A Contextual Theory on the Generation Effect

BY

MATTHEW P. McCURDY
B.S., Indiana University, 2012
M.S., University of Tennessee at Chattanooga, 2014
M.A., University of Illinois at Chicago, 2017

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Psychology
in the Graduate College of the
University of Illinois at Chicago, 2020

Chicago, Illinois

Defense Committee:

Eric D. Leshikar, Chair and Advisor
James W. Pellegrino
Gary E. Raney
Karl K. Szpunar, Ryerson University
R. Reed Hunt, University of Mississippi

ACKNOWLEDGMENTS

It is without doubt in my mind that I would not have not made it this far in my life without the help of an extraordinary group of people that I am truly grateful for. I write this section to you all. First, and foremost I'd like to thank my parents who have always been there to teach, push, and support me throughout my life. Through your guidance and examples, I have learned that I can accomplish any goal I set for myself with enough focus, hard work, persistence, and patience. Thanks for always being my biggest supporters no matter where life takes me. Second, I'd like to thank several important people that have helped me on my academic journey to this point. Foremost, I'd like to thank Eric Leshikar, for believing in me and giving me a chance to learn and grow throughout my PhD program as both a researcher and a person. I couldn't have done it without you. I'd also like to thank George Rebec, who gave me my first opportunity to engage in research, which fueled my passion for pursuing research as a career. I'd also like to thank my committee members on this project: Jim Pellegrino, Gary Raney, Karl Szpunar, and Reed Hunt. Thank you all for pushing me to learn more from this project. Each of you have helped me in unique ways throughout my time as a graduate student, and I am truly grateful for that. Third, I want to thank my friend, Travis Necessary, who always provides a useful interruption or visit at the right (and sometimes wrong) times. The breaks and visits are "necessary" though, and I appreciate them. Lastly, I would like to thank my wife, Courtney Zulauf-McCurdy, for always being there through the ups and downs of graduate school and life. It is much easier to endure and enjoy these ups and downs with a friend, rather than alone.

MPM

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

TAP	Transfer-Appropriate Processing
MAP	Material-Appropriate Processing
SAP	Subject-Appropriate Processing

SUMMARY

The generation effect is the memory benefit for information that is self-generated compared to information that is read. Several boundary conditions have been identified for this effect, suggesting the memory benefit for generating is not universal, but instead depends on certain experimental factors. Although several theories have been proposed to account for the generation effect throughout the study of this memory phenomenon, the existing theories often only apply under a specific set of experimental conditions, limiting their explanatory power. Based on principles of *contextualism*, I develop and empirically test a contextual theory of the generation effect that flexibly accounts for variations in the generation effect by considering interactions among four key experimental factors: encoding task, materials, memory test, and subject (abilities). To test the theory, I compared the size of the generation effect across 16 different experimental contexts (representing variation in how these four experimental factors interact with one another). The results of this study confirm that the generation effect is strongly influenced by the experimental context under which it is studied, revealing that the generation effect (i.e., the memory benefit over reading) varied from a 10% to 20% increase for a free recall memory test and from a 22% to 36% increase for a recognition memory test across the different experimental contexts. Further, analyses showed a four-way interaction between each of the four factors included in the theory, suggesting that these factors uniquely influence the outcomes of generation effect studies. The overall pattern of results, however, did not fully align with the predictions of the contextual theory, highlighting that more research is needed to refine the theory. The data from this project can be used as a guide for future investigations on the interactions among the four key factors to inform future theoretical work and advance understanding of the generation effect.

I. INTRODUCTION

Since the beginning of the modern era of psychology, experiments have been a favorite methodology for generating new knowledge about human memory (Ebbinghaus, 1885). Early on, researchers following Ebbinghaus carried out experiments under the assumption that through careful and systematic experimentation, one could discover “truths” about memory, or laws that govern the fundamentals of memory function (e.g., Thordike's Law of Effect; Thorndike, 1927). Although most formal memory laws have disappeared (Roediger, 2008), today’s literature on human memory is now rife with various memory “effects” that describe generalities about memory such as: more meaningful (or semantic) processing at encoding leads to better memory compared to less meaningful (or non-semantic) processing (i.e., levels of processing effect; Craik & Lockhart, 1972), spaced learning leads to better retention than massed learning (i.e., spacing effect; Ebbinghaus, 1885; Hintzman, 1974; Rohrer & Pashler, 2007), testing improves memory for previously learned materials compared to restudying (i.e., testing effect; Roediger & Karpicke, 2006), self-generating information improves memory over reading (i.e., generation effect; Slamecka & Graf, 1978), among others. Although these memory effects provide a more comprehendible understanding of memory, many (if not all) of these effects are limited by several boundary conditions (i.e., circumstances when the effect does not occur). As research continues to accumulate, it is becoming clearer that human memory may be too complex to be described by universal memory laws or effects, but instead a host of factors must be considered to make an informed prediction about how memory will perform. Indeed, as Roediger (2008) put it in his review on the relativity of memory, “... the great truth of the first 120 years of the empirical study of human memory is captured in the phrase ‘it depends’” (p. 228).

A. <u>**Contextualism and Memory**</u>

The philosophy that the results of an experiment are a function of the experimental context in which it is studied is sometimes referred to as *contextualism* (see Hoffman, 1986, for a historical review). Although the ideas of contextualism have been expressed throughout the history of experimental research (Laudan, 1977; McKeachie, 1974), one of the pioneers in advancing these ideas in the study of memory was James J. Jenkins. In 1979, Jenkins authored a book chapter in which he claims that the results of all memory experiments are *context sensitive* (i.e., influenced by certain experimental factors). In this chapter he proposes a tetrahedral model of memory experiments (Jenkins, 1979), which designates that experiments are primarily made up of four key factors: the *orienting* (*encoding) task*, the *criterial task (memory test)*, the *materials (used in the task)*, and *(abilities of the) subjects*. In Jenkins' model, the encoding task refers to any feature of the experiment that directs how subjects first interact with to-be-learned information, the memory test refers to the way subjects are asked to retrieve information, materials refer to the type of information to be learned and how that information is presented, and the subjects factor refers to the abilities (or inabilities) of the subject to engage in processing dictated by the experiment. Jenkins' thesis was that these factors interact "vigorously" with one another to influence memory performance and that these interactions should be considered when interpreting the results of any experiment (Jenkins, 1979).

At the time Jenkins wrote this chapter in 1979, many researchers had just begun investigating interactions between *two* of the four key factors he identified (hereafter referred to as "lower-order interactions"). In the decades that have followed, researchers have made great strides in advancing knowledge on the lower-order interactions between these factors (e.g., the encoding task by memory test interaction has been studied most, but some work has also looked at encoding task by materials and encoding task by subject interactions). Some findings on these lower-order interactions have been regarded as some of the most influential developments in the study of human memory to date, such as the concepts of transfer-appropriate processing (hereafter referred to as TAP; Morris et al., 1977) and encoding specificity (Tulving & Thomson, 1973). TAP suggests that memory performance will be best when there is greater overlap between processing at encoding and retrieval (and encoding specificity suggests that memory performance depends on the similarity between the conditions at encoding and retrieval; Tulving & Thomson, 1973). These two memory principles represent classic examples of the interaction between the *encoding task* and *memory test* factors, but importantly represent only one of many interactions that should be considered according to Jenkins (1979). Yet surprisingly, little to no work has been done to extend these ideas toward investigating the higher-order interactions among these key factors (i.e., simultaneously considering the interactions between more than two of the four key factors). Jenkins proposed this idea more than 40 years ago, and yet little work has followed up on this important point. Indeed, it is still common practice for researchers to disregard certain experimental factors in order to investigate a lower-order interaction of interest. For example, to investigate TAP (i.e., how the encoding task and memory test interact), researchers often purposefully hold the materials and subject variables constant in order to reduce or eliminate their influence on memory. But how might the principles of TAP be affected by changes in materials, or different subject groups? These are the types of questions yet to be fully addressed by empirical work, but seem to represent a crucial step forward in generating new knowledge about human memory. Thus, this project aims to examine how considering the higher-order interactions among four key experimental factors can provide a step forward toward a more nuanced understanding of human memory.

In the spirit of Jenkins and the ideas of contextualism, in this project I develop and test a "contextual theory" of one of the aforementioned memory effects that guide our understanding of memory, the *generation effect*. Decades of research has been devoted to studying the generation effect (for reviews see Bertsch et al., 2007; McCurdy et al., In press; Mulligan & Lozito, 2004), but an impoverished focus on the interactions among the four key experimental factors described by Jenkins (encoding task, memory test, materials, subjects) has led to a wide-range of outcomes across studies, without a commonly agreed upon theory to account for the variability. Thus, the generation effect makes for a prime target for developing a contextual theory to predict variations in memory performance across different experimental contexts by considering the simultaneous interactions among four key experimental factors.

In the next few sections, I will first provide some background on the generation effect by describing prior theoretical work attempting to account for why the effect occurs. Then, I will review some research that has examined how the lower-order interactions (between two of the four factors) influence the size of the generation effect. The conclusions from these studies are important as they will serve as the foundation for the contextual theory I propose. After presenting this previous research, I will introduce the contextual theory and describe some testable predictions that can be drawn from this theory. Then, I conduct an empirical test of these predictions to examine how well this theory can account for the variations in the generation effect under different experimental contexts.

B. **The Generation Effect**

The generation effect is the commonly observed memory benefit for information that is self-generated compared to information that is read (Slamecka & Graf, 1978). In a typical generation effect study, subjects learn word pairs in either a *generate* condition where subjects self-generate a target word from a cue (e.g., OPEN – C____), or a *read* condition where subjects read an intact word pair (e.g., REPLY – ANSWER). The common finding is that memory is better for words that are self-generated compared to read (i.e., a generation effect; Jacoby, 1978; Slamecka & Graf, 1978). This effect is evident across a variety of experimental conditions (Bertsch et al., 2007; McCurdy et al., In press), however, certain features of the experimental procedure used have been shown to largely influence the relative size of the memory benefit from self-generation. For example, the generation effect is often larger in experiments where subjects are not aware of an impending memory test, relative to when they are aware of a memory test (Begg et al., 1991; Watkins & Sechler, 1988). Furthermore, several boundary conditions for the effect have been identified, where generation no longer improves memory compared to reading. For example, the generation effect is often eliminated in a between-subject design using a free recall memory test (Begg & Snider, 1987; Schmidt, 1990; Schmidt & Cherry, 1989; Slamecka & Katsaiti, 1987). These examples, among other well-known boundary conditions (e.g., generating non-meaningful information; Johns & Swanson, 1988; McElroy & Slamecka, 1982; Nairne & Widner, 1987), importantly highlight that the generation effect (as with many memory effects) is sensitive to the exact experimental conditions used in a given experiment, or *experimental context*.

1. Theories of the Generation Effect

Throughout the history of research on the generation effect, several theories have been proposed to account for how self-generation improves memory over reading. The leading theory suggests that generation leads to better memory than reading because it enhances two types of encoding processes: item-specific and relational processing (i.e., the multi-factor theory; Hirshman & Bjork, 1988; McDaniel et al., 1988). Item-specific processing refers to encoding of the properties that are unique to the to-be-learned item (i.e., that distinguish that item from other studied items). In contrast, relational processing refers to the encoding of the relationship of the to-be-learned item to either a cue word (in paired associate learning; i.e., cue-target relational processing, Hirshman & Bjork, 1988), or of the relationship of that item to other items (i.e., target-target relational processing, McDaniel et al., 1990; McDaniel et al., 1988).

A recent meta-analysis on the generation effect assessed the empirical support for several extant theories (McCurdy et al., In press). This study found that although some theories, including the multi-factor theory, were supported to some degree, the existing theories could not fully account for the variation in the generation effect, suggesting that more comprehensive theories are needed. For instance, although it is clearly important to consider the processes engaged at encoding (e.g., the multi-factor theory), this factor alone is not the only component worth considering. As Jenkins' (1979) model suggests, it is important to consider how these processes at encoding interact with other factors. Indeed, the meta-analysis described above also identified several key moderators of the generation effect (e.g., generation task type, memory test, stimuli type; McCurdy et al., In press), many of which could be binned into one of the four key factors identified in Jenkins' (1979) model. Therefore, to attempt to account for more variance in the size of the generation effect across different experimental contexts, the theory I develop simultaneously considers the interaction among four factors (encoding task, memory test, materials, subjects). In line with other theoretical work that suggests simultaneously considering multiple factors often provides a more complete account than considering these factors in isolation (Healey & Kahana, 2016), a theory of the generation effect that accounts for four factors (as opposed to two) should account for more of the variance of the generation effect and more accurately predict when generation will lead to higher or lower memory performance.

2. Encoding (Generation) Task Processing and Generation Constraint

Although the focus of this project is on interactions, given that the generation effect primarily refers to a manipulation of the encoding task factor it may be useful to consider this factor in isolation before discussing how the encoding task interacts with other three experimental factors to be included in the contextual theory. Specifically, it is important to consider the processing induced by a generation task compared to reading, and the features of the generation task that can influence the type of processing induced by the task at encoding. As reviewed earlier, generation tasks are thought to induce both item-specific and relational processing at encoding (Hirshman & Bjork, 1988). The contextual theory I develop in this project will mainly focus on the distinction between these two types of encoding processes.[1] In order to understand how different encoding tasks might induce different encoding processes (item-specific, relational), in the next section I will discuss a factor known as generation constraint because this factor shows how different generation tasks can lead to more or less item-specific and relational processing, respectively.

One limitation of the multi-factor theory is that it broadly lumps all generation tasks together in terms of memory improvement that derives from item-specific and relational processing. More recent research, however, has shown that not all generation tasks are equal, and some features of generation tasks may influence the relative amount of item-specific or relational processing, which in turn can influence the magnitude of the generation effect. Specifically, one feature within the encoding (generation) task factor that has been shown to influence the magnitude of the generation effect is *generation constraint* (Fiedler et al., 1992; Gardiner et al.,

---

[1] Despite focusing on the distinction between item-specific and relational processing, the theory can also accommodate other types of task processing (e.g., conceptual, perceptual, phonemic, etc.; Jacoby, 1983; Payne et al., 1986), but for the purposes of this empirical work I will only focus on item-specific and relational processing.

1985; McCurdy et al., 2017, 2019; McCurdy et al., In press). Generation constraint can be thought of as the amount of information limiting what a subject can self-generate. A variety of generation tasks have been used in prior studies on the generation effect, with some tasks placing more or less constraints on what can be self-generated compared to other tasks. For example, a lower-constraint generation task might provide a cue word and ask subjects to generate an associated target word (e.g., generating semantic associates; OPEN – _____). In contrast, a higher-constraint generation task provides more information that limits (or constrains) how subjects generate the target item (e.g., solve the anagram: OPEN – COSEL). Importantly, prior work has shown that generation tasks with fewer constraints often lead to differential memory benefits compared to generation tasks with higher constraints (Fiedler et al., 1992; Gardiner et al., 1985; McCurdy et al., 2017, 2019).

The influence of generation constraint on memory performance is also subject to experimental context, however, where the memory effects for generation constraint differ depending on the type of memory test used. Prior work on generation constraint has found that the increased memory for lower-compared to higher-constraint generation is often limited to recall memory tests (and does not seem to influence item recognition), suggesting that this difference in memory may be tied to differences in processing between a lower- and higher-constraint generation task (McCurdy et al., 2019; McCurdy et al., 2020; McCurdy et al., In press). Given that prior work has shown that recognition and recall tests are sensitive to item-specific and relational processing respectively (Burns, 2006), these findings suggest that differences in generation constraint may influence the relative amount of item-specific versus relational processing induced. Specifically, more recent work has shown evidence that lower-constraint tasks lead to greater relational processing, whereas higher-constraint tasks lead to greater item-specific processing (McCurdy et al., 2019; McCurdy et al., 2020; McCurdy et al., In press).

Overall, what these findings suggest is that a successful theory of the generation effect must take into account differences between generation tasks (and the type of processing these tasks induce). Thus, the contextual theory developed in this project accounts for the type of processing induced by the task (i.e., the exact generation task used at encoding), and how that processing interacts with other experimental factors.

Next, I turn to a review of prior work examining the lower-order interactions between the encoding task and the other three experimental factors. In each section, I will describe the interaction, provide some examples from prior work that demonstrates the importance of this interaction, and then provide a "conclusion" that I will return to later in a discussion on higher-order interactions as it pertains to developing the contextual theory of the generation effect.

C. Lower-Order Interactions

1. Encoding Task by Memory Test Interaction – "Transfer-Appropriate Processing"

Of the various interactions I will pitch in this project (between encoding task, memory test, materials, and subjects), the interaction that has arguably been the most widely studied is between the encoding task and memory test. The memory principle of transfer-appropriate processing (TAP; Morris et al., 1977), was developed to describe how this interaction influences memory. Specifically, TAP is the idea that memory performance will be determined by the extent the processing engaged during encoding *matches* (or overlaps with) the processing required by the memory test. One way to assess TAP, is through the use of different memory tests for a given encoding task. Prior work has shown that different types of memory tests are sensitive to differences in processing induced at encoding (Hirshman & Bjork, 1988; Jacoby, 1983; Rabinowitz & Craik, 1986) For example, item recognition tests are thought to be sensitive to item-specific processing, whereas free recall tests are sensitive to relational processing (Burns, 2006). Thus, according to TAP, an encoding task where there is more processing overlap between the task and the memory test will yield better memory than an encoding task where there is less processing overlap between the task and the memory test.

As with many memory phenomena, TAP has been proposed to account for variation in the magnitude of the generation effect. Prior empirical work has shown that generation is often more effective in improving memory compared to read (control) tasks because there is a greater match in processing between generating and most memory tests (e.g., recognition, cued recall, free recall; deWinstanley & Bjork, 1997; Mulligan, 2011; Steffens & Erdfelder, 1998). One demonstration of the importance of this interaction to the generation effect comes from deWinstanley and Bjork (1997). In their study, they manipulated the instructions given at the time of encoding to focus on one type of processing or another (specifically, cue-target or target-target relational). At retrieval, subjects were given either a cued recall test (sensitive to cue-target processing), or a free recall test (sensitive to target-target processing). They found that the generation effect depended on both the encoding task instructions and the type of memory test used. Specifically, they found that instructions that focus on cue-target relations led to a larger generation effect on a cued recall test, relative to a free recall test. In contrast, instructions to focus on relations among target items during generation led to a larger generation effect on a free recall test, compared to a cued recall test (deWinstanley & Bjork, 1997). Although the empirical work I report below does not focus on cue-target versus target-target relational processing, this example from deWinstanley and Bjork (1997) demonstrates the importance of the interaction between encoding task and memory test, suggesting this interaction should be considered in a theory attempting to account for the generation effect.

Many demonstrations of this interaction between encoding task and memory test have been shown throughout the generation effect literature (McNamara & Healy, 1995; Mulligan, 2004, 2011; Mulligan & Lozito, 2006; Nieznański, 2012, 2014; Steffens & Erdfelder, 1998). Together, these studies have shown that the relative size of the generation effect is dependent on the amount of overlap between the encoding task processing and the processing required to perform well on the memory test given. As mentioned earlier, for each interaction I will make a conclusion based on prior findings that I will return to in a later section on higher-order interactions. The conclusion for the encoding task by memory test (TAP) interaction is: **A generation task will increase memory performance to the extent it induces processing that matches the processing the memory test is sensitive to**.

In this study, I will test the encoding task by memory test interaction by examining memory performance for a lower-constraint generation (relational) task and a higher-constraint generation (item-specific) task, using both a free recall test (relational), and item recognition test (item-specific). TAP suggests that the lower-constraint (relational) task should provide best memory on the free recall test because relational task and free recall tests represent a match in the type of processing required, while the higher-constraint (item-specific) task should provide best memory on the item recognition test because item-specific tasks and recognition tests represent a match in processing.

2. **Encoding Task / Materials Interaction – "Material-Appropriate Processing"**

Although it has received arguably less attention than the interaction between encoding task and memory test, some work has also investigated the interaction between the encoding task and materials (Einstein & Hunt, 1980; Einstein et al., 1990; Hunt & Einstein, 1981; McDaniel & Einstein, 1989; McDaniel et al., 1986). This work has led to a principle that McDaniel and Einstein (1989) dubbed "material-appropriate processing", hereafter referred to as MAP. In contrast to the principle of TAP described earlier, MAP suggests that memory is best when the processing induced by the encoding task is *different* from (or complementary to) the processing induced by the materials alone. In order to understand MAP, it might be helpful to give an illustrative empirical example. In their study, McDaniel et al. (1986) used two different types of materials and two different types of generation tasks, each designed to bias processing of different information. For their materials, they used a descriptive text (i.e., an excerpt from a textbook) which has been shown to invoke item-specific processing, and a fairy tale text which has been shown to invoke relational processing (Graf & Levy, 1984; Kintsch & Young, 1984). Prior work has shown that the causal chains in fairy tale texts are more clearly defined in narrative texts (e.g., fairy tales) which affords greater relational processing compared to descriptive texts, (Zelinski & Gilewski, 1988). In contrast, other work has shown that descriptive texts lead to greater encoding of the individual words in the text, affording greater item-specific processing relative to narrative texts (Graf & Levy, 1984). For the generation tasks, they had some subjects fill in letters of the text that were missing, a task designed to induce item-specific processing (i.e., emphasized the individual words of the text), while other subjects arranged scrambled sentences into their correct order, a task designed to induce relational processing (i.e., emphasized the relation among the words of the text). The results of the study showed that an item-specific generation task (e.g. missing letters) led to better memory for relational materials (e.g., fairy tale text), while the relational generation task (e.g., sentence reordering) led to better memory for item-specific materials (e.g., descriptive text). In other words, the generation task that induced processing that was *different* from (or complementary to) the processing induced by the materials led to the best memory performance (McDaniel et al., 1986).

The principles of MAP have also been demonstrated in word-list learning (Einstein & Hunt, 1980; Hunt & Einstein, 1981). Specifically, Einstein and Hunt (1980) showed that categorically related word lists naturally induce relational processing, whereas unrelated word lists tend to induce more item-specific processing. In this study, they showed that an encoding task that provided different (or complementary) processing from the materials led to better memory performance. Specifically, an item-specific encoding task led to better memory on a related word list, while a relational encoding task led to better memory on an unrelated list, a finding that has been replicated in later studies as well (Einstein & Hunt, 1980; Hunt & Einstein, 1981; Hunt & Seta, 1984). Theoretical work has suggested that a match in processing between encoding task and materials leads to redundant processing of the same information, affording weaker memory benefits than a mismatch where multiple types of information are processed together (McDaniel & Einstein, 1989; McDaniel et al., 1986). More recent work has shown that simultaneous encoding of both item-specific and relational information leads to better memory through *distinctiveness*, or the combined effect of processing similarity and difference (Hunt, 2006, 2012; Hunt & McDaniel, 1993). Specifically, it is thought that relational processing promotes the encoding of the similarity (or shared features; e.g., category membership) among items, while item-specific processing promotes the encoding of differences (or the unique features) that distinguish one item from another. At retrieval, both types of information are useful to make accurate memory judgments. Relational information is used to broadly define the episode to be remembered (i.e., brings to mind a group of items that share some feature, e.g., category membership), and item-specific information is used to distinguish the target item from other similar items encoded in that episode (Hunt, 2006, 2012; Hunt & McDaniel, 1993). Thus, distinctiveness can explain why experimental contexts where the encoding task and materials provide different (or complementary) processing lead to increased memory performance relative to contexts where there is a match in processing between the encoding task and materials.

Based on the evidence reviewed on MAP, the conclusion for this interaction is: **A generation task will increase memory performance to the extent it requires processing that is different from (or complementary to) the processing induced by the materials used**.

To test the encoding task by materials interaction, I will use two generation tasks and two types of materials that tap into item-specific and relational processing, respectively. Specifically, I will test memory performance for a lower-constraint generation (relational) task, a higher-constraint generation task (item-specific), with a related list structure (relational) and an unrelated list structure (item-specific). In the related list structure, multiple words belonging to a single category will be presented in succession, whereas in the unrelated list, words from the same category will be presented intermixed with words from other categories throughout the list. MAP suggests that the lower-constraint generation task should provide better memory with an unrelated list structure, whereas the higher-constraint generation task will provide better memory with a related list structure.

3. Encoding Task/Subject Interaction – "Subject-Appropriate Processing"

Although it may often be assumed, it is also important to consider the extent a subject can engage in the processing required by a given encoding task. This factor (perhaps more than the others) is typically ignored as many studies utilize a convenience sample of college-age students. Yet, in the cases where different subject populations are used, it is often found that the abilities of the subjects being studied can interact with the processing required by the encoding task to influence memory performance.

One example of the importance of the encoding task by subject ability interaction comes from research showing that differences in subjects' prior knowledge can influence how one engages with an encoding task, thus leading to differences in memory (Reardon et al., 1987). Specifically, Reardon and colleagues (1987) examined the generation effect for "experts" versus "novices" in the domain of the materials used. Psychology faculty members (experts) and college-level psychology students (novices) were given factual sentences about various psychological concepts. For half of the concepts, subjects generated the keyword of the concept, and for the other half, subjects read the keyword. Experts showed better memory for generated concepts compared to read concepts, but novices showed no differences between the two conditions (i.e., a generation effect for experts, but not novices; Reardon et al., 1987). The authors suggested that because of their increased background knowledge, generation led to more elaborative encoding for experts compared to novices who likely had a smaller network of knowledge in this domain. In other words, the subjects' prior knowledge influenced their ability to successfully engage in the generation task. This study highlights how differences in subjects' abilities (e.g., expertise) can determine the effectiveness of a generation task (see also Kalyuga, 2009, for a review on the influence of expertise on various encoding strategies).

Overall, prior work considering the interaction between encoding task and subjects provides the conclusion: **A generation task will increase memory performance to the extent it requires processing that the subject is able to successfully engage**. In conforming to the conventions adopted in describing the other two interactions reviewed earlier, I refer to this (encoding task by subjects) interaction as "subject-appropriate processing", hereafter referred to as SAP.

To test this interaction, I will examine memory performance for two groups of subjects that differ in their ability to process English words, and thus their ability to engage in the type of processing provided by the two generation tasks. Specifically, one group of subjects will be low-ability and another group will be high-ability based on an English vocabulary assessment.

D. **Higher-Order Interactions / Predictions of the Theory**

In the previous sections, I have reviewed prior work that has considered some lower-order (two-way) interactions between the encoding task, and the memory test, materials, and subject abilities individually. However, in order to more fully account for the variability in generation effect studies, my contextual theory considers higher-order interactions (i.e., interactions among more than 2 of the factors) to make predictions about memory performance. Thus, to develop this contextual theory I aim to use what we know from prior work about the lower-order interactions among these experimental factors, in order to make predictions about the higher-order interactions between the four experimental factors.

Before introducing the contextual theory, I will briefly revisit the primary "conclusions" from prior work on the lower-order interactions reviewed earlier. TAP suggests that an encoding task that induces processing that *matches* the processing required by the memory test will lead to increased memory performance. In contrast, MAP suggests that an encoding task that induces processing that is *different* from the processing invoked by the materials themselves will lead to increased memory performance. And finally, SAP suggests that encoding tasks that induce processing that *matches* the subjects' abilities will lead to increased memory performance. These conclusions will serve as the foundation of the contextual theory of the generation effect.

The contextual theory of the generation effect makes the prediction that the effects of these lower-order interactions (TAP, MAP, SAP) are additive, meaning that more "appropriate" processing will lead to better memory performance. Specifically, this theory predicts that an experimental context where TAP, MAP, and SAP are all "satisfied" (i.e., the encoding task induced processing that was appropriate given the memory test, materials, and subjects used), will lead to better memory performance compared to an experimental context where TAP, MAP, and SAP are all *not* "satisfied". In addition, this theory suggests that when some (but not all or none) of these interactions are satisfied, memory performance should fall somewhere in between these two extreme experimental contexts. Less work has compared the relative importance of each interaction (TAP, MAP, SAP) on memory performance. Thus, although the predictions of this contextual theory are based on the assumption that the relative contributions of these three lower-order interactions are equal, an empirical test of this theory should provide insight into whether one type of interaction may be more influential to memory than another.

In this project, I designed a study to test the predictions of the contextual theory of the generation effect described above. Specifically, I compare memory performance from two generation tasks that vary in the amount of constraint they provide (lower-constraint, higher-constraint) across various different experimental contexts to examine the ability of the proposed contextual theory to predict memory performance given the type of generation task, memory test, materials, and subjects. In addition, I also examine the relative contributions of the three lower-order interactions (TAP, MAP, SAP) to determine whether one type of interaction may be more influential to memory performance than another.

E. **Specific Aims and Hypotheses**

Table I shows an orthogonal list of the eight possible experimental contexts that correspond to the interactions between the encoding task and the other three factors (memory test, materials, subjects). In this table, a check mark (✓) indicates a satisfied lower-order interaction, and an “x” (☒) represents an unsatisfied lower-order interaction. This table represents a conceptual model for how this contextual theory takes into account higher-order interactions to predict memory performance (see **Table I**). To demonstrate how I will test this theory, I have taken the same structure of **Table I**, and layered on specific experimental details that correspond to the experimental manipulations I will use in this study. Specifically, **Tables II** and **III** show the predicted memory performance for a free recall test using a relational task (lower-constraint generation; **Table II**) and an item-specific task (higher-constraint generation; **Table III**), respectively. Similarly, **Tables IV** and **V** show the predicted memory performance for an item recognition memory test using a relational task (lower-constraint generation; **Table IV**) and an item-specific task (higher-constraint generation; **Table V**), respectively. Collecting data across two different types of encoding tasks reduces the likelihood that the results of this study are specific to one type of generation task (i.e., encoding processing). Indeed, one of the advantages of this contextual theory over existing generation effect theories is that it can account for differences between various generation tasks. Using both a lower-constraint generation task (relational processing), and a higher-constraint generation task (item-specific processing), I aim to test three hypotheses in this study:

Hypothesis 1: The generation effect (generate versus read) will be larger in experimental contexts where all interactions (TAP, MAP, SAP) are simultaneously satisfied compared to experimental contexts where no interactions are satisfied.

Hypothesis 2: The effects of transfer-appropriate, material-appropriate, and subject-appropriate processing are additive, such that as the number of "appropriate" processing types increase, the generation effect will increase (i.e., a positive relationship between the number of "appropriate" processing types satisfied and the size of the generation effect).

Hypothesis 3: The relative influence of transfer-appropriate, material-appropriate, and subject-appropriate processing is equal.

An alternative possibility to Hypothesis 3 is that one type of "appropriate" processing may be stronger than that of another type of "appropriate" processing, which is an idea that has not receive much consideration in the extant memory research. For example, it could be that the match in processing between encoding task and subject ability (SAP) is stronger than the match in processing between encoding task and memory test (TAP), potentially overpowering the effects of TAP in this scenario. Although the contextual theory predicts no differences in memory performance between "appropriate" processing types (TAP, MAP, or SAP), the prospect that one type of processing is more influential to memory performance than another is plausible, and therefore, testable. In the case that Hypothesis 3 is falsified, the results of this study would suggest that one type of "appropriate" processing may be more influential to memory performance than another, which would be an important advance in the field of learning and memory. Additionally, given that the predictions of the theory are based on whether or not the lower-order interactions are "satisfied", examining support for these three lower-order interaction across the different experimental contexts included in this study represents a concept check to make sure our experimental approach can truly assess the contextual theory.

II. METHOD

A. Subjects

Three hundred and sixty-four adults participated in this study. All subjects were recruited through the University of Illinois at Chicago introductory psychology course subject pool. Subjects gave written informed consent in accordance with the University of Illinois at Chicago Institutional Review Board and were compensated with course credit. Forty-two subjects were removed from the analyses[2], leaving a total N of 322 subjects (Age: M = 18.96, SD = 1.80, range: 17 – 33; 216 female) included in the reported analyses.

B. Stimuli

A total of 180 highly associated cue-target word pairs were used as stimuli. First, a total of 144 target words were selected from twelve categories from the Van Overschelde et al. (2004) norms. After selecting twelve target words from each of the twelve categories, a unique cue word was selected using the University of South Florida Free Association Norms (Nelson et al., 2004). Cue words were selected so that both the cue and target were highly related (Forward Association Strength: M = .35, SD = .24), and were not related to other target members of the category. The list of word pairs was counterbalanced so that each word pair occurred exactly once in each encoding task (generate, scramble, read), and as a "related" distractor item for each of the three tasks during the recognition test (i.e., 6 counterbalanced word lists in total). An additional 46 cue-target word pairs were selected from the University of South Florida Free Association Norms (Nelson et al., 2004). Ten of these word pairs served as examples in the

[2] The number of subjects removed (and percent of total sample) by removal criteria:
30 (8.2%) – did not follow instructions in the encoding phase
2 (0.5%) – did not complete the vocabulary/working memory phase
4 (1.1%) – generated greater than 75% idiosyncratic responses in the "lower-constraint" generation task
1 (0.3%) – 100% false alarm rate
5 (1.4%) – did not complete study/programming error

instructions/practice phase, and 36 of these words served as "unrelated" distractor items for the recognition memory test. These word pairs were selected such that they did not belong to any of the twelve categories used in the word lists, and were not related to any of the other word pairs. These ten example items and 36 unrelated distractor items were not counterbalanced (i.e., each subject was given the same ten examples, and the same 36 "unrelated" distractor items on the recognition test).

C. **Procedure**

Subjects completed eight phases in total throughout the experiment: Instructions/Practice, Encoding, Math Filler, Free Recall, Recognition, Vocabulary Assessments, Working Memory Assessment, Demographics/Language History Questionnaire. All phases of the experiment were presented on a PC monitor using E-Prime presentation software (Version 2.0.10; Psychology Software Tools, 2012), on a black background with white text (18-point, Times New Roman font). After the completion of each phase, subjects were given a 15-second break where they were instructed to rest their eyes. After 15 seconds, the program advanced to the next phase of the experiment. The total amount of time to complete all phases of the experiment was approximately one hour.

1. **Instructions / Practice**

Subjects were first given instructions and one example on how to respond to the word pairs during each of the encoding tasks (generate, scramble, read). For all encoding tasks, subjects were instructed to type both the cue and target word to ensure subjects attended both words across all encoding tasks. In the lower-constraint "generate" task (i.e., relational processing task), subjects were shown a cue word and the first letter of the associated target word followed by a blank line (e.g., antler – d_____) and were instructed to type the *cue* word, followed by the ENTER key to submit their response, and then type a word that was related to the cue word and began with the given letter (i.e., type the *target* word), followed by the ENTER key to submit their response. Responses during this phase were recorded by the computer program, and saved for use in the recognition memory test.[3] In the higher-constraint "scramble" task (i.e., item-specific processing task), subjects were shown a cue word followed by a scrambled target word (e.g., hospital – nrues) and were instructed to type the cue word, followed by the ENTER key to submit their response, and then to type the unscrambled target word, followed by the ENTER key. If the subject was unable to determine the scrambled word, they were instructed to leave the response blank, and that trial was later removed from all analyses. The first letter of the scrambled target word was always in its correct position to reduce the number of skipped trials,[4] as done previously (Foley & Foley, 2007; Foley et al., 1989; McCurdy et al., 2017, 2019). In the "read" task subjects were shown a cue and target word (e.g., dunk – basketball) and were instructed to type the cue word, followed by the ENTER key, and then to type the target word as shown on the screen, followed by the ENTER key. Once the subject typed both words for a given trial and pressed the ENTER key to submit their responses, the screen advanced and the subject was presented with a new word pair after a 500ms fixation. After the instructions, subjects were trained on a shortened version of the encoding phase with three practice trials for each task (nine practice trials in total). The practice word pairs were selected to be unrelated to each other (i.e., not related to any of the other practice items) and unrelated to the stimuli used in the experiment. Subjects were not instructed, trained on, or

[3] This procedure importantly allowed us to use the subject-generated *target* responses from the encoding phase on the recognition test, even if the response did not match the normed, expected target word.

[4] Pilot testing in prior work showed that using a scramble task where all letters of the target word were scrambled (hospital – surne) led to significantly more "skipped" trials than a scramble task where the first letter was always in the correct place but the remaining letters scrambled (hospital – nrues).

otherwise made aware of the free recall or recognition phases before the encoding task (i.e., subjects were not told there would be a memory test for these words).

2. **Encoding**

In the encoding phase, subjects saw a total of 72 word pairs across three encoding tasks: generate, scramble, and read (24 pairs in each task). Encoding trials were presented in six blocks of twelve trials each, and all trials within a block were of the same task (generate, scramble, or read). The order of the blocks was pseudorandom such that each task must have been presented exactly once before the second block of a task was presented. Before each block, an instruction prompt appeared for 3000ms indicating the next task (e.g., "Get ready to do the generate/scramble/read task.").

The organization of the trials within each encoding block (i.e., word list structure) was manipulated to induce relatively greater relational or item-specific processing. Subjects were randomly assigned to encode words in either a *related* list structure or an *unrelated* list structure. In the related list structure, the order of the trials within a block was arranged so that target words from the *same* category were presented in consecutive order (e.g., six words from the ANIMALS category were presented together, before moving on to six words from the category SPORTS presented together, making up a single block of twelve items total). **Figure 1** shows a schematic of a block of twelve trials for each encoding task in the related list structure. Alternatively, in the unrelated list structure, the order of trials within a block was arranged so that target words from *different* categories were presented consecutively (e.g., a word presented form the ANIMALS category, followed by a word from the SPORTS category, and so on for the rest of the twelve trials in the block). **Figure 2** shows a schematic of a block of twelve trials for each encoding task in the unrelated list structure. Prior work has shown that related word list structures induce greater relational processing whereas unrelated word list structures tend to induce greater item-specific processing (Einstein & Hunt, 1980; Hunt & Einstein, 1981), and this particular implementation has been used in prior work on the generation effect to manipulate relational versus item-specific processing between list structures (deWinstanley et al., 1996).

3. **Math Filler**

Following the encoding phase, subjects solved simple multiplication problems (e.g., 4 x 5 = ??) for exactly one minute. Subjects were shown a multiplication problem, and were instructed to type the answer and press the ENTER to submit their response. The filler task was designed to reduce any potential primacy and recency effects, and was non-verbal to limit any potential interference of the word pairs learned in the encoding phase. These data were not analyzed in this study.

4. **Free Recall**

In the free recall phase, subjects were given exactly five minutes to type as many target words from the experiment as they could remember. Subjects were instructed that once they had typed a single response (i.e., target word), to press the ENTER key to submit their response. After submitting their response, that response was added to a list displayed at the top of the screen. Thus, subjects were able to see the responses they had entered, but could not edit them, a procedure used previously (Huff & Bodner, 2014). The free recall test served as the *relational* memory test, as prior work has shown that free recall tests are primarily sensitive to relational information, especially of the relationships between target words in a list (Burns, 2006; Hunt & Einstein, 1981; McDaniel et al., 1988).

5. **Recognition**

In the recognition phase, subjects were shown all 72 target words from encoding (i.e., the words they generated, unscrambled, or read) and 108 "new" distractor words, one at a time, in a random order, for a total of 180 recognition trials. Of the 108 distractor words, 72 were "related" distractors, that came from the same twelve categories the target words came from. Thus, there were six "related" distractor words from each of the twelve categories of the target words at encoding. This importantly allows a unique false alarm rate to be calculated for each encoding task (generate, scramble, read), with 24 "old" target words (6 per category), and 24 "new" related distractors (six per category), corresponding to each of the three tasks. For example, if a subject encoded six target words from SPORTS category in the generate task, at recognition that subject would see the six target words from encoding, and six distractor words from the SPORTS category, and so on for the other three categories corresponding with that encoding task. The remaining 36 distractor items were "unrelated" distractors that were not related to any of the target words and "related" distractors, as described in the **Stimuli** section. For each recognition trial, an "old" target or "new" distractor word was presented on the screen, and subjects made a forced-choice judgment of whether the word was "old" (previously encoded) or "new" (a distractor item), using their index or middle finger, respectively, of their right hand. Each trial was separated by a 500ms fixation. The recognition test served as the *item-specific* memory test, as prior work has shown that recognition tests are primarily sensitive to item-specific information (Burns, 2006).

6. **Vocabulary Assessments**

Following the recognition test, subjects were given computerized versions of two vocabulary assessments (Raney; Shipley). A computerized version of the Raney Vocabulary Assessment (**Appendix A**) was presented first, followed by a computerized version of the

Shipley Vocabulary Test (**Appendix B**; Shipley, 1946). The maximum score on the Raney Vocabulary Assessment is 30, and the maximum score on the Shipley's Vocabulary test is 40. To assess subject vocabulary ability, the combined total score of the two measures was calculated for each subject (maximum score of 70). This score for each subject ranging from 0-70 was used to sort subjects into the "low-ability" (lower tertile) and "high-ability" (upper tertile) groups for analysis. A cutoff score was calculated for the lower tertile (i.e., the bottom third), and for the upper tertile (i.e., the upper third) for each list structure group separately (this was to ensure a similar sample size of low and high ability subjects across the two types of materials). Subjects scoring in the lower tertile were classified into the "low vocabulary ability" group, whereas subjects scoring the upper tertile range were classified into the "high vocabulary ability" group. Subjects scoring in the middle tertile (i.e., the middle third) were not analyzed in this study.

7. **Working Memory Assessment**

For the working memory assessment, subjects were given an auditory, computerized version of a backward digit span assessment (Woods et al., 2011). Subjects were given noise-canceling headphones and instructions for the task were presented on the screen before subjects began the assessment. Digitally recorded spoken digits (1-9) were obtained from MacWhinney et al. (2001). Spoken digits were delivered binaurally through the headphones at a pace of one digit per second. Using the procedure described in Woods et al. (2011), all subjects started with a span length of 3 digits and completed a total of exactly fourteen trials with the span length of each trial adjusted based on subject performance. Specifically, a single correct response increased the span length by one digit (e.g., from three digits to four digits), while two consecutive incorrect responses decreased the span length by one digit (e.g., from five digits to four digits). Subject's working memory was evaluated using a mean span scoring procedure as described in Woods et al. (2011). Mean working memory span was calculated as the expected list length where half of the trials would be successful (Killion et al., 2004): Mean span baseline was set at 0.5 digits less than the initial list length (i.e., 2.5 digits) and was incremented by the proportion of digit strings accurately reported at each succeeding list length. For example, adding the hit rate for each list length (e.g., 3 = 1.0, 4 = 1.0, 5 = 1.0, 6 = 1.0, 7 = 0.25, 8 = 0.33 and 9 = 0.0, sum = 4.58) to the baseline value of 2.5, would represent a mean span score of 7.08. This scoring procedure has been shown to provide more reliable and precise estimates of working memory, compared to traditional procedures and scoring methods (i.e., longest list length prior to two errors on a given length; Woods et al., 2011).

8. **Demographics / Language History Questionnaire**

In the final phase of the experiment, subjects were given computerized versions of a demographics (**Appendix C**) and language history (**Appendix D**) questionnaire. After completing the questionnaires, subjects were debriefed and left the laboratory. Responses to these questionnaires were used to evaluate potential performance differences based on age, gender, education history, and fluency of speaking, reading, and understanding the English language. Sample demographics are reported in **Table VI**.

III. RESULTS

A. Results Overview

In this section, I present the results of this study in the following order: First, I describe how participants were grouped into high- and low-ability based on their English vocabulary assessment scores (see **Vocabulary Measures**). These results represent the “subject ability” factor of the contextual theory, and thus are a critical prerequisite for the proposed analyses. Second, I describe the procedure used to calculate a generation effect score under each of the four within-subject conditions (see **Generation Effect Measures**). These generation effect scores were the primary dependent variable of interest and were used in the proposed analyses to test the main hypotheses of this study. Third, I report the results of the proposed analyses that assess the main hypotheses of this study that test the predictions of the contextual theory (see **Memory Performance**). Fourth, because the predictions of the contextual theory rely on understanding the encoding processing induced by each factor (and how these processes interact), I present data on independent measures of item-specific/relational processing to more fully investigate the contextual theory (see **Independent Indices of Item-Specific and Relational Processing**). These analyses provide a more nuanced understanding of the memory performance data and highlight important implications for future research aimed at refining the contextual theory.

B. Vocabulary Measures

In order to group subjects into high- and low-ability groups, the summed score of the two vocabulary assessments (Shipley Vocabulary + Raney Vocabulary) from each subject was used to calculate tertile cutoff scores (33.3%; 66.7%) within the related and unrelated list structure groups, separately. The lower and upper tertile cutoff scores were 37 and 42 (maximum possible score of 70) for the related list structure subjects, and 37 and 43 for the unrelated list structure subjects. Subjects scoring *within* these bounds (i.e., scoring between 37 – 43; 37 – 42) were not analyzed in this study, leaving a total N of 228 subjects included (Age: $M = 19.07$, $SD = 2.02$, range = 17 – 33; 156 females). The subjects included were organized into one of four groups corresponding to the two between-subject conditions (related list structure/high English vocab, $n = 61$; related list structure/low English vocab, $n = 55$; unrelated list structure/high English vocab, $n = 59$; unrelated list structure/low English vocab, $n = 53$). The means and standard deviations of vocabulary scores within each of the materials and subject groups are shown in **Table VII**. For both the related and unrelated list structures, the high and low English vocabulary groups were significantly different in their English language vocabulary mean scores, t's > 16.59, p's < .001, as expected.

C. **Generation Effect Measures**

A total of four generation effect scores were calculated for each subject corresponding to the four within-subject conditions (1. lower-constraint task/free recall memory test; 2. higher-constraint task/free recall memory test; 3. lower-constraint task/recognition memory test; 4. higher-constraint task/recognition memory test). For free recall, a generation effect score for each generation task (lower- and higher-constraint) was calculated as the proportion of generated items studied at encoding (24 per task) correctly recalled *minus* the proportion of read items studied at encoding (24) that were correctly recalled (i.e., [percent generate recalled – percent read recalled] = free recall generation effect score). For recognition, a generation effect score was calculated using false alarm-corrected memory performance. Each encoding task produced a hit rate (proportion of encoded items correctly identified as “old”), and a false alarm rate (proportion of “related” distractors incorrectly identified as “old”). For each of the three tasks, false alarm-corrected memory performance was calculated as the hit rate *minus* the false alarm rate. Then, a generation effect score was calculated for each generation task by taking the false alarm-corrected memory for the generate task *minus* the false alarm-corrected memory for the read task (i.e., [Generate hits – Generate false alarms] – [Read hits – Read false alarms] = recognition generation effect score).[5]

D. Memory Performance

In this section, I report several analyses to test the three primary hypotheses of this project. First, to assess Hypothesis 1, I conducted four separate independent samples t-tests corresponding to the four generation effect measures represented in **Tables II** through **V** (lower-constraint task/free recall memory test; higher-constraint task/free recall memory test; lower-constraint task/recognition memory test; higher-constraint task/recognition memory test) to examine memory differences between the experimental context with the maximum number of lower-order interactions satisfied compared to the experimental context with the minimum number of lower-order interactions satisfied. To assess Hypothesis 2, I conducted four regression analyses using each of the four generation effect measures as the dependent variable, and including the number of lower-order interactions satisfied as the predictor variable, to investigate whether increasing the number of lower-order interactions satisfied predicted an increase in the size of the generation effect (i.e., a positive relationship). Lastly, to assess Hypothesis 3, I conducted a four-way (2 x 2 x 2 x 2) mixed analysis of variance (ANOVA) with generation task (lower-constraint; higher-constraint) and memory test (free recall; recognition) as within-subject factors, and materials (related list structure; unrelated list structure) and subject ability (high English vocab; low English vocab) as between-subject factors. This analysis was designed to

[5] Raw response rates for the recognition memory test by encoding task and subject ability groups are reported in **Table VIII**.

assess the relative importance of each type of “appropriate” processing (TAP, MAP, SAP). Specifically, examining the interactions between the generation task variable and the memory test (TAP), materials (MAP), and subject ability (SAP), provides evidence about the relative influence of each interaction on the size of the generation effect. Further, this analysis represents a manipulation check to assess whether the design used is sufficient to fully assess the contextual theory (e.g., whether each of the lower order interactions were satisfied, respectively). The results of each of these analyses are reported below, separated by hypothesis. For each, I review the expected outcome before presenting the findings of the analysis.[6]

1. **Hypothesis 1: The generation effect will be larger in experimental contexts where *all* interactions (TAP, MAP, SAP) are simultaneously satisfied compared to *no* interactions satisfied.** For Hypothesis 1, I expected to find a larger generation effect in experimental contexts where the maximum number of lower-order interactions were satisfied compared to experimental contexts where the minimum number of lower-order interactions were satisfied. I used one-tailed t-tests to specify this directional hypothesis (maximum lower-order interactions satisfied > minimum lower-order interactions satisfied). Starting first with the lower-constraint task and free recall memory test data, the maximum number of interactions satisfied was three, while the minimum was one (see **Table II**). An independent samples t-test comparing memory performance under these two experimental contexts revealed memory performance was not greater for experimental contexts with three interactions satisfied (M = .160, SD = .142)

---

[6] Prior work has shown that lower-constraint generation tasks are associated with a greater likelihood for subjects to generate a response that is different from the normed expected response (compared to higher-constraint tasks), leading to potential item-selection effects (McCurdy et al., 2020). One way to account for item-selection effects is to only examine memory performance on trials where participants generated the normed expected target word. In addition to the analyses I report below, I ran the same set of analyses excluding any trials where the subject generated a response that was different from the normed target. This set of analyses (not reported) resulted in identical significance as the effects reported below, thus item-selection effects did not seem to significantly influence these data.

compared to experimental contexts with one interaction satisfied (*M* = .127, *SD* = .151), *t*(112) = 1.19, *p* (one-tailed) = .117, *d* = 0.22. For the higher-constraint task and free recall memory test data, the maximum number of interactions satisfied was two, and the minimum was zero (see **Table III**). This analysis indicated no differences between contexts with two interactions satisfied (*M* = .172, *SD* = .137) and zero interactions satisfied (*M* = .166, *SD* = .122), *t*(112) = 0.23, *p* (one-tailed) = .409, *d* = 0.04. Moving on to the recognition data, for experimental contexts with the lower-constraint task, the maximum number of interactions satisfied was two, while the minimum was zero (see **Table IV**). An independent samples t-test revealed no differences between experimental contexts with two interactions satisfied (*M* = .273, *SD* = .233) compared to zero interactions satisfied (*M* = .271, *SD* = .237), *t*(112) = 0.05, *p* (one-tailed) = .479, *d* = 0.01. Finally, for the higher-constraint task and recognition memory test, the maximum number of interactions satisfied was three, and the minimum was one (see **Table V**). An independent samples t-test revealed no significant differences between experimental contexts with three interactions satisfied (*M* = .301, *SD* = .236) and contexts with one interaction satisfied (*M* = .238, *SD* = .257), *t*(112) = 1.38, *p* (one-tailed) = .086, *d* = 0.26. Overall, despite the memory effects being in the expected direction, we found no substantial evidence supporting Hypothesis 1 in our data.

2. **Hypothesis 2: The effects of TAP, MAP, and SAP are additive.** For Hypothesis 2, I expected that as the number of lower-order interactions satisfied by the experimental context increased, the size of the generation effect would increase. In other words, I expected a positive relationship between the number of lower-order interactions satisfied and generation effect magnitude. To test this hypothesis, I conducted four linear regression analyses with the generation effect measure as the dependent variable and the number of lower-order

interactions satisfied as the predictor variable. First, for the lower-constraint task and free recall memory test data (**Table II**), the model did not account for a significant amount of variance, R^2 = .01, F(1, 226) = 1.28, p = .263, indicating that the number of interactions satisfied (β = .016, *SE* = .014) did not significantly predict memory performance in these data. Second, for the higher-constraint task and free recall memory test data (**Table III**), the model did not account for a significant amount of variance in memory performance, R^2 = .00, F(1, 226) = 0.06, p = .810, again indicating that the number of lower-order interactions satisfied (β = .003, *SE* = .013) did not predict memory performance. Third, for the lower-constraint task and recognition memory test data (**Table IV**), the model revealed that the number of lower-order interactions satisfied (β = .000, *SE* = .024) did not predict a significant amount of variance, R^2 = .00, F(1, 226) = 0.00, p = .981. Fourth, for the higher-constraint task and recognition memory test (**Table V**), the model again showed that the number of lower-order interactions satisfied (β = .031, *SE* = .022) did not predict memory performance, R^2 = .01, F(1, 226) = 1.89, p = .171. Overall, these analyses provide no support for Hypothesis 2, suggesting that the effects of each lower-order interaction (TAP, MAP, SAP) are not necessarily additive.

3. **Hypothesis 3: The relative influence of TAP, MAP, and SAP on memory performance is equal.** For Hypothesis 3, I predicted that the influence of each lower-order interaction on memory performance would be equal. To test this hypothesis, I conducted a 2 (generation task) x 2 (memory test) x 2 (materials) x 2 (subject ability) mixed ANOVA to examine each lower-order interaction included in the contextual theory. This analysis provides two pieces of information. First, finding a significant four-way interaction (among all four factors included) would provide evidence that each of the factors included in the contextual theory is important to consider in predicting the magnitude of the generation effect. Second,

examining the interactions between the generation task and memory test, generation task and materials, and generation task and subject ability factors provides evidence about the relative influence of TAP, MAP, SAP, respectively, in determining the size of the generation effect. Importantly, this analysis also serves as a sort of "manipulation check" to ensure that the current study design was sufficient enough to test the contextual theory (by checking that all the lower-order interactions were satisfied).

In line with central tenet of the contextual theory, the analysis revealed a significant four-way interaction, $F(1, 224) = 4.52$, $p = .035$, $\eta_p^2 = .020$, indicating that the size of the generation effect significantly differed based on the interactions between all four factors considered in the theory. Turning to the lower-order interactions of interest (TAP, MAP, SAP), the generation task by memory test (TAP) interaction was not significant, $F(1, 224) = 0.97$, $p = .191$, $\eta_p^2 = .014$, indicating that the size of the generation effect for the lower-constraint task and the higher-constraint task did not change based on the type of memory test given. In other words, both generation tasks led to similar generation effects regardless of the type of memory test (free recall or recognition), which is in contrast to the predictions of TAP. The interaction between generation task and materials (MAP) just missed significance, $F(1, 224) = 3.65$, $p = .057$, $\eta_p^2 = .016$, indicating that the size of the generation effect from the lower- and higher-constraint generation tasks marginally differed between the related list structure and unrelated list structure. Following up this analysis showed that for the lower-constraint generation task, the generation effect was numerically larger for the related list structure ($M = .241$, $SD = .226$), compared to the unrelated list structure ($M = .188$, $SD = .213$), $t(111) = 2.44$, $p = .071$, $d = 0.16$. For the higher-constraint task, however, the size of the generation effect was similar between the related ($M = .234$, $SD = .208$) and unrelated list structure ($M = .216$, $SD = .202$), $t(115) = 0.87$, $p = .819$, $d =$

0.06. These results are in contrast to the predictions of MAP. Lastly, the interaction between the generation task and subject ability (SAP) was significant, $F(1, 224) = 9.40$, $p = .002$, $\eta_p^2 = .040$ (see **Figure 3**). Follow-up analyses revealed that for the lower-constraint task, high English vocab subjects ($M = .248$, $SD = .220$) had a larger generation effect compared to low English vocab subjects ($M = .178$, $SD = .217$), $t(224) = 3.27$, $p = .007$, $d = 0.22$. However, for the higher-constraint task there was no difference in the size of the generation effect between the high English vocab subjects ($M = .232$, $SD = .202$) the low English vocabulary subjects ($M = .217$, SD = .209), $t(224) = 0.74$, $p = .458$, $d = 0.05$. The only other significant finding from this analysis was a main effect for subject ability, $F(1, 224) = 4.83$, $p = .029$, $\eta_p^2 = .016$, where high English vocabulary subjects ($M = .242$, SD = .153) generally had larger generation effects compared to low English vocabulary subjects ($M = .199$, SD = .145). Overall, this analysis does not support Hypothesis 3, and instead suggests that SAP had the strongest influence on the size of the generation effect in these data.[7]

E. **Indices of Item-Specific and Relational Processing**

Given that the predictions of the contextual theory are based on the interactions between different types of encoding processes, it is important to investigate the relative amount and type of processing induced by the different manipulations of the experimental factors included in this study. Prior work has identified several ways to assess item-specific and relational processing, independent of memory performance using free recall data (Burns, 2006). For item-specific

---

[7] I conducted a secondary analysis for Hypothesis 3 using the working memory measure (backward digit span) as the subject ability grouping factor (see **Table VI**). This analysis revealed a similar pattern of results, but with two changes. One, the generation task by materials (MAP) interaction was significant, $F(1, 214) = 4.97$, $p = .027$, $\eta_p^2 = .023$, with the difference between related list structure and unrelated list structure coming out significant for the lower-constraint generation task, $t(214) = 2.67$, $p = .048$, $d = 0.18$ (no differences for the higher-constraint task). Two, the generation task by subject ability (SAP) was no longer significant, $F(1, 214) = 0.92$, $p = .761$, $\eta_p^2 = .000$. It is interesting to note that a correlation analysis between English vocabulary scores and working memory revealed a significant, positive relationship, $r = .11$, $p = .048$. Thus, these analyses suggest that English vocabulary may play a significant role in influencing encoding processing and memory performance.

processing, I used the items per category (IPC) measure (Burns & Brown, 2000; Hunt & Seta, 1984), which uses the free recall data to calculate the total number of items recalled out of the number of categories recalled (e.g., if 14 items are recalled in total, and these 14 items come from four of the possible twelve categories, IPC would equal 14/4 = 3.5). This measure has been shown to be a reliable estimate of item-specific processing with a higher IPC score indicating greater item-specific processing (Jacoby, 1973; McDaniel et al., 1988; Sowder, 1973). For relational processing, I used the adjusted ratio of clustering (ARC) measure (Gerjuoy & Spitz, 1966; Roenker et al., 1971), which provides an index of relational processing while adjusting for differences in memory performance (Einstein & Hunt, 1980).[8] Using the free recall data, the ARC measure calculates how likely subjects are to recall items from the same category in successive recall attempts (i.e., clustering). These measures (IPC and ARC) are designed for use on between-subject manipulations, but given that the generation tasks were manipulated within-subjects in this study, I used the adapted versions of IPC and ARC developed by Klein and Kihlstrom (1986). Using this adaptation, a unique IPC score was calculated for each generation task (lower-constraint, higher-constraint) as the proportion of items recalled only out of the (four) categories encoded under each generation task. The adapted version of the ARC measure was used to produce a unique ARC score for each generation task by calculating the ratio of clustering by task (i.e., recalling items from the same task in consecutive recall attempts), accounting for chance clustering (see Klein & Kihlstrom, 1986). The IPC measure ranges from zero to six in this data (6 possible items to recall from each category), with a higher number

[8] There is a similar adjustment for the items per category measure (AIPC; Burns & Brown, 2000) that corrects for differences in memory performance. This measure, however, led to a large number of missing data when adapted to be calculated for each generation task, due to using only a small subset of the data. Given that there were no overall differences in memory performance between the two tasks, I used the unadjusted items per category (IPC) measure as the measure of item-specific processing. Furthermore, an analysis using this adjusted measure (AIPC) resulted in a similar pattern of results.

indicating more item-specific processing. For the ARC measure, any score above zero (0) can be interpreted as relational processing that is greater than chance, with a higher score indicating more relational processing (maximum score of 1).

In the analyses below, I first report two separate 2 x 2 x 2 mixed ANOVAs, one for the IPC scores (item-specific processing), and a second for the ARC scores (relational processing). Each analysis includes generation task type (lower-constraint, higher-constraint), materials (related list structure, unrelated list structure) and subject ability (high English vocab, low English vocab) as the independent variables. These analyses will provide estimates of observed item-specific and relational processing from the different experimental contexts tested in this study.[9] Then, to examine the assumption that different memory tests are sensitive to differences in item-specific and relational processing, I conducted a correlational analysis between the IPC and ARC scores, and memory performance on the recall and recognition memory tests, respectively. This analysis will provide evidence about the sensitivity of the free recall test and recognition test to relational and item-specific processing in the data.

To examine the extent that the different experimental contexts used in this study influenced the amount of item-specific processing engaged at encoding, I conducted a 2 x 2 x 2 mixed ANOVA on the IPC scores. I predicted that the higher-constraint task, unrelated list structure materials, and high English vocab subjects would lead to greater item-specific processing compared to the lower-constraint task, related list structure materials, and low English vocab subjects, respectively. The analysis revealed a main effect for generation task, $F(1, 203) = 67.16$, $p < .001$, $\eta_p^2 = .249$, indicating that the higher-constraint task ($M = 2.01$, $SD = 0.78$) led to

---

[9] The memory test factor was not included in these analyses because the different memory tests were not designed to induce a difference in relational versus item-specific processing at encoding, but rather to serve as different memory measures that are sensitive to each type of encoding processing.

significantly higher IPC scores compared to the lower-constraint task (M = 1.55, SD = 0.55), in line with my predictions. The main effect for materials was also significant, $F(1, 203) = 7.39$, p = .007, η_p^2 = .035. The direction of this effect was in contrast to the predictions, however, where the related materials (M = 1.88, SD = 0.75) led to significantly higher IPC scores than the unrelated materials (M = 1.68, SD = 0.67). Finally, the main effect for subject-ability just missed significance, $F(1, 203) = 3.57$, p = .060, η_p^2 = .017, with the high English vocab group (M = 1.85, SD = 0.71) having marginally higher IPC scores compared to the low English vocab group (M = 1.70, SD = 0.71). None of the interactions were significant (F's < 0.54, p's > .465). Overall, this analysis revealed that the higher-constraint task, related list structure materials, and (numerically) high English vocab subjects led to higher estimates of item-specific processing at encoding.

To examine differences in relational processing between the three factors, I conducted a second 2 x 2 x 2 ANOVA on the ARC scores. In this analysis, I predicted that the lower-constraint generation task, related list structure materials, and high English vocab subjects would induce greater relational processing relative to the higher-constraint generation task, unrelated list structure materials, and high English vocab subjects, respectively. The analysis revealed no main effect for generation task type, $F(1, 110) = 0.00$, p = .971, η_p^2 = .000, suggesting there were no differences in relational processing between the two generation tasks. Follow up analyses, however, importantly showed that ARC scores for the lower-constraint (M = .023, SD = .670), $t(109) = 0.32$, p = .745, and higher-constraint generation tasks (M = .019, SD = .856), $t(109)$ = 0.28, p = .782, were not significantly higher than chance clustering (0). The main effect for materials was significant, $F(1, 110) = 20.90$, p < .001, η_p^2 = .160, such that the related list structure materials (M = .260, SD = .674) led to greater ARC scores than unrelated list structure materials (*M* = -.218, *SD* = .847). Finally, the main effect for subject group was not significant, $F(1, 110) = 0.00$, $p = .993$, $\eta_p^2 = .000$, indicating no differences in ARC scores between the high English vocab subjects (*M* = .020, *SD* = .709) and the low English vocab subjects (*M* = .021, *SD* = .879). Follow-up analyses showed that neither group had ARC scores greater than chance (0), *t*'s < .291, *p*'s > .772. Overall, the results of this analysis indicate that the related list structure materials induced greater relational processing relative to the unrelated list structure materials, however the generation task type and subject ability factors did not seem to significantly induce relational processing in these data.

Next, I conducted a correlation analysis to examine the relationship between the measures of item-specific (IPC) and relational processing (ARC) and memory performance on the free recall and recognition memory tests. Specifically, I correlated the IPC and ARC scores for both generation tasks with performance on both the free recall test (free recall hits) and the recognition test (hits minus false alarms). I predicted that the free recall test would be sensitive to relational processing, whereas the recognition test would be sensitive to differences in item-specific processing. Starting first with item-specific processing, IPC scores were significantly, positively associated with recognition memory performance for both generation tasks (*r*'s > .228, *p*'s < .001), in line with my predictions. Interestingly, the relationship between IPC scores and free recall performance was also significant and positive for both tasks (*r*'s > .698, *p*'s < .001). This suggests that item-specific processing was associated with significant increases in both recognition and free recall performance. Turning to relational processing, I examined the relationship between ARC scores and free recall and recognition memory performance for both generation tasks. This analysis revealed that ARC scores (relational processing) were significantly associated with free recall memory performance for both tasks (*r*'s > .234, *p*'s <

.001). For the recognition memory test, there was a positive relationship between ARC scores and memory performance for the lower-constraint task (r = .226, p < .001), but this association was not significant for the higher-constraint task (r = .102, p = .130). Overall, this correlation analysis revealed that both types of memory tests (free recall, recognition) are generally sensitive to both types of processing (item-specific, relational).

To summarize, these analyses provide critical evidence about the original assumptions about the experimental contexts tested in this study and the relative amount of item-specific and relational processing induced under these contexts. Starting first with the two generation tasks used, these findings suggest that both tasks led to minimal (not greater than chance) relational processing, whereas the higher-constraint task lead to greater item-specific processing compared to the lower-constraint task. For the materials, results suggest that the related list structure materials led to both greater relational and item-specific processing relative to the unrelated list structure materials. As for subject-ability, the analyses suggest that high English vocabulary subjects engaged in (marginally) greater item-specific processing, with no differences in relational processing. Additionally, the correlation analysis revealed that increases in performance on both types of memory tests are generally sensitive to both item-specific and relational processing at encoding. These findings have important implications for the interpretation of the data presented in the **Memory Performance** section, and can help make sense of these findings. I discuss this topic more in-depth in the discussion. Most notably is the implication of finding that some of the core assumptions were not met by the current experimental design. These assumptions are necessary in order to truly test the contextual theory.

IV. DISCUSSION

The goal of this study was to develop and test a contextual theory that accounts for variations in the size of the generation effect by considering the interactions among four key experimental factors (generation task, memory test, materials, subject ability). The data reported in this study support the idea that each of the four factors play a role in determining the size of the generation effect. The predictions of the theory about how these factors interact were largely unsupported, however, indicating that there is still more research to be done toward understanding exactly *how* the interactions among these factors influence the size of the generation effect. Despite limited support for the hypotheses in this study, the data reported here provide a starting point for research going forward to better understand how these factors influence memory. The discussion will be broken down into three main sections. In the first section, I will interpret the data presented in this study and discuss the implications of these findings. The second section will focus on questions that still remain about the contextual theory, and I discuss ways that future research can address these questions. Finally, in the third section I focus on some broader implications of developing a contextual theory of the generation effect by considering how this type of theory can advance our understanding of the generation effect relative to currently existing theories.

A. Present Findings

1. **Hypothesis 1**. The results of this study did not support Hypothesis 1, which predicted that the generation effect would be larger under experimental contexts where the maximum number of lower-order interactions were satisfied compared to contexts where the minimum number of these interactions were satisfied. Although the pattern of results showed that the generation effect was numerically higher in experimental contexts with more interactions

satisfied compared to contexts with fewer interactions satisfied, these differences were not statistically significant. Furthermore, the data showed that the experimental contexts with the largest generation effect and smallest generation effect were not the contexts where all of the interactions were satisfied and none of the interactions were satisfied, respectively. One likely explanation for this pattern of results is that one or more of the lower-order interactions were not necessarily “satisfied” by the experimental contexts as predicted based on assumptions from prior work. “Satisfaction” of a lower-order interaction in this study was determined *a priori* based on assumptions about the underlying encoding processes induced by a particular manipulation (e.g., a lower-constraint generation task induces greater relational processing, while a higher-constraint task induces greater item-specific processing). The analyses reported on the independent measures of item-specific (IPC) and relational processing (ARC), however, suggest that some of the *a priori* assumptions did not pan out for the manipulations used in this study. For instance, it was assumed that a lower-constraint generation tasks induces relational processing at encoding, based on evidence from prior work (McCurdy et al., 2019; McCurdy et al., 2020). Data from the measure of relational processing (ARC), however, indicated that the lower-constraint generation task did not induce relational processing greater than chance in this study, providing evidence against this assumption. The encoding processes induced by a given experimental manipulation are critical because the theory’s predictions are based on the interactions of these processes, not the specific manipulations used in the present study. Thus, the finding that some of the experimental manipulations used in this study did not result in the expected type of encoding processes could explain why the data did not support the predictions of Hypothesis 1.

Although evidence suggests that the lower-constraint generation task induced a different type of processing than expected, the observed processing from the higher-constraint generation task aligned more closely with its assumptions. As predicted based on prior work, the higher-constraint generation task led to increased item-specific processing (as measured by IPC scores) relative to the lower-constraint generation task. Thus, despite the potential issues in evaluating the theory from the lower-constraint generation task, the experimental contexts using the higher-constraint task (Tables **III** and **V**) could be considered more reliable in this data set. In the analyses testing Hypothesis 1, the set of experimental contexts reported in **Table V** showed a mean difference in the size of the generation effect of about 6.3% between the context with the maximum (3) and the context with the minimum (1) number of interactions satisfied, yielding the largest effect size (d = .23) out of all the analyses examining Hypothesis 1 (see **Memory Performance**). Despite this, the mean difference was still not statistically significant given the small effect size, which is not consistent with Hypothesis 1. Although the data did not fully support the predictions of this hypothesis, it is encouraging that the trends in the data were in the expected direction. Future work using experimental designs that meet these necessary underlying assumptions (e.g., a lower-constraint task that induces relational processing, etc.) might be necessary before we see support for Hypothesis 1.

2. **Hypothesis 2.** The second hypothesis of this study that predicted a positive, linear relationship between the number of lower-order interactions satisfied and the size of the generation effect (i.e., more interactions satisfied = larger generation effect) was also not supported in this study. Although some of the data showed a weak trend in the expected direction, very little variance in the size of the generation effect across experimental contexts was accounted for by the number of interactions satisfied. Similar to Hypothesis 1, the most

likely explanation for this is that the expected processing from the different manipulations used in this study did not always align with the observed processing (see **Indices of Item-Specific and Relational Processing**). The finding that the lower-constraint task did not induce relational processing is especially problematic for the evaluation of this hypothesis because each lower-order interaction involved the processing between the generation task and the other three factors included in the theory. Thus, this particular failed assumption influenced not just one interaction, but all three lower-order interactions. In addition to the generation task factor, manipulations of the other factors (e.g., memory test, materials) designed to induce specific types of processing were also not supported by the data. In order to truly test Hypothesis 2 (and other predictions of the contextual theory), it is critical that the manipulations used influence processing in the expected way in order to satisfy (or dissatisfy) the different lower-order interactions. The data show this was not the case with the current study design, suggesting future work may be necessary to design experimental contexts where all three lower-order interactions are satisfied in order to truly test this theory.

Given the implications of the observed processing not matching the expected processing of some of the manipulations in this study, an important question to consider is why some of these assumptions did not come out. Regarding the lower-constraint generation task, several studies have shown that generation tasks with fewer constraints lead to enhanced relational processing relative to higher-constraint tasks (McCurdy et al., 2017, 2019; McCurdy et al., 2020), yet in this study there was no evidence supporting that the lower-constraint generation task induced relational processing (based on the ARC score measure of relational processing). One reason this particular assumption may not have been met in this study is due to differences in types of relational processing. In the literature, the term *relational processing* has been used broadly to refer to the processing of the relationship between the target word and several different types of details, such as a cue word (i.e., cue-target relational processing; Hirshman & Bjork, 1988), other target words presented in the same list of stimuli (i.e., target-target relational processing; McDaniel et al., 1990; McDaniel et al., 1988), and even extraneous details (e.g., font color, background color) associated with the target item (i.e., context-target relational processing; Greenwald & Johnson, 1989; Marsh, 2006). Prior research on lower-constraint generation tasks, has primarily examined relational processing of the *cue-target relationship*, concluding that lower-constraint generation enhances the association between the cue and target word (McCurdy et al., 2017, 2019; McCurdy et al., 2020). In the present study, however, the cue-target relationship (semantic relationship) was different from the target-target relationship among different target items in the list (categorical relationship). It is possible that the lower-constraint task induced greater cue-target relational processing (as it has been shown to do in prior work), leading subjects to encode the semantic association between the cue and target more deeply at the expense of processing the categorical relationship among the target items. Prior work has suggested that subjects will use whatever information is most salient to help them generate, leading to differences in encoding processing and ultimately resulting in different patterns of memory performance (deWinstanley & Bjork, 1997; deWinstanley et al., 1996; McDaniel et al., 1990; McDaniel et al., 1988). Importantly, in the present study the semantic relationship between cue and target was incompatible with the categorical relationship among the items, potentially explaining why the lower-constraint generation task did not appear to induce relational processing based on the ARC score measure (a measure of categorical relational processing; Burns, 2006). Future work should consider using a memory test that is sensitive to cue-target relational processing (such as a cued recall test) may represent a better way to assess the contextual theory using this particularly lower-constraint generation task. Alternatively, a different lower-constraint generation task may be used that is designed to induce processing of the target-target relationship (a prospect I discuss later in the discussion).

Another point worth considering in attempting to understand why Hypothesis 2 was not supported, is the fact that the memory tests were manipulated within-subjects in the current design. It is possible that having the free recall test before the recognition test could have influenced processing measured in the recognition task. Specifically, it could be that accessing information in the free recall test induced additional relational processing for items that were recalled, then in the later recognition test, this additional relational processing may have influenced performance. This speculation is at least partially supported by the analyses of observed processing (IPC and ARC scores), which showed that increased relational processing (as measured by ARC scores) was associated with higher performance on the recognition memory test, in contrast to the expectation that recognition memory tests are primarily sensitive to item-specific processing. The idea that the free recall test may have influenced processing by acting as a second encoding attempt may also explain why we found no evidence of transfer-appropriate processing in the current data. Future work should consider using a single, independent memory test (or a between-subject design) to reduce this possibility and to provide a more stringent test of Hypothesis 2.

3. **Hypothesis 3.** Another goal of this study was to investigate the relative influence of the three lower-order interactions that make up the contextual theory (TAP, MAP, SAP) on the size of the generation effect. Hypothesis 3 predicted that each lower-order interaction would be relatively equal in explaining variations in the generation effect, however, the data showed that SAP, followed by MAP, accounted for the most variation, while TAP was found to have no

significant influence on the size of the generation effect. Results showed that the subject ability factor played the largest role for the lower-constraint generation task (see **Figure 3**). Specifically, high English vocabulary subjects benefited from the lower-constraint generation task significantly more than the low English vocabulary group, whereas both groups had a similar generation effect for the higher-constraint generation task. This finding is in line with other research showing that subject ability can influence the effectiveness of encoding strategies (McDaniel et al., 2002; Mulligan et al., 2018). Akin to the results reported in the present study, McDaniel et al. (2002) found that high-ability subjects (as measured by reading ability) showed larger generation effects compared to low-ability subjects when using a generation task that induced relational processing, whereas both subject ability groups showed similar memory performance from a generation task that induced item-specific processing. The authors proposed that low-ability subjects may be less effective at noticing and using relational information at encoding (selective deficit hypothesis). Thus, for a generation task that induces greater relational processing, the extra demands placed on encoding relational information may be less beneficial for low-ability subjects compared to high-ability subjects. As an example, for high-ability subjects who are assumed to be able to efficiently engage in relational processing, a generation task that induces more effortful processing of that information likely leads to a deeper encoding of relational details, and in turn a richer memory representation that can enhance later retrieval. For low-ability subjects, however, to the extent that processing relational information is inherently effortful due to their inability or inefficiency to engage in relational information, a generation task that places extra demands on this type of processing will hinder, rather than enhance the storage of relational information (McDaniel et al., 2002). This idea may be relevant to lower-constraint generation task used in the present study, which is thought to rely more heavily on relational information to generate an item. The low English vocabulary subjects may have benefited less from the lower-constraint task because they were less familiar with the relationship between the cue and target words, compared to the high English vocabulary subjects who showed robust memory benefits from the lower-constraint task. This explanation is in line with the subject-appropriate processing claim that a generation task will be beneficial to the extent that it induces processing that the subject is able to successfully engage in.

Another possible explanation for differences in the effectiveness of encoding strategies between low- and high-ability subjects is that low-ability subjects are less likely to spontaneously engage in processing induced by a given encoding task (Mulligan et al., 2018; Schindler et al., 2019; Wenzel & Reinhard, 2019). The important distinction between this explanation and the specific deficit hypothesis (proposed above) is that low-ability subjects are able to engage in a specific type of processing, but they are less likely to spontaneously engage in the processing on their own. When given an encoding task that guides them to engage in a certain type processing, however, low-ability subjects typically benefit to the same level of high-ability subjects (Mulligan et al., 2018; Schindler et al., 2019). This idea could also potentially explain why high-ability subjects performed better than low-ability subjects on the lower-constraint task, but there were no differences between the groups on the higher-constraint task in the current study. By design, the lower-constraint generation task supplies less information about the target word, which may require more self-initiated processing in order to effectively generate a response. This is in contrast to the higher-constraint task where there is more guidance about what should be generated, and thus less self-initiated processing is necessary to generate a response.

Overall, the finding of differences in the effectiveness of generation tasks based on subject abilities is important for the application of the generation effect in educational settings. One of the goals of this study is to provide a framework that can predict when the generation effect is most effective, given the conditions under which it is being used. The finding that SAP was the most influential interaction aligns with the growing amount of recent research indicating the importance of considering the abilities of the learner to determine the type of strategy that will be most beneficial to learning (Lin et al., 2018; Marsh & Butler, 2013; McDaniel & Einstein, 2005; McDaniel et al., 2002; Mulligan et al., 2018; Schindler et al., 2019; Wenzel & Reinhard, 2019). A final note worth mentioning in regard to the importance of the subject ability factor in determining the size of the generation effect is how "ability" is defined. In this study, I used English vocabulary as the ability factor, but researchers have measured subject abilities in a variety of ways including reading ability (McDaniel et al., 2002), cognitive functioning (Mulligan et al., 2018), and intelligence (Wenzel & Reinhard, 2019). Interestingly, in the present study when subjects were grouped by working memory capacity (instead of English vocabulary) the effect of SAP was no longer evident (see **Footnote 6**). This suggests that the type of ability measure used may be an important consideration in determining the influence of SAP in predicting the generation effect. The finding of differences in the size of the generation effect when using an English vocabulary test as the subject-ability measure (but not with working memory) is in line with the principles of SAP, that subject's ability should be directly tied to the processing induced by the generation task, as opposed to a general ability measure. It may be that working memory is not critical for the processing required by generation tasks, but vocabulary is. This idea converges with theoretical work proposing that semantic activation is a core memory mechanism underlying the generation effect (McCurdy et al., In press). Research has shown the generation effect only improves memory when the generation task activates sematic information, and does not improve memory for meaningless information (i.e., non-words; Gardiner & Hampton, 1985; Johns & Swanson, 1988; McElroy & Slamecka, 1982). Thus, if generating requires access to semantic information to be beneficial, then it makes sense that subjects with higher vocabulary ability would benefit to a greater extent as we saw in this study, due to their potentially larger semantic stores compared to subjects with lower vocabulary ability.

Beyond SAP, our data showed no evidence of a TAP effect and some evidence in contrast to the principle of MAP (matched processing between encoding task and materials improved memory to a greater extent than mismatched processing). Given that these principles have been so widely-supported in prior work, it is unlikely that these principles do not influence memory performance, but rather it is more likely that these principles did not apply in the current study design. Regarding the contextual theory, it is critical to develop an experimental design where the lower-order interactions are more strongly satisfied (or dissatisfied) to truly test the predictions of the theory. Thus, despite finding little support for the contextual theory in this study, it is still possible that the contextual theory can effectively explain variation in the generation effect. In the next section, I describe some potential avenues of future research that can more fully test the predictions of this theory going forward.

B. **Future Directions and Research Questions**

One of the limitations of this study was that there were several underlying assumptions that were not met (e.g., no evidence of transfer-appropriate processing, lower-constraint task did not induce sufficient relational processing, etc.). In order to truly test the contextual theory, these core assumptions need to be met first. Thus, one important direction for future research is to more rigorously test the predictions of the contextual theory by using manipulations that more tightly control the type of processing induced for each factor (especially encoding task, materials). For instance, in this study the lower-constraint generation task could be altered so that a category label serves as the cue word (e.g., animal-d__; animal-c__), instead of a unique semantic associate (as used in the current study), providing a stronger induction of relational processing of the categorical relationship among items (target-target relational processing). This small methodological change should yield a stronger "match" in processing between both the materials and free recall memory test, two assumptions that were not met in the present design. Designing studies with stronger "matches" and "mismatches" between the four experimental factors will enhance tests of the contextual theory going forward.

Relatedly, understanding the type and amount of encoding processing induced by the specific manipulations used in a given study is critical in order to test (and to use) the predictions of the theory. Thus, advancing some of the independent measures of underlying memory processes will be an important area of future research. Several measures exist for the distinction between item-specific and relational processing, which are nicely summarized and contrasted in (Burns, 2006). However, there are still limitations in the use of these existing measures. For one, many of the relational processing measures are tailored for categorical processing among different items in a list, therefore these measures are not useful for measuring other types of relational processing (e.g., semantic relations between the cue and target). Additionally, many of these measures are designed for use in between-subject manipulations, and need to be adapted for within-subject manipulations. The use of these measures becomes complicated in mixed designs, however, as in the present study, reducing the accuracy and usefulness of these measures. Research aimed at continuing to develop these measures of encoding processes will further the capability of future studies to more accurately test the predictions of the theory.

Finally, a primary goal of developing this contextual theory was to advance knowledge about the "higher-order" interactions between the four key factors included in the theory. The data in this study revealed a four-way interaction between all four factors, indicating that memory performance was indeed influenced by the interaction among these factors. This finding suggests that future work should go beyond examining lower-order interactions if we want to truly begin to understand the complexities of human memory. Limited research has shown evidence of a three-way interaction between encoding task, materials, and subject abilities (McDaniel et al., 2002), however studies like these (investigating the interactions among more than two key experimental factors) are rare in the literature. It is possible (and perhaps likely) that the fundamental memory concepts we currently use to guide our understanding of memory (TAP, MAP, SAP), change (i.e., are stronger, weaker, or no longer apply) across a variety of experimental contexts. Future work aimed at understanding how these concepts adapt across different contexts would be an important step forward in beginning to advance our understanding of the complexities of human memory.

C. **Toward a Better Account of the Generation Effect**

Decades of research on the generation effect has produced several theories that explain why the generation effect occurs while other theories attempt to explain specific boundary conditions of the generation effect (for a review, see McCurdy et al., In press). The most prominent theory on the generation effect (the multi-factor theory) suggests that the memory benefits from self-generating can be explained by enhanced item-specific and relational processing at encoding relative to reading (Hirshman & Bjork, 1988; McDaniel et al., 1990; McDaniel et al., 1988). Recent discoveries on the generation effect, however, reveal that this explanation of the generation effect cannot fully account for the complexities in the generation effect data. For example, some generation tasks may induce more or less item-specific and relational processing, respectively, yet there is no facet of the multi-factor theory to account for differences among generation tasks. The contextual theory developed in this study advances on the existing multi-factor theory by accounting for the *extent* a generation task induces item-specific and relational processing, respectively. This allows for more specificity in predicting the size of the generation effect, whereas the multi-factor theory simply predicts whether a generation effect will occur or not.

Another limitation of the multi-factor theory is that it only takes into account the processing induced by the generation task to explain the generation effect (see Jacoby, 1983; Mulligan, 2004; 2011, for a TAP account considering two factors; encoding task and memory test). As we have shown in this study, a host of contextual factors interact with the processing induced by the generation task to determine memory performance. Thus, a generation effect theory that aims to predict the size of the generation effect across contexts should be able to account for these interactions. The contextual theory introduced in this study considers four factors to explain variations in the generation effect. In line with other theoretical work that suggests considering multiple factors simultaneously often provides a more complete account than considering these factors in isolation (Healey & Kahana, 2016), the contextual theory of the generation effect aims to provide a more nuanced explanation of the generation effect.

V. CONCLUSION

Decades of research has shown that the generation effect is a robust memory effect, however, the magnitude of this effect varies widely across experimental contexts (McCurdy et al., In press). The contextual theory introduced in this study provides a framework to understand and predict variations in the generation effect. The findings from this study show that the generation effect is indeed largely influenced by experimental context, including the type of generation task, memory test, materials, and subject abilities. Although the present study offers a clear indication that each of these factors is uniquely important to consider, there is still more knowledge to be discovered about how these factors interact to influence memory. Future work should continue to investigate the interactions among these four key experimental factors and how these interactions influence memory performance. Lastly, my goal is for this project to serve as a paradigm for future research to examine the interactions among these factors, paving the way for taking the next step in advancing our understanding of the generation effect and other memory effects.

CITED LITERATURE

Begg, I., & Snider, A. (1987). The generation effect: Evidence for generalized inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(4), 553-563.

Begg, I., Vinski, E., Frankovich, L., & Holgate, B. (1991). Generating makes words memorable, but so does effective reading. *Memory & Cognition, 19*(5), 487-497.

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

Burns, D. J. (2006). Assessing distinctiveness: Measures of item-specific and relational processing. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and Memory* (pp. 109-130). Oxford, NY: Oxford University Press.

Burns, D. J., & Brown, C. (2000). The Category Access Measure of Relational Processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(4), 1057-1062.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior, 11*(6), 671-684.

deWinstanley, P. A., & Bjork, E. L. (1997). Processing instructions and the generation effect: A test of the multifactor transfer-appropriate processing theory. *Memory, 5*(3), 401-422.

deWinstanley, P. A., Bjork, E. L., & Bjork, R. A. (1996). Generation effects and the lack thereof: The role of transfer-appropriate processing. *Memory, 4*(1), 31-48.

Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology*: University Microfilms.

Einstein, G. O., & Hunt, R. R. (1980). Levels of processing and organization: Additive effects of individual-item and relational processing. *Journal of Experimental Psychology: Human Learning and Memory, 6*(5), 588-598.

Einstein, G. O., McDaniel, M. A., Owen, P. D., & Cote, N. C. (1990). Encoding and recall of texts: The importance of material appropriate processing. *Journal of Memory and Language, 29*(5), 566-581.

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

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

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

Gardiner, J. M., & Hampton, J. A. (1985). Semantic memory and the generation effect: Some tests of the lexical activation hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(4), 732.

Gardiner, J. M., Smith, H. E., Richardson, C. J., Burrows, M. V., & Williams, S. D. (1985). The generation effect: Continuity between generating and reading. *The American Journal of Psychology, 98*(3), 373-378.

Gerjuoy, I., & Spitz, H. (1966). Associative clustering in free recall: intellectual and developmental variables. *American Journal of Mental Deficiency, 70*(6), 918-927.

Graf, P., & Levy, B. A. (1984). Reading and remembering: Conceptual and perceptual processing involved in reading rotated passages. *Journal of Memory and Language, 23*(3), 405-424.

Greenwald, A. G., & Johnson, M. M. (1989). The generation effect extended: Memory enhancement for generation cues. *Memory & Cognition, 17*(6), 673-681.

Healey, M. K., & Kahana, M. J. (2016). A four-component model of age-related memory change. *Psychological Review, 123*(1), 23.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in Cognitive Psychology, The Loyola Symposium* (pp. 77-99). Potomac, MD: Erlbaum.

Hirshman, E., & Bjork, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 484-494.

Hoffman, R. R. (1986). Context and contextualism in the psychology of learning. *Cahiers de psychologie cognitive, 6*(2), 215-232.

Huff, M. J., & Bodner, G. E. (2014). All varieties of encoding variability are not created equal: Separating variable processing from variable tasks. *Journal of Memory and Language, 73*, 43-58.

Hunt, R. R. (2006). The concept of distinctiveness in memory research. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and Memory* (pp. 3-25). Oxford, NY: Oxford University Press.

Hunt, R. R. (2012). Distinctive processing: The co-action of similarity and difference in memory. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 56, pp. 1-46). San Diego, CA: Academic Press.

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior, 20*(5), 497-514.

Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language, 32*(4), 421-445.

Hunt, R. R., & Seta, C. E. (1984). Category size effects in recall: The roles of relational and individual item information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(3), 454-464.

Jacoby, L. L. (1973). Test appropriate strategies in retention of categorized lists. *Journal of Verbal Learning and Verbal Behavior, 12*(6), 675-682.

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

Jacoby, L. L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior, 22*(5), 485-508.

Jenkins, J. J. (1979). Four points to remember: A tetrahedral model of memory experiments. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of Processing in Human Memory* (pp. 429-446). Hillsdale, NJ: Erlbaum.

Johns, E. E., & Swanson, L. G. (1988). The generation effect with nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(1), 180-190.

Kalyuga, S. (2009). The Expertise Reversal Effect. In S. Kalyuga (Ed.), *Managing Cognitive Load in Adaptive Multimedia Learning* (pp. 58-80). Hershey, PA: IGI Global.

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America, 116*(4), 2395-2405.

Kintsch, W., & Young, S. R. (1984). Selective recall of decision-relevant information from texts. *Memory & cognition, 12*(2), 112-117.

Klein, S. B., & Kihlstrom, J. F. (1986). Elaboration, organization, and the self-reference effect in memory. *Journal of Experimental Psychology: General, 115*(1), 26-38.

Laudan, L. (1977). Progress and its problems: Towards a theory of scientific growth. Berkeley. *University of California Press. Leahey, TH (1992). The mythical revolutions of American psychology. American Psychologist, 47*, 308-318.

Lin, C., McDaniel, M. A., & Miyatsu, T. (2018). Effects of flashcards on learning authentic materials: The role of detailed versus conceptual flashcards and individual differences in structure-building ability. *Journal of Applied Research in Memory and Cognition, 7*(4), 529-539.

MacWhinney, B., James, J. S., Schunn, C., Li, P., & Schneider, W. (2001). STEP—A system for teaching experimental psychology using E-Prime. *Behavior Research Methods, Instruments, & Computers, 33*(2), 287-296.

Marsh, E. J. (2006). When does generation enhance memory for location? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(5), 1216-1220.

Marsh, E. J., & Butler, A. C. (2013). Memory in educational settings. *Oxford handbook of cognitive psychology*.

McCurdy, M. P., Leach, R. C., & Leshikar, E. D. (2017). The generation effect revisited: Fewer generation constraints enhances item and context memory. *Journal of Memory and Language, 92*, 202-216.

McCurdy, M. P., Leach, R. C., & Leshikar, E. D. (2019). Fewer constraints enhance the generation effect for source memory in younger, but not older adults. *Open Psychology, 1*(1), 168-184.

McCurdy, M. P., Sklenar, A. M., Frankenstein, A. N., & Leshikar, E. D. (2020). Fewer constraints increase the generation effect for item and source memory through enhanced relational processing. *Memory*, 1-19.

McCurdy, M. P., Viechtbauer, W., Sklenar, A. M., Frankenstein, A. N., & Leshikar, E. D. (In press). Theories of the generation effect and the impact of generation constraint: A meta-analytic review. *Psychonomic Bulletin & Review*.

McDaniel, M. A., & Einstein, G. O. (1989). Material-appropriate processing: A contextualist approach to reading and studying strategies. *Educational Psychology Review, 1*(2), 113-145.

McDaniel, M. A., & Einstein, G. O. (2005). Material Appropriate Difficulty: A Framework for Determining When Difficulty Is Desirable for Improving Learning *Experimental cognitive psychology and its applications.* (pp. 73-85): American Psychological Association.

McDaniel, M. A., Einstein, G. O., Dunay, P. K., & Cobb, R. E. (1986). Encoding difficulty and memory: Toward a unifying theory. *Journal of Memory and Language, 25*(6), 645-656.

McDaniel, M. A., Hines, R. J., & Guynn, M. J. (2002). When text difficulty benefits less-skilled readers. *Journal of Memory and Language, 46*(3), 544-561.

McDaniel, M. A., Riegler, G. L., & Waddill, P. J. (1990). Generation effects in free recall: Further support for a three-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(5), 789.

McDaniel, M. A., Waddill, P. J., & Einstein, G. O. (1988). A contextual account of the generation effect: A three-factor theory. *Journal of Memory and Language, 27*(5), 521-536.

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

McKeachie, W. J. (1974). The decline and fall of the laws of learning. *Educational Researcher, 3*(3), 7-11.

McNamara, D. S., & Healy, A. F. (1995). A procedural explanation of the generation effect: The use of an operand retrieval strategy for multiplication and addition problems. *Journal of Memory and Language, 34*(3), 399-416.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*(5), 519-533.

Mulligan, N. W. (2004). Generation and memory for contextual detail. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(4), 838-855.

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

Mulligan, N. W., & Lozito, J. P. (2004). Self-generation and memory. *The Psychology of Learning and Motivation: Advances in Research and Theory, 45*, 175-214.

Mulligan, N. W., & Lozito, J. P. (2006). An asymmetry between memory encoding and retrieval: Revelation, generation, and transfer-appropriate processing. *Psychological Science, 17*(1), 7-11.

Mulligan, N. W., Rawson, K., Peterson, D., & Wissman, K. (2018). The Replicability of the Negative Testing Effect: Differences Across Participant Populations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(5), 752-763.

Nairne, J. S., & Widner, R. L. (1987). Generation effects with nonwords: The role of test appropriateness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(1), 164-171.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36*(3), 402-407.

Nieznański, M. (2012). Effects of generation on source memory: A test of the resource tradeoff versus processing hypothesis. *Journal of Cognitive Psychology, 24*(7), 765-780.

Nieznański, M. (2014). The role of reinstating generation operations in recognition memory and reality monitoring. *Polish Psychological Bulletin, 45*(3), 363-371.

Payne, D. G., Neely, J. H., & Burns, D. J. (1986). The generation effect: Further tests of the lexical activation hypothesis. *Memory & Cognition, 14*(3), 246-252.

Psychology Software Tools, I. (2012). E-Prime 2.0.8.

Rabinowitz, J. C., & Craik, F. I. (1986). Specific enhancement effects associated with word generation. *Journal of Memory and Language, 25*(2), 226-237.

Reardon, R., Durso, F. T., Foley, M. A., & McGahan, J. R. (1987). Expertise and the generation effect. *Social Cognition, 5*(4), 336-348.

Roediger, H. L. (2008). Relativity of remembering: Why the laws of memory vanished. *Annual Review of Psychology, 59*, 225-254.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological science, 17*(3), 249-255.

Roenker, D., Thompson, C., & Brown, S. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin, 76*(1), 45-48.

Rohrer, D., & Pashler, H. (2007). Increasing retention without increasing study time. *Current Directions in Psychological Science, 16*(4), 183-186.

Schindler, J., Schindler, S., & Reinhard, M.-A. (2019). Effectiveness of Self-Generation during Learning Is Dependent on Individual Differences in Need for Cognition. *Frontline Learning Research, 7*(2), 23-39.

Schmidt, S. R. (1990). A test of resource-allocation explanations of the generation effect. *Bulletin of the Psychonomic Society, 28*(1), 93-96.

Schmidt, S. R., & Cherry, K. (1989). The negative generation effect: Delineation of a phenomenon. *Memory & Cognition, 17*(3), 359-369.

Shipley, W. (1946). The Shipley-Hartford Vocabulary Test. *Shipley Institute of Living Scale*.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4*(6), 592-604.

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

Sowder, C. (1973). Mediated unlearning. *Journal of Experimental Psychology, 100*(1), 50-57.

Steffens, M. C., & Erdfelder, E. (1998). Determinants of positive and negative generation effects in free recall. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 51*(4), 705-733.

Thorndike, E. L. (1927). The law of effect. *The American Journal of Psychology, 39*(1/4), 212-222.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*(5), 352-373.

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language, 50*(3), 289-335.

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

Wenzel, K., & Reinhard, M.-A. (2019). Relatively unintelligent individuals do not benefit from intentionally hindered learning: The role of desirable difficulties. *Intelligence, 77*, 101405.

Woods, D. L., Kishiyama, M. M., Yund, E. W., Herron, T. J., Edwards, B., Poliva, O., . . . Reed, B. (2011). Improving digit span assessment of short-term verbal memory. *Journal of Clinical and Experimental Neuropsychology, 33*(1), 101-111.

Zelinski, E. M., & Gilewski, M. J. (1988). Memory for Prose and Aging: A Meta-Analysis. In M. L. Howe & B. C. J (Eds.), *Cognitive Development in Adulthood* (pp. 133-158). New York, NY: Springer.

APPENDIX A

Raney Vocabulary Assessment

Vocabulary Test (Version 6/09/2004) **Subject __________**

Directions: Choose the BEST definition for each word.

1. **ASCEND**
   A. to go up or mount
   B. consent
   C. improve with time
   D. to leave behind
   E. to replace a leader

2. **WARY**
   A. tired out
   B. rude; uncouth
   C. perturbed
   D. brand-new
   E. cautious; careful

3. **NURTURE**
   A. helped by man
   B. to feed or nourish
   C. to educate
   D. to protect by nature
   E. to cook

4. **INFINITESIMAL**
   A. very long
   B. very slow
   C. well defined
   D. uncompromising
   E. very small

5. **BELLIGERENT**
   A. informative
   B. blunt
   C. tiring
   D. war-like
   E. pro-active

6. **INDIFFERENT**
   A. similar
   B. unconcerned
   C. diffident
   D. solicitous
   E. opposite

7. **PERJURE**
   A. to save from indignity
   B. to improve or rectify
   C. to demand support
   D. to lie under oath
   E. day by day

8. **VERBOSE**
   A. slow
   B. impressive
   C. complicated
   D. wordy
   E. meaningless

9. **OPAQUE**
   A. transparent
   B. slippery
   C. impenetrable by light
   D. gem-like
   E. financially well-off

10. **SYNTHESIS**
    A. musical rendition of a written work
    B. a theory of immoral behavior
    C. the combination of parts to form a whole
    D. watching or guarding
    E. properties of artificial chemicals

11. **SPONTANEITY**
    A. unwanted laughter
    B. uncontrollable danger
    C. unplanned action
    D. unneeded socialism
    E. stand-up attitude

12. **VALIDATE**
    A. to prove
    B. to get paid back
    C. to expire
    D. to run away
    E. to complete successfully

13. **SUBORDINATE**
    A. to hypothesize in abstract
    B. to practice with instruction
    C. to levy upon others
    D. to go on vacation
    E. to rank in importance

14. **MEAGER**
    A. not full, inadequate
    B. to beg
    C. without self-respect
    D. in good shape, healthy
    E. wise, full of advice

APPENDIX A (continued)

15. EQUIVOCAL
A. premier, establishing new precedent
B. popular, known by everyone
C. exciting, causing a commotion
D. peculiar, one of a kind
E. uncertain, having two meanings

16. REBUKE
A. to dispute
B. poor reputation
C. to scold harshly
D. to stop at midpoint
E. to overfill

17. ECLECTIC
A. providential
B. of religious origins
C. purified
D. out of fashion
E. from various sources

18. TERSE
A. concise
B. private
C. angry
D. outdated
E. harsh-sounding

19. ILLUSORY
A. bright
B. deceptive
C. unhealthy
D. making a reference to
E. sometimes friendly, sometimes undependable

20. DIVULGE
A. to discourage
B. to pay for
C. to turn away
D. to reveal
E. to infiltrate

21. REPROVE
A. to reverse an argument
B. to be clean of
C. to express disapproval
D. to grovel for forgiveness
E. to encourage hope

22. IMPLAUSIBLE
A. could happen at any moment
B. not believable
C. unyielding
D. considered tactless
E. to serve or worship

23. INCONTROVERTIBLE
A. useless
B. prone to trouble making
C. indisputable
D. successful
E. unprotected

24. QUERY
A. excavation
B. prey
C. inquiry
D. strange occurrence
E. strange, odd

25. DISPERSE
A. to seize one's assets
B. to live in exile
C. to break up and scatter
D. to weaken connections
E. to make vacant

26. VACILLATE
A. to prepare for action; lubricate
B. to show indecision; to waver
C. to hold firmly, to be stubborn
D. to wait until the last second, delay
E. to scatter; to create chaos

27. SUPERFLUOUS
A. gay, happy
B. reserved, waiting
C. trivial; unimportant
D. unnecessary; excessive
E. undecided; variable

28. AUTONOMOUS
A. unknown identity
B. having many names
C. uncontrollable
D. independent existence
E. self-confidence

29. PRECEDENT
A. an expectation
B. most important event
C. a leader
D. a prior occurrence
E. a forgotten time

30. BOLSTER
A. to disagree, strongly
B. to defend, proudly
C. to reinforce, strengthen
D. to agonize, repeatedly
E. brutalize, mercilessly

APPENDIX B

Shipley Vocabulary Assessment

On each line, select the one word that means the same thing, or most nearly the same thing, as the word in CAPITAL letters.

TALK	draw	eat	speak	sleep
PERMIT	allow	sew	cut	drive
PARDON	forgive	pound	divide	tell
COUCH	pin	eraser	sofa	glass
REMEMBER	swim	recall	number	defy
TUMBLE	drink	dress	fall	think
HIDEOUS	silvery	tilted	young	dreadful
CORDIAL	swift	muddy	leafy	hearty
EVIDENT	green	obvious	skeptical	afraid
IMPOSTER	conductor	officer	book	pretender
MERIT	deserve	distrust	fight	separate
FASCINATE	welcome	fix	stir	enchant
INDICATE	defy	excite	signify	bicker
IGNORANT	red	sharp	uninformed	precise
FORTIFY	submerge	strengthen	vent	deaden
RENOWN	length	head	fame	loyalty
NARRATE	yield	buy	associate	tell
MASSIVE	bright	large	speedy	low
HILARITY	laughter	speed	grace	malice
SMIRCHED	stolen	pointed	remade	soiled
SQUANDER	tease	belittle	cut	waste
CAPTION	drum	ballast	heading	ape
FACILITATE	help	turn	strip	bewilder
JOCOSE	humorous	paltry	fervid	plain
APPRISE	reduce	strew	inform	delight
RUE	eat	lament	dominate	cure
DENIZEN	senator	inhabitant	fish	atom
DIVEST	dispossess	intrude	rally	pledge
AMULET	charm	orphan	dingo	pond
INEXORABLE	untidy	involatile	rigid	sparse
SERRATED	dried	notched	armed	blunt
LISSOM	moldy	loose	supple	convex
MOLLIFY	mitigate	direct	pertain	abuse
PLAGIARIZE	appropriate	intend	revoke	maintain
ORIFICE	brush	hole	building	lute
QUERULOUS	maniacal	curious	devout	complaining
PARIAH	outcast	priest	lentil	locker
ABET	waken	ensue	incite	placate
TEMERITY	rashness	timidity	desire	kindness
PRISTINE	vain	sound	first	level

APPENDIX C

Demographics Questionnaire

Subject number ___________
Experimenter ___________
Date ___________

Demographics Questionnaire

Date of Birth: ________________ **Gender:** O 1 Male O 2 Female

---

1. **Education completed (check highest level)**
   - O 1 Less than high school graduate (highest grade completed? ________)
   - O 2 High school graduate/G.E.D.
   - O 3 Some college, or trade, technical, or business school (how many years? _________)
   - O 4 Bachelor's degree
   - O 5 Some graduate work (how many years? _____)
   - O 6 Master's degree
   - O 7 M.D., J.D., Ph.D., other advanced degree

2. **Ethnicity**
   - O 1 Hispanic or Latino
   - O 2 Not Hispanic or Latino

3. **Race**
   - O 1 American Indian/Alaskan Native
   - O 2 Asian
   - O 3 Native Hawaiian/Other Pacific Islander
   - O 4 Black/African American
   - O 5 White/Caucasian
   - O 6 Multiracial (please specify ____________)
   - O 7 Other (please specify ________________)

4. **Is English your native and primary language?**
   - O 1 Yes
   - O 2 No (please specify your native/primary language _______________________)

APPENDIX D

Language History Questionnaire

Language History Questionnaire

Subject # ________ (Version 12-12-2001)

Sex _________ Age _________ What country were you born in? _______________________

Years living in U.S. _________ Years in U.S. Schools _________

(1) What is the FIRST language you spoke? If your parents spoke two languages to you, list BOTH languages.

(2) List from MOST fluent to LEAST fluent all of the languages that you know (write on the back of this page if you need more space). Note that the language you learned first is not necessarily the language you now know best. Specify the age at which you began to learn the language (if it is your native language you should specify age as "birth") and where you learned it (e.g., school, home, church).

	Language	Age learned	Location learned
Most fluent	___________________	__________	______________________________
	___________________	__________	______________________________
	___________________	__________	______________________________
Least fluent	___________________	__________	______________________________

(3) Answer the following questions. Complete only those questions that apply to you.
At what age did you begin <u>speaking</u> **English**? __________
At what age did you begin <u>reading</u> **English**? __________
At what age did you begin <u>speaking</u> your most fluent language **OTHER THAN** English? __________
At what age did you begin <u>reading</u> your most fluent language **OTHER THAN** English? __________

(4) Complete the following ratings. If you think you are more proficient in either English or your OTHER language, your ratings should reflect this difference. Answer only those questions that apply to you.

	NOT fluent									VERY fluent
For ENGLISH:										
How fluent are you in <u>speaking</u>?	1	2	3	4	5	6	7	8	9	10
How fluent are you in <u>understanding</u>?	1	2	3	4	5	6	7	8	9	10
How fluent are you in <u>reading</u>?	1	2	3	4	5	6	7	8	9	10
For your most fluent language OTHER THAN English:										
How fluent are you in <u>speaking</u>?	1	2	3	4	5	6	7	8	9	10
How fluent are you in <u>understanding</u>?	1	2	3	4	5	6	7	8	9	10
How fluent are you in <u>reading</u>?	1	2	3	4	5	6	7	8	9	10

TABLE I
Predicted Memory Performance as a Function of the Number of Lower-Order Interactions Satisfied

Predicted Memory Performance	Number of Interactions Satisfied		Lower Order Interactions: Generation task *match* Memory Test (TAP)	Generation task *mismatch* Materials (MAP)	Generation task *match* Subject Abilities (SAP)
Higher	**3**		✓	✓	✓
	2	a	☒	✓	✓
		b	✓	☒	✓
		c	✓	✓	☒
	1	a	☒	☒	✓
		b	☒	✓	☒
		c	✓	☒	☒
Lower	**0**		☒	☒	☒

Note. This table shows an orthogonal list of the eight possible experimental contexts that correspond to the interactions between the generation task and the other three factors (memory test, materials, subject abilities). In this table, a check mark indicates a satisfied lower-order interaction, and an "x" represents an unsatisfied lower-order interaction. According to the proposed Contextual Theory of the Generation Effect, memory performance should increase, as the number of lower-order interactions satisfied increases. Under the assumption that the relative amount of match (or mismatch) across the lower-order interactions is equal, this theory predicts no differences in the relative memory performance between rows of the same number (2_a, 2_b, 2_c; 1_a, 1_b, 1_c).

TABLE II
Predicted and Observed Memory Performance by Number of Interactions Satisfied for
Lower-Constraint (Relational) Generation Task and Free Recall (Relational) Memory Test

			Lower-Order Interactions			**Observed Memory Performance**	
Predicted Memory Performance	**Number of Interactions Satisfied**		**Generation task *match* Memory Test (TAP)**	**Generation task *mismatch* Materials (MAP)**	**Generation task *match* Subject Abilities (SAP)**	$M_{\text{GenEffect}}$ (*SD*)	***n***
Higher	**3**		✓ Free Recall	✓ Unrelated List	✓ High Eng. Vocab	.160 (.142)	59
	2	b	✓ Free Recall	☒ Related List	✓ High Eng. Vocab	.196 (.168)	61
	2	c	✓ Free Recall	✓ Unrelated List	☒ Low Eng. Vocab	.097 (.137)	53
Lower	**1**	c	✓ Free Recall	☒ Related List	☒ Low Eng. Vocab	.127 (.151)	55

Note. $M_{\text{GenEffect}}$ = mean difference between generation task and read control task, averaged across *n* subjects. *SD* = Standard deviation across *n* subjects. *n* = number of subjects contributing to this row. "Observed memory performance" represents the proportion of correctly recalled items out of all items encoded.

TABLE III

Predicted and Observed Memory Performance by Number of Lower-Order Interactions Satisfied for **Higher-Constraint (Item-Specific) Generation Task and Free Recall (Relational) Memory Test**

Predicted Memory Performance	Number of Interactions Satisfied		Lower-Order Interactions: Generation task *match* Memory Test (TAP)	Lower-Order Interactions: Generation task *mismatch* Materials (MAP)	Lower-Order Interactions: Generation task *match* Subject Abilities (SAP)	Observed Memory Performance: $M_{GenEffect}$ (*SD*)	*n*
Higher	**2**	a	☒ Free Recall	✓ Related List	✓ High Eng. Vocab	.172 (.137)	61
	1	a	☒ Free Recall	☒ Unrelated List	✓ High Eng. Vocab	.162 (.152)	59
	1	b	☒ Free Recall	✓ Related List	☒ Low Eng. Vocab	.154 (.155)	55
Lower	**0**		☒ Free Recall	☒ Unrelated List	☒ Low Eng. Vocab	.166 (.122)	53

Note. $M_{GenEffect}$ = mean difference between generation task and read control task, averaged across *n* subjects. *SD* = Standard deviation across *n* subjects. *n* = number of subjects contributing to this row. "Observed memory performance" represents the proportion of correctly recalled items out of all items encoded.

TABLE IV
Predicted and Observed Memory Performance by Number of Lower-Order Interactions Satisfied for
Lower-Constraint (Relational) Generation Task and Recognition (Item-Specific) Memory Test

			Lower-Order Interactions			Observed Memory Performance	
	Number of Interactions Satisfied		**Generation task *match* Memory Test (TAP)**	**Generation task *mismatch* Materials (MAP)**	**Generation task *match* Subject Abilities (SAP)**	**$M_{GenEffect}$ (*SD*)**	***n***
Higher	**2**	a	☒ Item Recognition	✓ Unrelated List	✓ High Eng. Vocab	.273 (.233)	59
Predicted Memory Performance	**1**	a	☒ Item Recognition	☒ Related List	✓ High Eng. Vocab	.362 (.260)	61
	1	b	☒ Item Recognition	✓ Unrelated List	☒ Low Eng. Vocab	.217 (.271)	53
Lower	**0**		☒ Item Recognition	☒ Related List	☒ Low Eng. Vocab	.271 (.237)	55

Note. $M_{GenEffect}$ = mean difference between generation task and read control task, averaged across *n* subjects. *SD* = Standard deviation across *n* subjects. *n* = number of subjects contributing to this row. "Observed memory performance" represents the proportion of correctly recognized "old" items (hits), minus the proportion of "lure" items incorrectly identified as "old" (false alarms).

TABLE V
Predicted and Observed Memory Performance by Number of Lower-Order Interactions Satisfied for
Higher-Constraint (Item-Specific) Generation Task and Recognition (Item-Specific) Memory Test

			Lower-Order Interactions			**Observed Memory Performance**	
	Number of Interactions Satisfied		**Generation task *match* Memory Test (TAP)**	**Generation task *mismatch* Materials (MAP)**	**Generation task *match* Subject Abilities (SAP)**	**$M_{GenEffect}$ *(SD)***	***n***
Higher	**3**		✓ Item Recognition	✓ Related List	✓ High Eng. Vocab	.301 (.236)	61
Predicted Memory Performance	**2**	b	✓ Item Recognition	☒ Unrelated List	✓ High Eng. Vocab	.295 (.223)	59
	2	c	✓ Item Recognition	✓ Related List	☒ Low Eng. Vocab	.308 (.238)	55
Lower	**1**	c	✓ Item Recognition	☒ Unrelated List	☒ Low Eng. Vocab	.238 (.257)	53

Note. $M_{\text{GenEffect}}$ = mean difference between generation task and read control task, averaged across *n* subjects. *SD* = Standard deviation across *n* subjects. *n* = number of subjects contributing to this row. "Observed memory performance" represents the proportion of correctly recognized "old" items (hits), minus the proportion of "new" distractor items incorrectly identified as "old" (false alarms).

TABLE VI
Sample Demographics by Materials (Related List Structure, Unrelated List Structure) and Subject Ability (High Eng. Vocab, Low Eng. Vocab)

	Related List Structure		**Unrelated List Structure**	
	High Eng. Vocab	**Low Eng. Vocab**	**High Eng. Vocab**	**Low Eng. Vocab**
n	61	55	59	53
Females	40	33	40	43
% Native English Speaker	70.5%	53.7%	66.1%	47.2%
% Bilingual	24.6%	34.5%	28.8%	26.4%

TABLE VII
Means and Standard Deviations of Sample Characteristics
by Materials (Related List Structure, Unrelated List Structure) and Subject Ability (High Eng. Vocab, Low Eng. Vocab)

	Related List Structure		**Unrelated List Structure**	
	High Eng. Vocab	**Low Eng. Vocab**	**High Eng. Vocab**	**Low Eng. Vocab**
n	61	55	59	53
Age	19.08 (1.92)	18.93 (1.97)	19.39 (2.25)	18.85 (1.89)
Shipley Vocab (MAX = 40)	29.00 (2.37)[a,b]	21.27 (4.44)[a,d]	28.66 (2.53)[c,d]	22.64 (2.18)[b,c]
Raney Vocab (MAX = 30)	17.57 (2.66)[f,g]	9.49 (2.20)[f,i]	18.68 (3.25)[h,i]	10.64 (2.60)[g,h]
Vocab Total (MAX = 70)	46.57 (3.67)[j,k]	30.76 (5.64)[j,m]	47.34 (4.75)[l,m]	33.28 (3.55)[k,l]
Working Memory (MAX = 9)	6.13 (1.20)	5.56 (1.34)	6.09 (1.19)	5.71 (1.16)

Note. Cells within a row with matching superscript letters are significantly different ($p < .001$). Vocab Total = Shipley Vocab + Raney Vocab. Working memory = mean span length on a backward digit span task (Killion et al., 2004; Woods et al., 2011). Differences in working memory between groups were not significant ($p > .05$).

TABLE VIII

Means and Standard Deviations of Recognition Memory Test Response Rates of Studied and Lure Items by Encoding Task Condition and Subject Ability Group

High English Ability

Studied Items			**Lure Items**		
Task	**Old (Hit)**	**New (Miss)**	**Task**	**Old (False Alarm)**	**New (Correct Reject)**
Generate	.833 (.373)	.167 (.373)	**Generate**	.144 (.367)	.856 (.352)
Scramble	.834 (.372)	.166 (.372)	**Scramble**	.165 (.371)	.835 (.371)
Read	.473 (.499)	.527 (.499)	**Read**	.102 (.303)	.898 (.303)
			New (unrelated lures)	.098 (.297)	.902 (.297)

Low English Ability

Studied Items			**Lure Items**		
Task	**Old (Hit)**	**New (Miss)**	**Task**	**Old (False Alarm)**	**New (Correct Reject)**
Generate	.795 (.404)	.205 (.404)	**Generate**	.177 (.382)	.823 (.382)
Scramble	.843 (.364)	.157 (.364)	**Scramble**	.196 (.397)	.804 (.397)
Read	.520 (.500)	.480 (.500)	**Read**	.147 (.354)	.853 (.354)
			New (unrelated lures)	.139 (.346)	.861 (.346)

Note. Means are collapsed across Materials factor (related list structure, unrelated list structure).

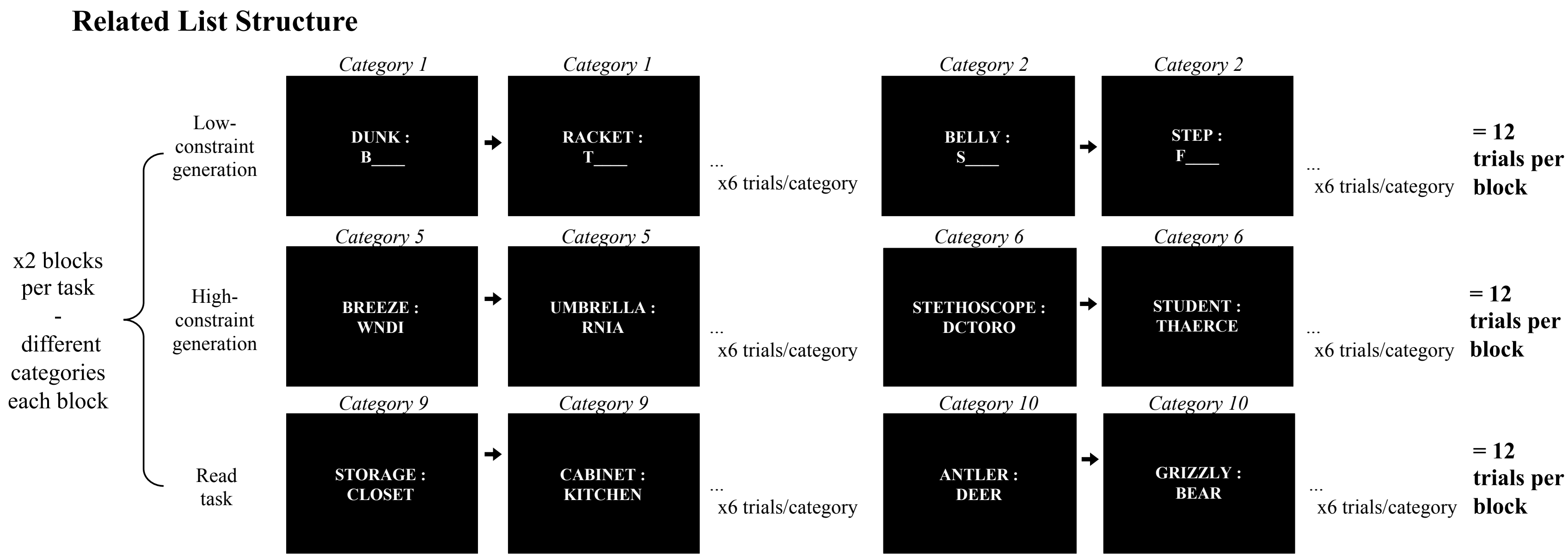

Figure 1. Trial schematic for the related list structure.

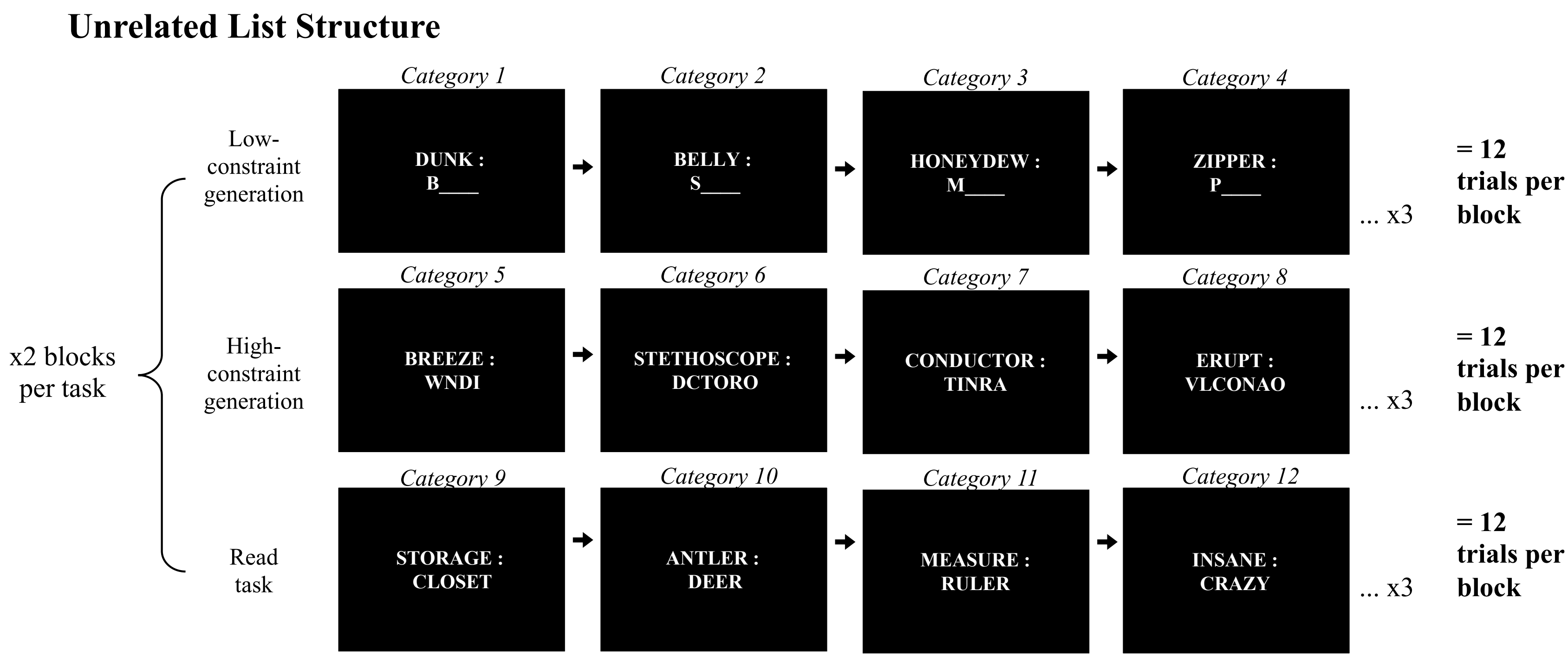

Figure 2. Trial schematic for unrelated list structure.

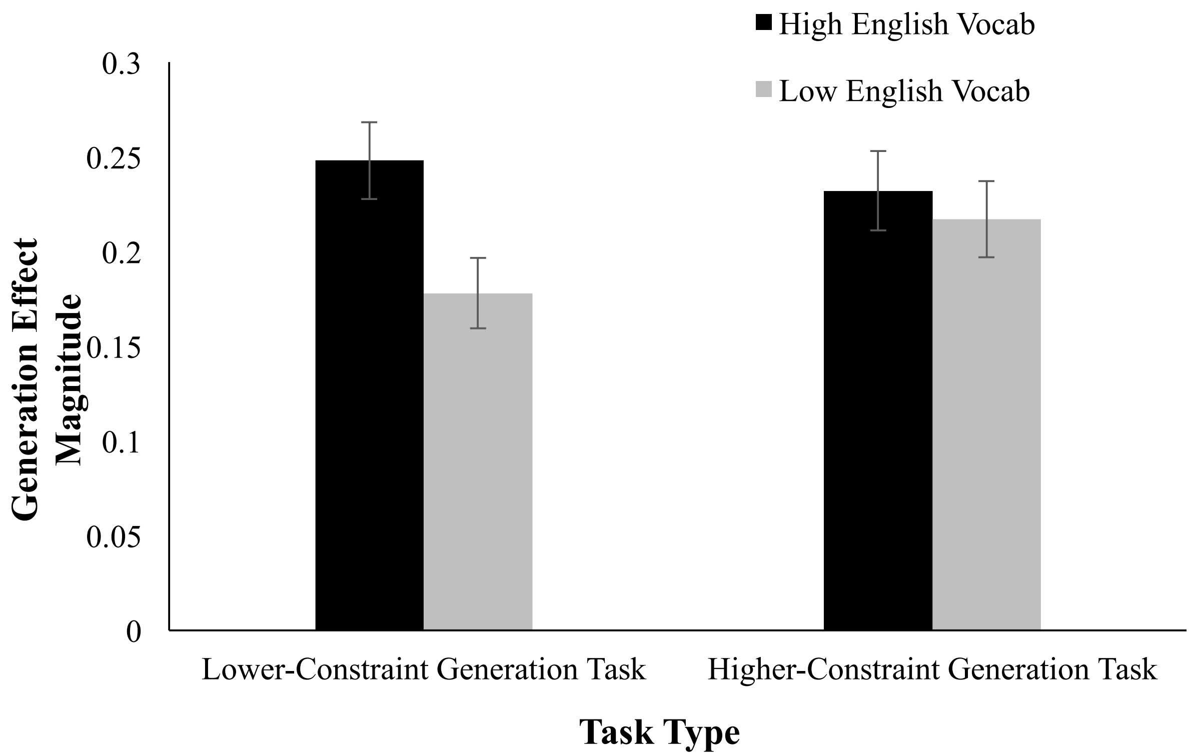

Figure 3. Generation effect magnitude by type of generation task and subject ability. Graph shows the significant interaction between task type and subject ability, highlighting the influence of subject appropriate processing (SAP) in the data. The difference between High English Vocab and Low English Vocab is significant for the lower-constraint generation task ($p < .05$). "Generation Effect Magnitude" is the percent memory improvement above the read control task, collapsed across free recall and recognition memory data. Error bars represent standard error of the mean.

CURRICULUM VITAE

MATTHEW P. MCCURDY

University of Illinois at Chicago || Department of Psychology
1007 W Harrison St. (M/C 285), Chicago, IL, 60607
mmccur3@uic.edu || 765-894-3066 || www.matthewmccurdy.wix.com/research

EDUCATION

University of Illinois at Chicago
Ph.D. Candidate, Cognitive Psychology August 2014 – Present
Dissertation: Developing and Testing a Contextual Theory of the Generation Effect
Advisor: Eric D. Leshikar, Ph.D.

M.A. Cognitive Psychology December 2016
Thesis: Fewer Generation Constraints Enhances the Generation Effect in Younger, but not Older Adults

The University of Tennessee at Chattanooga
M.S. Experimental Psychology August 2012 – May 2014
Thesis: Factors Predicting Depression Scores in End-Stage Renal Disease Patients
Advisor: Irene N. Ozbek, Ph.D.

Indiana University Bloomington
B.S. Psychology August 2008 – May 2012

AWARDS AND HONORS

Leonard Eron Award for Outstanding Scholarly Achievement	May 2020
Paul D. Doolen Scholarship for the Study of Aging (Alternate)	October 2019
Dallas Aging and Cognition Conference Travel Award	January 2019
Psychonomics Graduate Travel Award	November 2018
Psychonomics Society International Graduate Accommodation Award	May 2018
APA Division 20 (Adult Development and Aging) -- Featured Student	Fall 2017
APA Division 20 (Adult Development and Aging) Best Poster Award	August 2017
Christopher Keys Early Outstanding Research Achievement	April 2017
MPA Graduate Student Paper Award	April 2017
Sallie P. Asche Conference Travel Award	January 2017
UIC Provost's Award for Graduate Research	May 2016 – 2017
Psychonomics Society International Graduate Accommodation Award	May 2016

PEER-REVIEWED PUBLICATIONS

McCurdy, M.P., Viechtbauer, W., Frankenstein, A.N., Sklenar, A.M., & Leshikar, E.D. (In press) Theories of the Generation Effect and the Impact of Generation Constraint: A Meta-Analytic Review. *Psychonomic Bulletin and Review*.

Frankenstein, A.N., **McCurdy, M.P.**, Sklenar, A.M., Pandya, R., Szpunar, K.K., & Leshikar, E.D. (In press). Future thinking about social targets: The influence of prediction outcome on memory. *Cognition*.

McCurdy, M.P., Sklenar, A.M., Frankenstein, A.N., & Leshikar, E.D. (2020) Fewer generation constraints increase the generation effect for Item and Source Memory through Enhanced Relational Processing. *Memory*, 1-19.

McCurdy, M.P., Leach, R.C., & Leshikar, E.D. (2019). Fewer generation constraints enhance the generation effect for source memory in younger, but not older adults. *Open Psychology*, *1*(1), 168-184.

Leach, R.C., **McCurdy, M.P.**, Trumbo, M.C., Matzen, L.E., & Leshikar, E.D. (2018). Differential Age Effects of Transcranial Direct Current Stimulation on Associative Memory. *The Journals of Gerontology: Series B*, gby003-gby003. doi:10.1093/geronb/gby003

McCurdy, M.P., Leach, R.C., & Leshikar, E.D. (2017). The generation effect revisited: Fewer generation constraints enhances item and context memory. *Journal of Memory and Language*, *92*, 202-216.

Leshikar, E.D., Leach, R.C., **McCurdy, M.P.**, Trumbo, M.C., Sklenar, A.M., Frankenstein, A.N., & Matzen, L.E. (2017). Transcranial direct current stimulation of dorsolateral prefrontal cortex during encoding improves recall but not recognition memory. *Neuropsychologia, 106*, 390-397.

Leach, R.C., **McCurdy, M.P.**, Trumbo, M.C., Matzen, L.E., & Leshikar, E.D. (2016). Transcranial stimulation over the left inferior frontal gyrus increases false alarms in an associative memory task in older adults. *Healthy Aging Research*, *5*, 1-6.

MANUSCRIPTS UNDER REVIEW OR IN PREPARATION

(* indicates undergraduate mentee)

McCurdy, M.P., Frankenstein, A.N., Sklenar, A.M., Urban Levy, P., & Leshikar, E.D. (Under review) Examining the relationship between generation constraint and memory.

McCurdy, M.P. & Leshikar, E.D. (In preparation). Contextualism and Memory: Developing a Contextual Theory of the Generation Effect.

McCurdy, M.P., Pandya, R., Leach, R.C., & Leshikar, E.D. (In preparation). Adaptive memory: Test of the future simulation hypothesis.

*Giannakopoulos, K.L., **McCurdy, M.P.**, Sklenar, A.M., Frankenstein, A.N., & Leshikar, E.D. (Under review). Lower Constraint Testing Increases the Testing Effect.

Urban Levy, P., Frankenstein, A. N., Sklenar, A.M., **McCurdy, M.P.**, & Leshikar, E.D. (In preparation). A memory advantage for prosocial behaviors.

*Villasenor, J., Sklenar, A.M., Frankenstein, A. N., Urban Levy, P., **McCurdy, M.P.**, & Leshikar, E.D. (In preparation). Value directed memory effects on item and context memory.

Frankenstein, A., **McCurdy, M.P.**, Sklenar, A. M., Urban-Levy, P., & Leshikar, E. D. (In preparation). The relationship between retrieval practice, self-efficacy, and memory.

Sklenar, A.M., Pérez, J., **McCurdy, M.P.**, Frankenstein, A.N., & Leshikar, E.D. (Revise and resubmit). Similarity to the self influences memory for social targets. Manuscript in revisions at *Journal of Experimental Social Psychology*.

Sklenar, A.M., Perez, J., **McCurdy, M.P.**, Frankenstein, A.N., Motyl, M., & Leshikar, E.D. (Under review). Person Memory Mechanism Underlying Approach and Avoidance Judgments of Social Targets.

Ilenikhena, G., *Narmawala, H., Sklenar, A.M., **McCurdy, M.P.**, Gutchess, A., & Leshikar, E.D., (Under review). STOP SHOUTING AT ME: The influence of case and self-referencing on explicit and implicit memory. Manuscript submitted for publication at *Psychonomic Bulletin and Review*.

*Meyers, Z.R., **McCurdy, M.P.**, Leach, R.C., & Leshikar, E.D. (Under review). Effects of survival processing on item and context memory: Enhanced memory for survival-relevant details. Manuscript submitted for publication at *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

INVITED TALKS

McCurdy, M.P. (September, 2019) *Improving the Generation Effect Through Fewer Generation Constraints: The Role of Relational Processing and Aging.* Talk given at the Kable Lab, University of Pennsylvania, Philadelphia, PA.

McCurdy, M.P. (March, 2017). *Fewer generation constraints improve memory benefits from self-generation.* Talk given at Association for Neuropsychological Student Training (ANST) meeting, University of Illinois at Chicago Chapter.

McCurdy, M.P. (February, 2017). *Generation constraint and its impact on the generation effect in older adults*. Talk given at Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign.

E-BOOK CHAPTERS

McCurdy, M.P. (2017). ANOVA (afex): Within Subjects and Mixed Designs. In Demos, A. & Salas, C. (Eds.), *A Language, not a Letter: Learning Statistics in R* (Chapter 15). Available at: http://ademos.people.uic.edu/index.html.

SCHOLARLY PRESENTATIONS

(* indicates undergraduate mentee)

McCurdy, M.P., Frankenstein, A.N., Sklenar, A.M., Urban Levy, P., & Leshikar, E.D. (May 2020). *Experimental Context Accounts for Variations in the Generation Effect: Developing and Testing a Contextual Theory*. Poster to be presented at the Association for Psychological Science Annual Convention, Chicago, IL, USA

McCurdy, M.P., Sklenar, A.M., Frankenstein, A.N., Urban Levy, P., & Leshikar, E.D. (April 2020). *A Contextual Model Accounting for Variations in the Generation Effect*. Talk to be given at the Midwestern Psychological Association Annual Meeting, Chicago, IL, USA.

Frankenstein, A. N., **McCurdy, M. P.**, Sklenar, A. M., Urban Levy, P., & Leshikar, E. D. (April 2020). *Does self-efficacy underlie the memory benefit of retrieval practice?* Talk to be given at the Annual Meeting of the Midwestern Psychological Association, Chicago, IL, USA.

*Giannakopoulos, K.L., **McCurdy, M.P.**, Sklenar, A.S., Frankenstein, A.N., Urban Levy, P., Leshikar, E.D. (March 2020). *Lower Constraint Testing Enhances the Testing Effect for Context Memory for Location but Not Color.* Poster presented at the UIC Psychology Department Cross Program Conference, Chicago, IL, USA.

Frankenstein, A. N., **McCurdy, M. P.**, Sklenar, A. M., & Leshikar, E.D. (February 2020). *How is self-efficacy related to the retrieval practice effect?* Poster presented at the Society for Personality and Social Psychology Annual Convention, New Orleans, LA, USA.

Sklenar, A. M., **McCurdy, M. P.**, Frankenstein, A. N., Urban Levy, P., & Leshikar, E.D. (February 2020). *Memory for impressions based on traits and beliefs affects approach/avoidance decisions.* Poster presented at the Society for Personality and Social Psychology Annual Convention, New Orleans, LA, USA.

McCurdy, M.P., Frankenstein, A.N., Sklenar, A.M., Urban Levy, P., & Leshikar, E.D. (November 2019). *Developing and Testing a Contextual Theory of the Generation Effect*. Poster presented at the Annual Meeting of the Psychonomics Society, Montreal, QC, Canada.

Frankenstein, A.N., **McCurdy, M.P.**, Sklenar, A.M., Urban Levy, P., & Leshikar, E.D. November 2019). *Does self-efficacy contribute to the retrieval practice effect?* Poster presented at the Annual Meeting of the Psychonomic Society, Montreal, Quebec, Canada.

Sklenar, A.M., **McCurdy, M.P.**, Frankenstein, A.N., Urban Levy, P., & Leshikar, E.D. (November 2019). *Differences in the effect of self-similarity of behaviors and beliefs on impression memory.* Poster presented at the Annual Meeting of the Psychonomic Society, Montreal, Quebec, Canada.

McCurdy, M.P., Viechtbauer, W., Frankenstein, A.N., Sklenar, A.M., & Leshikar, E.D. (April 2019). *A Meta-Analytic Review of the Theories of the Generation Effect and the Impact of Generation Constraint*. Talk given at the Midwest Psychological Association Annual Meeting, Chicago, IL, USA.

*Giannakopoulos, K.L., **McCurdy, M.P.,** Sklenar, A.M., Frankenstein, A. N., Leshikar, E.D. (April, 2019). *Low Constraint Testing Heightens the Testing Effect.* Poster presented at the Midwest Psychological Association Annual Meeting, Chicago, IL, USA.

Sklenar, A.M., **McCurdy, M.P.,** Frankenstein, A. N., Leshikar, E.D. (April, 2019). *Person Memory Mechanism Influences Decision to Approach/Avoid Others.* Talk given at Midwest Psychological Association Annual Meeting, Chicago, IL, USA.

Sklenar, A.M., **McCurdy, M.P.,** Frankenstein, A. N., Leshikar, E.D. (February, 2019). *Person Memory Mechanism Contributes to Approach and Avoidance Judgments.* Poster presented at the Annual Meeting of the Society for Personality and Social Psychology, Portland, OR, USA.

Frankenstein, A. N., **McCurdy, M.P.,** Sklenar, A.M., & Leshikar, E.D. (February, 2019). *That Was Unexpected: How Expectancy Confirmation and Violation Affect Memory for Social Behaviors*. Poster presented at the Annual Meeting of the Society for Personality and Social Psychology, Portland, OR, USA.

McCurdy, M.P., Viechtbauer, W., Frankenstein, A.N., Sklenar, A.M., & Leshikar, E.D. (January 2019). *Fewer Generation Constraints Enhance Memory for Self-Generated Information Differently for Younger Adults and Older Adults*. Poster presented at the Dallas Aging and Cognition Conference, Dallas, TX, USA.

McCurdy, M.P., Viechtbauer, W., Frankenstein, A.N., Sklenar, A.M., & Leshikar, E.D. (November 2018). *Theories of the Generation Effect and the Impact of Generation Constraint: A Meta-Analytic Review*. Poster presented at the Annual Meeting of the Psychonomics Society, New Orleans, LA, USA.

Sklenar, A.M., **McCurdy, M.P.,** Frankenstein, A. N., Leshikar, E.D. (November, 2018). *Belief Similarity Interferes with Self-Similarity Memory Effect.* Poster presented at the Annual Meeting of the Psychonomics Society, New Orleans, LA, USA.

Frankenstein, A. N., **McCurdy, M.P.,** Sklenar, A.M., & Leshikar, E.D. (November, 2018). *Predicting Social Behaviors: Do We Remember Better When We're Wrong?* Poster presented at the Annual Meeting of the Psychonomics Society, New Orleans, LA, USA.

McCurdy, M.P., Frankenstein, A.N., Sklenar, A.M., Leshikar, E.D. (May, 2018). *Improving the Generation Effect: Fewer Generation Constraints enhance Relational Processing.* Poster presented at the International Psychonomics Society Annual Meeting, Amsterdam, Netherlands.

Sklenar, A.M., **McCurdy, M.P.,** Frankenstein, A. N., Leshikar, E.D. (May, 2018). *The Effect of Similarity to the Self on Impression Memory. P*oster presented at the International Psychonomics Society Annual Meeting, Amsterdam, Netherlands.

McCurdy, M.P. & Leshikar, E.D. (April 2018). *Improving the generation effect through fewer generation constraints.* Talk given at the Midwest Psychological Association Annual Meeting, Chicago, IL.

Sklenar, A.M., Motyl, M., **McCurdy, M.P.,** Frankenstein, A. N., Leshikar, E.D. (April, 2018). *Memory for impressions of social targets influences subsequent social distance judgments.* Poster presented at the Midwest Psychological Association Annual Meeting, Chicago, IL.

*Giannakopoulos, K.L., **McCurdy, M.P.,** Sklenar, A.M., Frankenstein, A. N., Leshikar, E.D. (April, 2018). *Enhanced Testing Effect Through Fewer Test Constraints.* Poster presented at the Student Research Forum, UIC, Chicago, IL.

Frankenstein, A. N., Motyl, M., **McCurdy, M.P.,** Sklenar, A.M., Leshikar, E.D. (April, 2018). *Prediction Outcome and Memory for Consistent versus Inconsistent Social Behaviors.* Poster presented at the Midwest Psychological Association Annual Meeting, Chicago, IL.

*Giannakopoulos, K.L., **McCurdy, M.P.,** Sklenar, A.M., Frankenstein, A. N., Leshikar, E.D. (March, 2018) *Improving the Testing Effect Through Fewer Test Constraints.* Poster presented at the Cross Program Conference, Department of Psychology, University of Illinois at Chicago, Chicago, IL.

Frankenstein, A. N., **McCurdy, M.P.,** Sklenar, A.M., Leshikar, E.D. (March, 2018). *Effects of trait diagnosticity and prediction outcome on Memory.* Poster presented at the Society for Personality and Social Psychology's Annual Meeting, Atlanta, GA.

Sklenar, A.M., Motyl, M., **McCurdy, M.P.,** Frankenstein, A. N., Leshikar, E.D. (March, 2018). *Memory for impressions of others influence social distancing judgments.* Poster presented at the Society for Personality and Social Psychology's Annual Meeting, Atlanta, GA.

McCurdy, M.P. & Leshikar, E.D. (March 2018). *Assessing the influence of generation constraints on the generation effect.* Talk given to the Cognitive Program, UIC Department of Psychology.

McCurdy, M.P., Frankenstein, A.N., Sklenar, A.M., & Leshikar, E.D. (November, 2017). *Fewer generation constraints enhance the generation effect vie cue-target associative strengthening.* Poster presented at the Annual Meeting of the Psychonomics Society, Vancouver, B.C., Canada.

Sklenar, A.M., **McCurdy, M.P.,** Frankenstein, A. N., Leshikar, E.D. (November, 2017). *Extending the self-similarity effect in impression memory.* Poster presented at the Psychonomics Society Annual Meeting, Vancouver, BC.

Frankenstein, A. N., **McCurdy, M.P.,** Sklenar, A.M., & Leshikar, E.D. (November, 2017). *Future thinking: Influence of prediction outcome on memory.* Poster presented at the Psychonomics Society Annual Meeting, Vancouver, BC.

McCurdy, M.P., Leach, R., & Leshikar, E.D. (August, 2017). *Fewer generation constraints impacts the generation effect for younger but not older adults*. Poster presented at the American Psychological Association Annual Convention, Washington, D.C.

McCurdy, M.P., Leach, R., & Leshikar, E.D. (April, 2017). *Fewer generation constraints enhances the generation effect*. Poster presented at the Midwest Psychological Association Annual Meeting, Chicago, IL.

Leach, R., **McCurdy, M.P.,** & Leshikar, E.D. (April, 2017). *Transcranial stimulation effects on face-name associative memory*. Poster presented at the Midwest Psychological Association Annual Meeting, Chicago, IL.

*Meyers, Z., **McCurdy, M.P.,** Leach, R., & Leshikar, E.D. (April, 2017). *Survival processing: Item and context memory enhancement*. Poster presented at the Student Research Forum, Chicago, IL.

*Meyers, Z., **McCurdy, M.P.,** Leach, R., & Leshikar, E.D. (April, 2017). *Effects of survival processing on context memory*. Poster presented at the Midwest Psychological Association Annual Meeting, Chicago, IL.

McCurdy, M.P. (March, 2017). *Fewer generation constraints improve the generation effect: Impacts for item and context memory, recognition and recall.* Talk given at 2017 Cross Program Conference, University of Illinois at Chicago, Chicago, IL, USA.

McCurdy, M.P., Leach, R.C. & Leshikar, E.D. (January, 2017). *Level of constraint influences the generation effect for item and context memory in younger but not older adults*. Poster presented at the Dallas Aging and Cognition Conference, Dallas, TX, USA.

Leach, R.C., **McCurdy, M.P.,** Trumbo, M.C., Matzen, L.E., & Leshikar, E.D. (January, 2017). *Differential age effects of transcranial stimulation on face-name associative memory*. Poster presented at the Dallas Aging and Cognition Conference, Dallas, TX, USA.

McCurdy, M.P., Leach, R.C. & Leshikar, E.D. (November, 2016). *Fewer constraints enhances the generation effect for context memory: Benefits for Source and Color.* Poster presented at Annual Meeting of the Psychonomics Society, Boston, MA, USA.

McCurdy, M.P., & Leshikar, E.D. (September, 2016). *Fewer constraints enhances the generation effect for context memory.* Talk given to the Cognitive Program, UIC Psychology, Chicago, IL, USA.

McCurdy, M.P., Leach, R.C., & Leshikar, E.D. (May, 2016). *The generation effect revisited: Enhancements for item and context memory.* Poster presented at the International Meeting of the Psychonomics Society, Granada, Spain.

McCurdy, M.P., Leach, R.C., & Leshikar, E.D. (May, 2016). *Enhancing the generation effect: Unconstrained generation improves item and context memory.* Poster presented at the annual meeting of the Association for Psychological Science, Chicago, IL, USA.

Leach, R.C., **McCurdy, M.P**., & Leshikar, E.D. (May, 2016). *Transcranial stimulation of dorsolateral prefrontal cortex improves face-name associative memory*. Poster presented at the annual meeting of the Association for Psychological Science, Chicago, IL, USA.

Perez, J., Leach, R.C., **McCurdy, M.P.,** Motyl, M., Leshikar, E.D. (April 2016). *In search of a person memory mechanism underlying ideological migration*. Poster presented at the UIC Student Research Forum. Chicago, IL, USA.

McCurdy, M.P., Leach, R.C., & Leshikar, E.D. (November, 2015). *Unconstrained generation improves the generation effect: Benefits for item and context Memory.* Poster presented at the Psychonomics Society Annual Meeting, Chicago, IL, USA.

McCurdy, M.P., & Leshikar, E.D. (April, 2015). *Self-Generation benefits item and context memory.* Talk given to the Cognitive Program, UIC Psychology, Chicago, IL, USA.

*Pandya, R., Leach, R., **McCurdy, M.P.**, Leshikar, E.D. (April, 2015). *Adaptive memory: Test of the future simulation hypothesis.* Poster presented at 2015 *Student Research Forum*, Chicago, IL, USA.

McCurdy, M.P., Jones, J., & Ozbek, I.N. (April, 2014*). Depression scores as a predictor of olfactory sensitivity in end-stage renal disease patients*. Poster presented at 2014 *Association of Chemoreception Sciences (AChemS) Annual Meeting*.

Jones, J., **McCurdy, M.P.**, & Ozbek, I.N. (April, 2014). *Variability in olfactory detection thresholds in end-stage renal disease patients*. Poster presented at 2014 *Association of Chemoreception Sciences (AChemS) Annual Meeting*.

DEPARTMENTAL SERVICE

UIC Cognitive Psychology Program
Cognitive Program Graduate Assistant August 2015 – May 2016

RESEARCH EXPERIENCE

Functional Aging Brain Laboratory, University of Illinois at Chicago
Graduate Research Assistant August 2014 – Present

STEM Education, The University of Tennessee at Chattanooga
Graduate Research Assistant August 2012 – May 2014
Supervisors: Sandy Watson, Ed.D.; Steve Kuhn, Ph.D.

Department of Psychology, The University of Tennessee at Chattanooga
Graduate Research Assistant August 2012 – May 2014

Preclinical Neuropsychopharmacology Lab, Indiana University
Undergraduate Research Assistant March 2010 – August 2012
Supervisor: George Rebec, Ph.D.

TEACHING EXPERIENCE

University of Illinois at Chicago
Instructor – Introduction to Psychology 1 term
Teaching Assistant – Introduction to Research Methods 5 terms
Teaching Assistant – Statistical Methods in Behavioral Sciences 3 terms
Teaching Assistant – Cognition and Memory 5 terms
Teaching Assistant – Cognition and Memory Research Lab 2 term
Teaching Assistant – Introduction to Psychology 3 terms

The University of Tennessee at Chattanooga
Instructor – Introduction to Psychology 2 terms
Teaching Assistant – Research Methods Lab 2 terms

PROFESSIONAL DEVELOPMENT

SBSRI Methods Series: Meta-Analysis Workshop — September 2017
University of Illinois at Urbana-Champaign

Longitudinal Data Analysis — Fall 2017
University of Illinois at Chicago – Biostatistics / School of Public Health

Hierarchical Linear Modeling — Spring 2019
University of Illinois at Chicago – School of Education

Research Methods in Learning Sciences — Spring 2019
University of Illinois at Chicago – Learning Science Research Institute

SOFTWARE PROFICIENCIES

R; SAS; SPSS; JASP; E-Prime; MS Office

PROFESSIONAL MEMBERSHIPS

Psychonomic Society
American Psychological Association
APA Division 20: Adult Development and Aging
Association for Psychological Science (APS)
Midwestern Psychological Association (MPA)