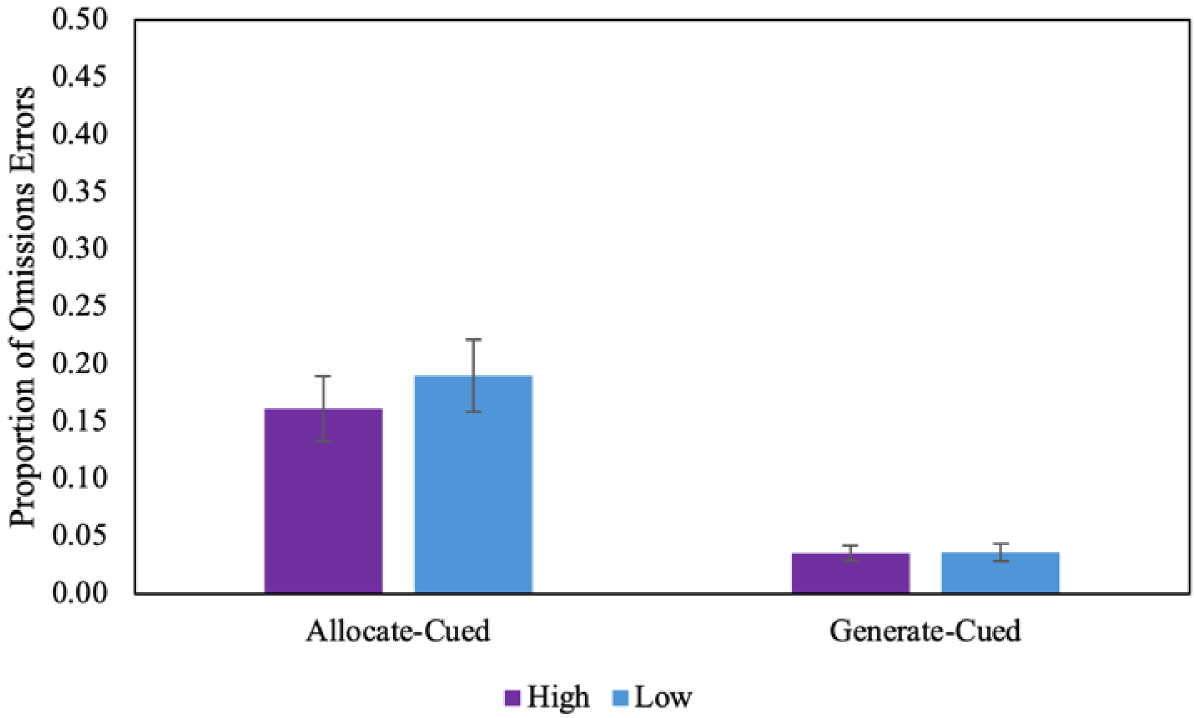


Figure B11.

Experiment 2 Mean Number of Associated Targets by Condition and WM.

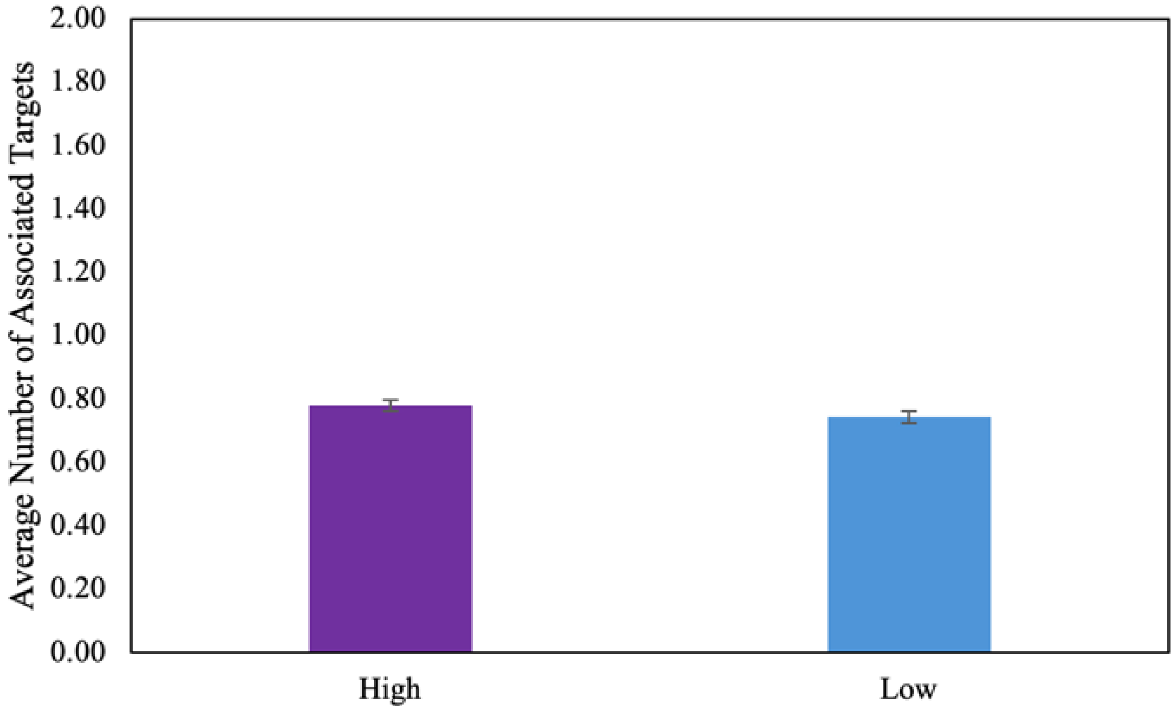


Figure B12.

Experiment 2 Mean Cumulative Associative Strength by Condition and WM.

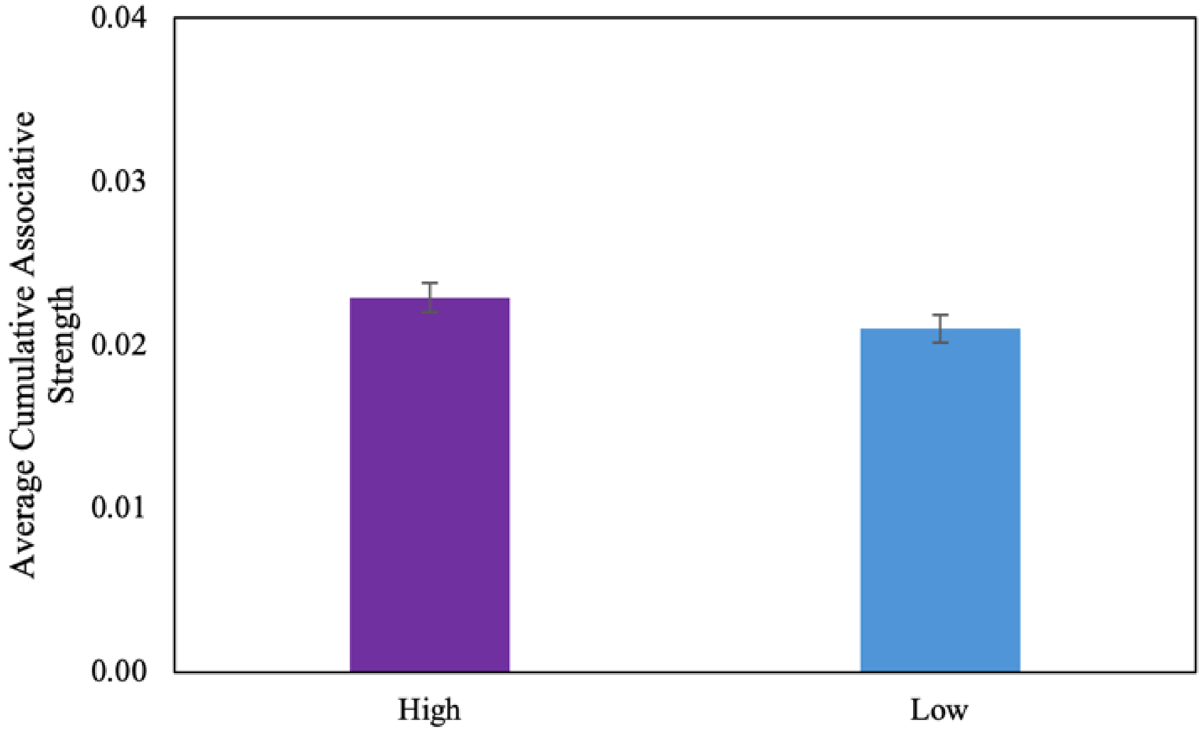


Figure B13.

Experiment 2 Proportion of Cues in the Database by Condition and WM.

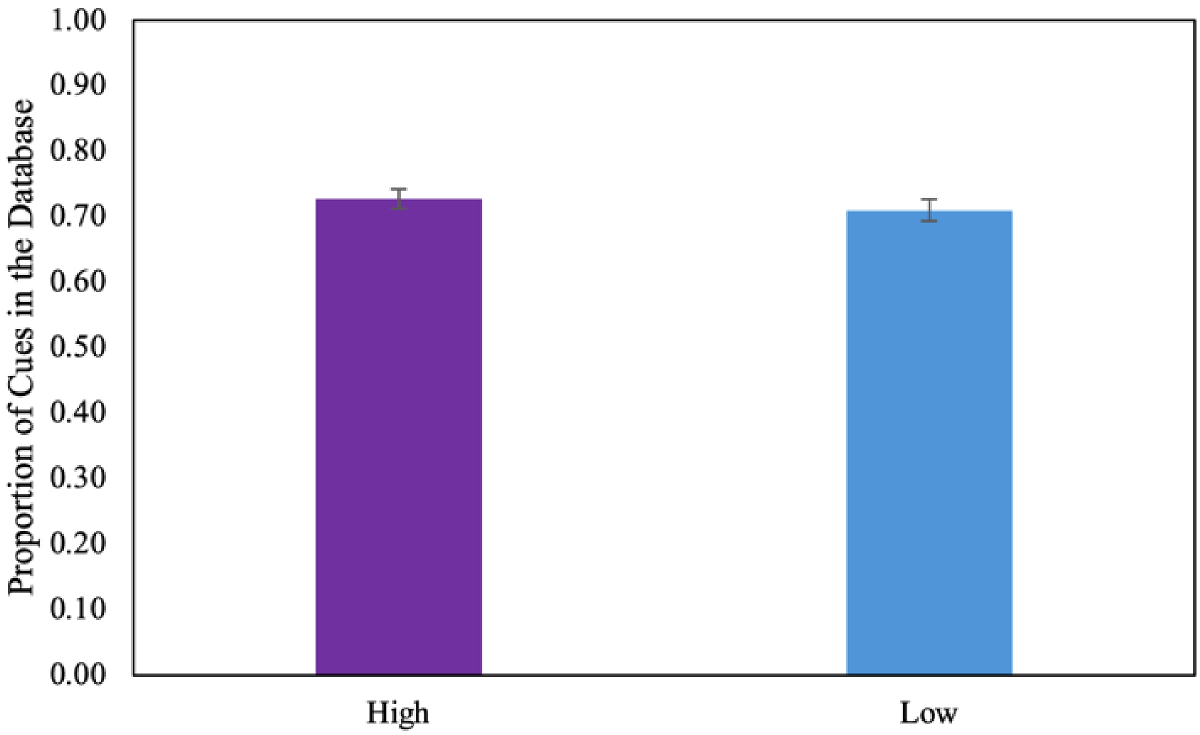
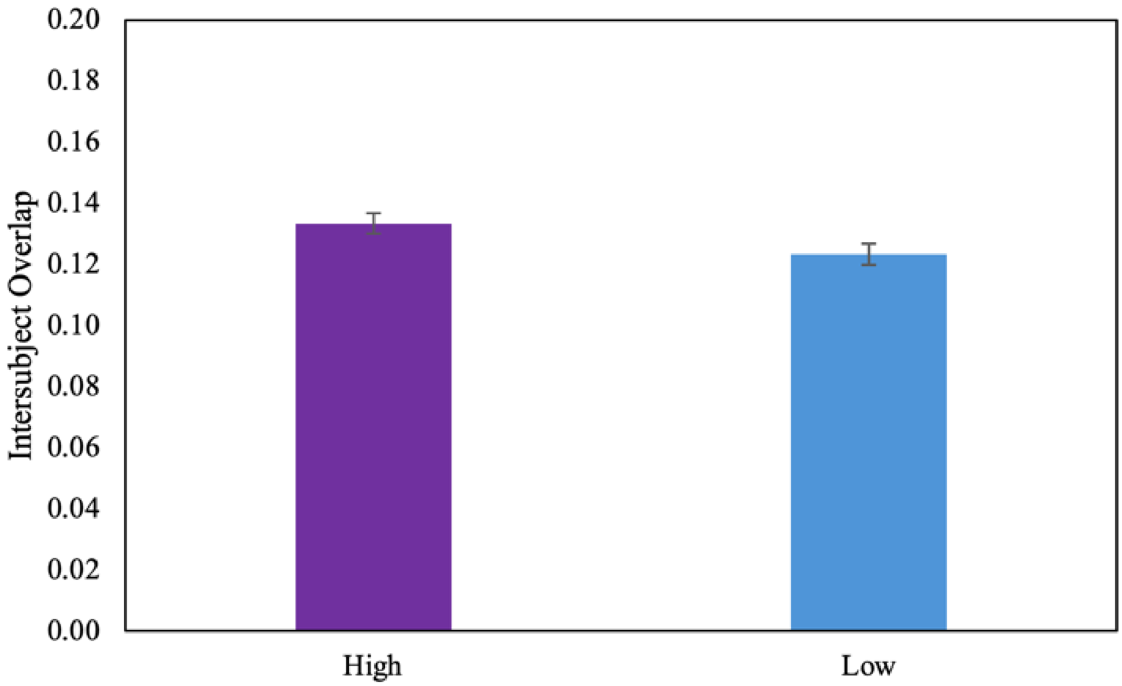


Figure B14.

Experiment 2 Intersubject Overlap by Condition and WM.



REFERENCES

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1-29. https://doi.org/10.1146/annurev-psych-120710-100422

Bailey, H., Dunlosky, J., & Kane, M. J. (2008). Why does working memory span predict complex cognition? Testing the strategy affordance hypothesis. *Memory & Cognition*, *36*, 1383-1390. https://doi.org/10.3758/MC.36.8.1383

Bartsch, L. M., Loaiza, V. M., Jäncke, L., Oberauer, K., & Lewis-Peacock, J. A. (2019). Dissociating refreshing and elaboration and their impacts on memory. *Neuroimage*, *199*, 585-597. https://doi.org/10.1016/j.neuroimage.2019.06.028

Bartsch, L. M., & Oberauer, K. (2021). The effects of elaboration on working memory and long-term memory across age. *Journal of Memory and Language*, *118*, 104215. https://doi.org/10.1016/j.jml.2020.104215

Bartsch, L. M., Singmann, H., & Oberauer, K. (2018). The effects of refreshing and elaboration on working memory performance, and their contributions to long-term memory formation. *Memory & Cognition*, *46*, 796-808. https://doi.org/10.3758/s13421-018-0805-9

Bousfield, W. A., & Sedgewich, C. H. W. (1944). An analysis of sequences of restricted associative responses. *The Journal of General Psychology*, 30(2), 149-165. https://doi.org/10.1080/00221309.1944.10544467

Cokely, E. T., Kelley, C. M., & Gilchrist, A. L. (2006). Sources of individual differences in working memory: Contributions of strategy to capacity. *Psychonomic Bulletin & Review*, *13*, 991-997. https://doi.org/10.3758/BF03213914

Cowan, N. (1999). An embedded-process model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62-101). New York: Cambridge University Press.

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin and Review, 24*, 1158-1170. https://doi.org/10.3758/s13423-016-1191-6

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology, 51*(1), 42-100. https://doi.org/10.1016/j.cogpsych.2004.12.001

Dunlosky, J., & Kane, M. J. (2007). The contributions of strategy use to working memory span: A comparison of strategy assessment methods. *Quarterly Journal of Experimental Psychology*, *60*(9), 1227-1245. https://doi.org/10.1080/17470210600926075

Engle, R.W., Kane, M.J., & Tuholski, S.W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102-134). New York: Cambridge University Press.

Goodhew, S. C., & Edwards, M. (2019). Translating experimental paradigms into individual-differences research: Contributions, challenges, and practical recommendations. *Consciousness and Cognition*, *69*, 14-25. https://doi.org/10.1016/j.concog.2019.01.008

Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology, 33*, 457–461. https://doi.org/10.1111/j.1469-8986.1996.tb01071.x

Granholm, E., Morris, S. K., Sarkin, A. J., Asarnow, R. F., & Jeste, D. V. (1997). Pupillary responses index overload of working memory resources in schizophrenia. *Journal of Abnormal Psychology, 106*, 458–467. https://doi.org/10.1037/0021-843X.106.3.458

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 923 –941. https://doi.org/10.1037/0278-7393.25.4.923

Kaakinen, J. K., & Hyönä, J. (2007). Strategy use in the reading span test: An analysis of eye movements and reported encoding strategies. *Memory*, *15*(6), 634-646. https://doi.org/10.1080/09658210701457096

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working-memory capacity as variation in executive attention and control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 21-48). New York: Oxford University Press.

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(2), 336-358. https://doi.org/10.1037/0278-7393.26.2.336

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. Science, 154, 1583–1585. https://doi.org/10.1126/science.154.3756.1583

Mäantylä, T. (1986). Optimizing cue effectiveness: Recall of 500 and 600 incidentally learned words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(1), 66.

Mäntylä, T., & Nilsson, L. G. (1988). Cue distinctiveness and forgetting: Effectiveness of self-generated retrieval cues in delayed recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 502-509.

McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory & Cognition*, *29*, 10-17. https://doi.org/10.3758/BF03195736

McVay, J. C., & Kane, M. J. (2009). Conducting the train of thought: working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 196–204. https://doi.org/10.1037/a0014104.

Miller, A. L., Gross, M. P., & Unsworth, N. (2019). Individual differences in working memory capacity and long-term memory: The influence of intensity of attention to items at encoding as measured by pupil dilation. *Journal of Memory and Language*, *104*, 25-42. https://doi.org/10.1016/j.jml.2018.09.005

Miller, A. L., & Unsworth, N. (2018). Individual differences in working memory capacity and search efficiency. *Memory & Cognition*, *46*(7), 1149-1163. https://doi.org/10.3758/s13421-018-0827-3

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory*, *10*(5-6), 389-395. https://doi.org/10.1080/09658210244000216

Nelson, D. L., Brooks, D. H., & Borden, R. C. (1974). Effects of formal similarity: Phonetic, graphic, or both? *Journal of Experimental Psychology, 103*(1), 91–96. https://doi.org/10.1037/h0036821

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. http://www.usf.edu/FreeAssociation/.

Oberauer, K. (2009). Design for a working memory. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 51, pp. 45–100). Elsevier. https://doi.org/10.1016/S00797421(09)51002-X

Peavler, W. S. (1974). Pupil size, information overload, and performance differences. *Psychophysiology, 11*, 559–566. https://doi.org/10.1111/j.1469-8986.1974.tb01114.x

Raaijmakers, J. G., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In *Psychology of learning and motivation* (Vol. 14, pp. 207-262). Academic Press.

Robison, M. K., & Unsworth, N. (2017). Working memory capacity, strategic allocation of study time, and value-directed remembering. *Journal of Memory and Language*, *93*, 231-244. https://doi.org/10.1016/j.jml.2016.10.007

Redick, T. S., Shipstead, Z., Meier, M. E., Montroy, J. J., Hicks, K. L., Unsworth, N., ... & Engle, R. W. (2016). Cognitive predictors of a common multitasking ability: Contributions from working memory, attention control, and fluid intelligence. *Journal of Experimental Psychology: General*, *145*(11), 1473-1492. https://doi.org/10.1037/xge0000219

Robison, M. K., Miller, A. L., Wiemers, E. A., Ellis, D. M., Unsworth, N., Redick, T. S., & Brewer, G. A. (2024). What makes working memory work? A multifaceted account of the predictive power of working memory capacity. *Journal of Experimental Psychology: General*, *153*(9), 2193.

Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, *126*(3), 211-227.

Shipstead, Z., Redick, T. S., Hicks, K. L., & Engle, R. W. (2012). The scope and control of attention as separate aspects of working memory. *Memory*, *20*(6), 608-628. https://doi.org/10.1080/09658211.2012.691519

Souza, A. S., & Oberauer, K. (2018). Does articulatory rehearsal help immediate serial recall?. *Cognitive Psychology*, *107*, 1-21. https://doi.org/10.1016/j.jml.2020.104215

Spillers, G. J., & Unsworth, N. (2011). Variation in working memory capacity and temporal–contextual retrieval from episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1532-1539. https://doi.org/ 10.1037/a0024852

Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology*, *86*(2), 255-262. https://doi.org/10.1037/h0029997

Tullis, J. G., & Benjamin, A. S. (2015). Cue generation: How learners flexibly support future retrieval. *Memory & Cognition*, *43*(6), 922-938. https://doi.org/10.3758/s13421-015-0517-3

Tullis, J. G., & Finley, J. R. (2021). What characteristics make self-generated memory cues effective over time?. *Memory*, *29*(10), 1308-1319. https://doi.org/10.1080/09658211.2021.1979585

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, *5*(4), 381-391. https://doi.org/10.1016/S0022-5371(66)80048-8

Turley-Ames, K. J., & Whitfield, M. M. (2003). Strategy training and working memory task performance. *Journal of Memory and Language*, *49*(4), 446-468. https://doi.org/10.1016/S0749-596X(03)00095-0

Unsworth, N. (2007). Individual differences in working memory capacity and episodic retrieval: Examining the dynamics of delayed and continuous distractor free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(6), 1020 – 1034. https://doi.org/10.1037/0278-7393.33.6.1020

Unsworth, N. (2009a). Individual differences in self-initiated processing at encoding and retrieval: A latent variable analysis. *Quarterly Journal of Experimental Psychology*, *62*(2), 257-266. https://doi.org/10.1080/174702108023730

Unsworth, N. (2009b). Examining variation in working memory capacity and retrieval in cued recall. *Memory*, *17*(4), 386-396. https://doi.org/10.1080/09658210902802959

Unsworth, N. (2009c). Variation in working memory capacity, fluid intelligence, and episodic recall: A latent variable examination of differences in the dynamics of free recall. *Memory & Cognition*, *37*, 837-849. https://doi.org/10.3758/MC.37.6.837

Unsworth, N. (2016). Working memory capacity and recall from long-term memory: Examining the influences of encoding strategies, study time allocation, search efficiency, and monitoring abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(1), 50 –61. https://doi.org/10.1037/xlm0000148

Unsworth, N. (2019). Individual differences in long-term memory. *Psychological Bulletin*, *145*(1), 79 –139. https://doi.org/10.1037/bul0000176

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2011). Variation in working memory capacity and episodic memory: Examining the importance of encoding specificity. *Psychonomic Bulletin & Review*, *18*, 1113-1118. https://doi.org/10.3758/s13423-011-0165-y

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2013). Working memory capacity and retrieval from long-term memory: The role of controlled search. *Memory & Cognition*, *41*, 242-254. https://doi.org/10.3758/s13421-012-0261-x

Unsworth, N., & Engle, R. W. (2006). A temporal–contextual retrieval account of complex span: An analysis of errors. *Journal of Memory and Language*, *54*(3), 346-362. https://doi.org/10.1016/j.jml.2005.11.004

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology, 71,* 1-26. https://doi.org/10.1016/j.cogpsych.2014.01.003

Unsworth, N., Miller, A. L., & Strayer, D. L. (2024). Individual differences in attention control: A meta-analysis and re-analysis of latent variable studies. *Psychonomic Bulletin & Review*, 1-47. https://doi.org/10.3758/s13423-024-02516-1

Unsworth, N., & Redick, T. S. (2017). Working memory and intelligence. In J. Wixted (Ed.), Cognitive Psychology of Memory, *Vol. 2 of Learning and Memory: A Comprehensive Reference*, 2nd edition, Byrne, J. H. (ed.). pp. 163-180. Oxford: Academic Press. https://doi.org/10.1016/B978-0-12-809324-5.21041-9

Unsworth, N., & Spillers, G. J. (2010). Variation in working memory capacity and episodic recall: The contributions of strategic encoding and contextual retrieval. *Psychonomic Bulletin & Review*, *17*, 200-205. https://doi.org/10.3758/PBR.17.2.200

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2012). Working memory capacity and retrieval limitations from long-term memory: An examination of differences in accessibility. *Quarterly Journal of Experimental Psychology*, *65*(12), 2397-2410. https://doi.org/10.1080/17470218.2012.69043

VanArsdall, J. E., & Blunt, J. R. (2022). Analyzing the structure of animacy: Exploring relationships among six new animacy and 15 existing normative dimensions for 1,200 concrete nouns. *Memory & Cognition*, *50*(5), 997-1012. https://doi.org/10.3758/s13421-021-01266-y

Van Gerven, P. W. M., Paas, F., Van Merrienboer, J. J. G., & Schmidt, H. G. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology, 41*, 167–174. https://doi.org/10.1111/j.1469-8986.2003.00148.x

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*(7067), 500-503. https://doi.org/10.1038/nature04171

Watkins, M. J., & Bloom, M. J. (1999). Selectivity in memory: An exploration of willful control over the remembering process. Unpublished manuscript.

Wheeler, R. L., & Gabbert, F. (2017). Using self-generated cues to facilitate recall: A narrative review. *Frontiers in Psychology*, *8*, 1830. https://doi.org/10.3389/fpsyg.2017.01830

Whitney, P., Ritchie, B. G., & Clark, M. B. (1991). Working-memory capacity and the use of elaborative inferences in text comprehension. *Discourse Processes*, *14*(2), 133-145. https://doi.org/10.1080/01638539109544779

Wixted, J. T., & Rohrer, D. (1993). Proactive interference and the dynamics of free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(5), 1024–1039. https://doi.org/10.1037/0278-7393.19.5.1024

Wixted, J. T., & Rohrer, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. *Psychonomic Bulletin & Review*, *1*, 89-106. https://doi.org/10.3758/BF03200763

PUBLICATION

Fordyce, A. L., Redick, T. S., Bedwell, J. P., & Karpicke, J.D. (2025). Individual differences in working memory ability and the benefit of retrieval practice. *Journal of Memory and Language, 144*, 1-16. https://doi.org/10.1016/j.jml.2025.104664