2013

Can interpolated testing reduce retrieval-induced forgetting?

Matthew R. Erdman
Iowa State University

Can interpolated testing reduce retrieval-induced forgetting?

by

Matthew R. Erdman

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Psychology

Program of Study Committee:
Jason C. K. Chan, Major Professor
Veronica Dark
Rob West
Fred Lorenz
Gary Phye

Iowa State University

Ames, Iowa

2013

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABSTRACT

Testing can improve retention of tested information (see Roediger & Karpicke, 2006a for review), but it can also impair memory for nontested, related information: an effect termed retrieval-induced forgetting (Anderson, Bjork, & Bjork, 1994). To my knowledge, retrieval-induced forgetting has only been shown in experiments where participants study all exemplars and then perform one retrieval practice phase (see Anderson, 2003; Raaijmakers & Jakab, 2013; Storm & Levy, 2012 for reviews). Researchers have demonstrated that interpolating memory tests during the learning phase can reduce the amount of proactive interference one experiences when one learns new information (Szpunar, McDermott, & Roediger, 2008). In this dissertation, I examined how studying information in multiple blocks rather than one block influenced the buildup of proactive interference. Previous researchers (Szpunar et al., 2008) examined how testing can reduce the buildup of practice interference when all previously studied information was tested. I examined whether testing only a subset of the exemplars in a block would inoculate participants against the buildup of proactive interference. Furthermore, I examined how the presentation order of the Rp- exemplars (i.e., nontested exemplars from the tested categories) relative to the Rp+ (tested) exemplars influenced the magnitude of retrieval-induced forgetting. Retrieval-induced forgetting researchers generally endorse one of two accounts: the inhibition (Anderson, 2003) or blocking account (Raaijmakers & Jakab, 2013). I predicted that if blocking drives retrieval-induced forgetting, I would find retrieval-induced forgetting regardless of presentation order of the Rp- and Rp+ exemplars. However, if inhibition drives retrieval-induced forgetting, I would only find retrieval-induced forgetting in conditions where Rp-exemplars were presented prior to retrieval practice. In three experiments, to-be-learned information was presented either in one block followed by retrieval practice (the cumulative retrieval practice condition), or presented in four blocks. For the latter participants, some were given math instead of retrieval practice between each block (the interim math condition), some participants studied the Rp- exemplars in blocks prior to the presentation and retrieval practice of the Rp+ exemplars (the high competition condition), and the remaining participants studied and received retrieval practice over the Rp+ exemplars before learning the Rp- exemplars (the low competition condition). Then participants either had a 20-minute (Experiments 1 and 3) or 10-minute distractor period (Experiment 2). Finally, participants either were given a category cued recall test (Experiment 1) or a category-plus-stem cued recall test (Experiments 2 and 3). In Experiment 1, participants demonstrated retrieval-induced forgetting in the cumulative retrieval practice, high competition, and low competition conditions. However, in Experiment 2, participants only demonstrated retrieval-induced forgetting in the cumulative retrieval practice and high competition conditions. The results of Experiment 3 were inconsistent with Experiments 1 and 2 and may have been due to chance. The current dissertation provides evidence for both the inhibition and blocking accounts.

CHAPTER 1. INTRODUCTION

Testing can be a powerful learning tool. Roediger, Putnam, and Smith (2011) discussed 10 different benefits of testing. The first, which they described as the direct benefit of testing, refers to the finding that people often remember more information when tested over that information compared to when they do not take a test (Glover, 1989; Spitzer, 1939) or when they restudy the information (Carpenter, Pashler, & Vul, 2006; Nungester & Duchastel, 1982; Roediger & Karpicke, 2006a). This testing effect has been demonstrated with word pairs (Carpenter et al., 2006; Carpenter, Pashler, Wixted, & Vul, 2008; Toppino & Cohen, 2009), sentences (Chan, 2009), prose (Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008; Chan, 2009; Chan, McDermott, & Roediger, 2006), and visuospatial materials (Carpenter & Pashler, 2007).

Along with this direct benefit, testing also produces indirect benefits. That is, testing can identify gaps in knowledge, improve learning in the next study period, improve the organization of knowledge, improve transfer of knowledge to a nontested domain, improve memory for nontested material, improve metacognitive monitoring, provide feedback to instructors, encourage students to study, and reduce proactive interference (Roediger et al., 2011). Of these indirect benefits, the most relevant to the current dissertation is that testing can reduce proactive interference during learning (see Szpunar, McDermott, & Roediger, 2008). Due to the myriad benefits of testing on memory retention, researchers have advocated for increased use of testing in classrooms (Leeming, 2002; McDaniel, Anderson, Derbish, & Morrisette, 2007; Roediger & Karpicke, 2006b). Despite the many positive influences of testing, one negative consequence is that testing can impair memory of nontested but related information – an effect named retrieval-induced forgetting.[1]

In a typical study of its kind (see Anderson, Bjork, & Bjork, 1994 for an example), participants study category-exemplar pairs (e.g., Weather – Blizzard, Weather – Monsoon, Body – Arm, Body – Leg) for a later memory test. Afterwards, participants take an initial test on some of the word pairs (e.g., Weather – Blizzard), perform a distractor task during a retention interval, and then they are tested on all the word pairs (e.g., Weather – Blizzard, Weather – Monsoon, Body – Arm, Body – Leg). Interestingly, researchers often find that the nontested exemplars that are related to the tested ones are recalled at a lower rate than the unrelated exemplars. This is known as the retrieval-induced forgetting effect. A fuller understanding of the positive and negative effects of testing on memory performance is important to maximize the potential benefits of applying frequent testing to education (see Leeming, 2002; McDaniel et al., 2007; and Roediger & Karpicke, 2006b).

To my knowledge, studies that have demonstrated retrieval-induced forgetting have always employed a single retrieval practice phase after all the word pairs have been studied. That is, participants typically study all the to-be-remembered materials before

[1] Though Roediger et al. (2011) suggested that retrieval-induced forgetting was a negative side effect of testing, other researchers have argued that retrieval-induced forgetting may be beneficial – it may help us forget information that is no longer relevant (for more information, see Anderson & Levy, 2009; Harris, Sutton, & Barnier, 2010; Schacter, 2001).

they perform retrieval practice over a subset of these materials. However, researchers have not examined the impact of performing multiple retrieval practice phases (each with their own unique set of tested word pairs) on final test performance. Instead of studying all word pairs and then performing retrieval practice on them, an alternative is that one studies one-quarter of the pairs, performs retrieval practice over some of those pairs, studies a different quarter of the word pairs, performs retrieval practice over those pairs, and so on.

This method might better mirror the way retrieval practice is implemented in the real world. For example, a student may be required to read the first three chapters of a book for an upcoming test. The student may read Chapter 1, answer the Chapter 1 review questions, read Chapter 2, answer the Chapter 2 review questions, and so on. Typically, the review questions will not cover all of the material that the student reads in each chapter. Of course, retrieval practice schedules can be even more fine-grained than is described here (e.g., one may try to recall information from a particular section of a chapter before studying the remainder of a chapter, thus breaking down the study and retrieval practice cycles into even smaller chunks). In the current dissertation, I sought to understand how performing interim retrieval practice as opposed to performing cumulative retrieval practice affects memory for the unpracticed materials.

Researchers have also examined how taking interim tests during learning affects subsequent learning (Szpunar et al., 2008; Weinstein, McDermott, & Szpunar, 2011; Wissman, Rawson, & Pyc, 2011). Specifically, interim testing refers to taking a test over the just studied material prior to studying the next set of materials. Interim testing leads to an effect that I will refer to as test-potentiated learning,[2] which is increased memory of new material and reduced prior list intrusions. Here, intrusions refer to recall of materials from an inappropriate source (e.g., recalling a word from List 1 when asked to recall words from only List 2). The direct benefit of testing, improved memory for tested materials, is well established in the lab (see Roediger & Karpicke 2006 for a literature review) and in the classroom (Leeming, 2002; McDaniel, Anderson, Derbish, & Morrisette, 2007; Spitzer, 1939). Nonetheless, educators may still be reluctant to adopt testing because of a fear that only directly tested information will receive memorial benefits. However, there is evidence that testing can benefit nontested information as well. Chan, McDermott, and Roediger (2006) demonstrated that testing can improve memory for nontested, related information.[3] Furthermore, test-potentiated learning

---

[2] This term was borrowed from Izawa (1971). She found that testing can *potentiate* learning. That is, testing can improve learning on the next study trials. Though Izawa noted that test potentiation occurs for the same materials (e.g., taking a test where you see *Weather* – Bl____ and then study *Weather* – Blizzard), the current dissertation examines benefits for new material.

[3] Though this may seem at odds with the retrieval-induced forgetting literature, Chan (2009) identified variables that determine whether testing produces retrieval-induced forgetting or retrieval-induced facilitation. When the final test occurs 24 hours after retrieval practice and related category exemplar pairs were encode on consecutive trials to promote integrative processing, no retrieval-induced forgetting was found. However,

provides another method for improving memory of nontested information. Here, I examined how testing a subset of information over multiple tests influences memory, so the current dissertation has theoretical and practical applications for both retrieval-induced forgetting and test-potentiated learning. In the following sections, I review these two literatures. I then describe the present study in detail.

Empirical review of retrieval-induced forgetting

Retrieval-induced forgetting has been a heavily researched phenomenon (see Anderson, 2003). The effect has been demonstrated using paired associates (Aslan, Bäuml, & Pastötter, 2007; Anderson, Bjork, & Bjork, 1994; Anderson, Bjork, & Bjork, 2000; Anderson & Spellman, 1995; Hicks & Starns, 2004; Verde & Perfect, 2011), sentences (Anderson & Bell, 2001; Chan, 2009, Gómez-Ariza, Lechuga, Pelegrina, & Bajo, 2005; Macrae & MacLeod, 1999), and prose (Carroll, Campbell-Ratcliffe, Murnane, & Perfect, 2007; Chan, 2009; Little, Storm, & Bjork, 2011). Furthermore, retrieval-induced forgetting has been found in recall (Anderson et al., 1994; Chan, 2009; Jakab & Raaijmakers, 2009; Shivde & Anderson, 2011; Storm & Nestojko, 2010) and recognition (Hicks & Starns, 2004; Spitzer & Bäuml, 2009; Verde, 2004; Verde & Perfect, 2011). Typically, retrieval-induced forgetting experiments are completed with word pairs (i.e., Anderson et al., 1994; Jonker, Seli, & MacLeod, 2012; Williams & Zacks), but Garcia-Bajos, Migueles, and Anderson (2009) found retrieval-induced

---

when the final test occurs 20 minutes after retrieval practice and when integration of the exemplars was made difficult, Chan found retrieval-induced forgetting.

forgetting with more complex, video materials in a misinformation effect paradigm (see also MacLeod, 2002).

In their seminal paper, Anderson et al. (1994) asked participants to remember category-exemplar pairs (e.g., *Weather* – Blizzard, *Weather* – Monsoon, *Body* – Arm, and *Body* – Leg). After participants studied the pairs, they were given a category-plus-two letter stem cued recall test (e.g., *Weather* – Bl____) over half of the exemplars in half of the categories (the *Rp+* condition; e.g., *Weather* – Blizzard), whereas half of the exemplars from these categories were not tested (the *Rp-* condition; e.g., *Weather* – Monsoon). All of the exemplars from the remaining categories were not tested (the *Nrp* condition; e.g., *Body* – Arm, *Body* – Leg). After a 20-minute distractor phase, participants were given a category cued recall test over all the exemplars. Specifically, participants were presented with a category cue (e.g., *Body*, *Weather*), and they attempted to recall all studied exemplars that were presented under that category. Anderson et al. found that participants recalled a higher proportion of Rp+ exemplars than Nrp exemplars (a testing effect). Moreover, participants recalled fewer Rp- exemplars than Nrp exemplars, an effect they termed retrieval-induced forgetting. Although several explanations have been offered to explain this phenomenon, researchers generally endorse the inhibition account (Anderson & Levy, 2009; Storm, Bjork, & Bjork, 2008), the blocking account (Camp, Pecher, & Schmidt, 2005; Raaijmakers & Jakab, 2013; Williams & Zacks, 2001), or a combination of the two (Storm & Levy, 2012).

The inhibition account and blocking account make different assumptions about whether memory for the Rp- exemplars is impaired. Inhibition account proponents (Anderson, 2003; Storm, Bjork, & Bjork, 2008) argue that retrieval-induced forgetting results from competition that occurs during retrieval practice. For example, assume that participants study the words "*Weather* – Blizzard, *Weather* – Monsoon, and *Body* – Leg," and perform retrieval practice on "*Weather* – Bl____." Two possible answers (Blizzard and Monsoon) may compete for retrieval because of the presence of the category cue *Weather*. To resolve this competition, "Monsoon" is inhibited to facilitate recall of "Blizzard." Due to this lingering inhibition, participants are less likely to recall "Monsoon" during the final test later. That is, after inhibition, the representation of "Monsoon" is weakened in long-term memory (see Anderson, 2003). Additionally, "Blizzard" is better remembered because it has received retrieval practice.

Blocking account proponents offer a different explanation for the impaired recall of the Rp- exemplars. Like inhibition account advocates, blocking account advocates (Williams & Zacks, 2001) assume that "*Weather* – Blizzard" is better remembered than Nrp exemplars due to retrieval practice. However, blocking account advocates assume that representation of "*Weather* – Monsoon" is not inhibited. Instead, they assume that because "*Weather* – Blizzard" is better remembered than "*Weather* – Monsoon," "Blizzard" will be more easily recalled when prompted with the cue "*Weather"* on the final test. Because "Blizzard" is remembered so well, it interferes with participants' attempts to recall "Monsoon" until either time runs out or the participants abandon their attempt to recall "Monsoon."

There are two main differences between the inhibition and blocking account. Firstly, the inhibition account assumes that the impairment of the Rp- exemplars occurs during retrieval practice, but the blocking account assumes that the impairment occurs during the final test. Secondly, the inhibition account assumes that the Rp- exemplar itself becomes harder to access, but the blocking account assumes that the Rp- information's memory strength is unaffected. Though the two explanations are not mutually exclusive, much research has been conducted to separate the two accounts (Anderson, 2003; Anderson et al., 2000; Storm & Levy, 2012; Verde, 2004; Williams & Zacks, 2001). Between the inhibition and blocking accounts, the latter is considered more parsimonious because it does not require an extra psychological process (i.e., inhibition) to reduce the memory strength of the Rp- exemplar. Therefore, blocking account proponents tend to provide evidence for their account by offering evidence against assumptions made by the inhibition account (see Camp, Pecher, & Schmidt, 2007; Verde & Perfect, 2011; Williams & Zacks, 2001). The major assumptions of the inhibition account include cue independence, retrieval specificity, strength independence, and interference dependence.

Cue independence

One assumption of the inhibition account is cue independence, which refers to the idea that retrieval inhibition causes a reduction of memory strength of the exemplar itself rather than a reduction of the strength of the association between the category and the exemplar. Because the exemplar itself is less accessible, it should demonstrate impairment relative to an Nrp exemplar regardless of how the exemplars are cued. On the contrary, the blocking account would lead one to predict that retrieval-induced forgetting is contingent on presentation of the studied category cue (or the ability to use a final test prompt to recall the originally studied category) that allows the Rp+ exemplar to intrude when one attempts to recall the Rp- exemplars.

Initially, researchers used cross-category cues (see Anderson & Spellman, 1995) to demonstrate cue independence. That is, half of the Nrp exemplars shared characteristics with the Rp+ exemplars. These Nrp exemplars, along with Rp- exemplars, demonstrated retrieval-induced forgetting. For example, participants might have studied "*Red* – Blood, *Red* – Tomato, *Food* – Strawberry, and *Food* – Crackers." Participants then performed retrieval practice over "*Red* – Tomato" (but not over any of the *Food* items). Notably, "*Red* – Blood" (a Rp- exemplar) and "*Food* – Strawberry" (a Nrp exemplar that shared characteristics with a Rp+ exemplar, namely the color red) were recalled at a lower rate than "*Food* – Crackers" (a Nrp exemplar that did not share characteristics with a Rp+ exemplar). This experiment thus demonstrated cross-category retrieval-induced forgetting, which violates assumptions made by the blocking account because presentation of the category name "*Food*" should not allow the strengthened exemplar "Blood" to block retrieval of "Strawberry."

More recently, researchers have used item-specific extralist cues to demonstrate cue-independence. That is, instead of using studied category names, an unstudied category cue is supplied to probe recall. Aslan et al. (2007) provided a straightforward demonstration of item specific independent cues. Participants studied category-exemplar pairs and then performed category-plus-stem cued recall over the Rp+ exemplars. For example, participants might have studied a set of exemplars that included "*Weapon* – Grenade, *Weapon* – Knife, *Fruit* – Orange, and *Fruit* – Banana." Participants then performed retrieval practice over some exemplars (e.g., *Weapon* – Gr____). After a one-minute distractor task, memory for all studied exemplars was prompted by new, extralist cues. For example, to cue "Grenade" on the final test, participants were shown "*Loud* -

Gr____." Despite the absence of the studied category name (e.g., "Weapon"), recall probability of the Rp- exemplars was lower than that of the Nrp exemplars.

Camp, Pecher, Schmidt, and Zeelenberg (2009) have challenged whether extralist cues provide unequivocal evidence for cue independent forgetting. First, Camp et al. asked participants to study single words that will later appear as cues in cue-target word pairs (the *initial study phase*; e.g., *Rope*). For half of the cue – target pairs, the cue was presented in the initial study phase, while the other half of the cues were not presented. After this initial study phase, participants studied cue-target pairs like "Rope – Sailing" and "Sunflower – Yellow" (the *second study phase*). Notably, the cues were semantically related to the targets. Finally, participants were given an item specific independent cue final test. For example, the participants might see *"Sport –* ___" to cue "Sailing." Camp et al. found that participants remembered more targets if they had studied the cues (e.g., Rope) in the initial study phase.

The researchers argued that if participants used independent probes (e.g., Sport) to recall the initially studied cues (e.g., Rope) before they attempted to recall the targets (rather than using independent cues to directly recall studied targets), then participants should have remembered more targets that were related to cues presented in the initial study phase. If participants did not use the extralist cue (e.g., *Sport*) to recall the initial cue (e.g., *Rope*), the number of times that the initial cue (e.g., *Rope*) was presented should be irrelevant to final test performance. When this logic is applied to retrieval-induced forgetting, it suggests that independent cue retrieval-induced forgetting could occur because the extra list cue activates the studied category cue, which leads to intrusion of the Rp+ exemplars. This intrusion of the Rp+ exemplar then could block retrieval of the

Rp- exemplars. In other words, blocking proponents could argue that participants are likely using the independent cue to recall the initially studied category cue and then using the initial category cue to recall a specific exemplar. Therefore, blocking account proponents argue that cue independence does not necessarily violate the interference account.

Recently, Huddleston and Anderson (2012) have challenged Camp et al.'s (2009) critique of cue independence. Specifically, when Huddleston and Anderson used materials in which the independent probe and initial cue were not semantically related, no difference in recall probabilities was observed regardless of the number of times a cue was studied. Specifically, Huddleston and Anderson asked participants to judge the relatedness of the independent cues and initially studied categories using materials from Camp et al. and Anderson and Green (2001). Participants judged Camp et al.'s independent cues and categories to be more related than Anderson and Green's cues and categories. Furthermore, Huddleston and Anderson found that when they used Anderson and Green's materials, the recall probability of exemplars did not increase when participants had repeatedly studied the initial category (as opposed to what Camp et al. found). Therefore, Huddleston and Anderson concluded that there was no evidence that participants used the independent cue to recall the initially studied category and argue that cue independence still provides evidence for inhibition.

Inhibition account proponents have argued that retrieval-induced forgetting with a recognition final test provides another example of cue independence (Anderson, 2003; Aslan and Bäuml, 2011; Levy & Anderson, 2012; Storm & Levy, 2012). In a study by Hicks and Starns (2004), participants studied category-exemplar word pairs and then performed category-plus-stem cued recall over the Rp+ exemplars. After a 10-minute distractor task, participants performed an old-new recognition test over all the exemplars and distractors. Hicks and Starns found that participants correctly recognized Nrp exemplars more often than Rp- exemplars – a retrieval-induced forgetting effect.

This finding provides support for the inhibition account, which assumes that the representation of the Rp- exemplars is weakened by retrieval practice of related items.[4] The blocking account assumes that other exemplars interfere with recall of the Rp- exemplars. If the Rp- exemplar is presented for recognition, it should be readily accessible – no other exemplars should interfere with one's ability to remember an exemplar that is displayed to that individual. Consequently, one would predict retrieval-induced forgetting in recognition if one subscribed to the inhibition (but not blocking) account. Retrieval-induced forgetting in recognition has been reported numerous times (e.g., Spitzer & Bäuml, 2009; Verde, 2004; Verde & Perfect, 2011; but see Koutstaal, Schacter, Johnson, & Galuccio, 1999 for failure to find retrieval-induced forgetting in recognition). Although independent cue retrieval-induced forgetting is not always replicable (Camp, Pecher, & Schmidt, 2007; Williams & Zacks, 2001), it continues to be considered a strong piece of evidence for the inhibition account.

---

[4] Inhibition account supporters argue that the representation of a word is inhibited. However, this does not mean that every meaning of a single word will be inhibited. A copy cue presented during the recognition test may not map perfectly onto the memory representation that is established during the time of encoding (see Thomson & Tulving, 1973).

Nonetheless, some researchers have questioned whether recognition is definitive evidence of cue independence (Verde & Perfect, 2011). Explicitly, Verde and Perfect found retrieval-induced forgetting in recognition when the test was self-paced. However, the effect disappeared when participants were forced to respond to each exemplar within 750 ms. Verde and Perfect argued that if the Rp- exemplar is less accessible due to inhibition, retrieval-induced forgetting should be found in a recognition test regardless of time constraints.

Recently, Raaijmakers and Jakab (2012) and Verde (2012) have argued that some interference memory models can explain retrieval-induced forgetting in recognition. Their models assume that Rp+ exemplars can interfere with Rp- exemplar recognition because information related to the Rp- exemplar will receive activation during the final recognition test. That is, when a participant is probed about a Rp- exemplar (e.g., Monsoon), related concepts are automatically activated, which will likely include the Rp+ exemplars from that category (e.g., Blizzard), thus causing interference.[5] Clearly, whether or not recognition is evidence of independent cue retrieval-induced forgetting is still under debate.

Though cue independence has been contested (Camp et al., 2007; Camp et al., 2009; Raaijmakers & Jakab, 2013; Verde, 2012; Verde & Perfect, 2011; Williams & Zacks, 2001), Storm and Levy (2012) argued that retrieval-induced forgetting in

---

[5] Models that Raaijmakers and Jakab (2013) discussed assume that recognition is a based on a single process, familiarity, but other researchers have argued that recognition uses two processes, familiarity and recollection (see Yonelinas, 2002 for review).

recognition and cue independence may be unnecessary for the inhibition account if researchers adopt a more general definition of inhibition. Explicitly, inhibition account proponents have primarily argued that the representation of an exemplar or word itself is weakened. Alternatively, researchers could define inhibition as a process that reduces the memory strength of an exemplar, an association, or any other information to resolve competition between memories. However, this alteration would eliminate two methods (recognition and independent cue final tests) that are currently used to distinguish between the blocking and inhibition accounts. Nonetheless, Storm and Levy's idea is relatively new and will need further exploration (and this broader definition also makes the inhibition account more similar to the blocking account).

Strength independence

A second assumption of the inhibition account is strength independence, which states that increasing memory strength of the Rp+ exemplars does not increase the magnitude of retrieval-induced forgetting. This assumption has been examined in multiple ways. For example, Aslan and Bäuml (2011) failed to find a correlation between the magnitude of the testing effect and the magnitude of retrieval-induced forgetting. Aslan and Bäuml examined retrieval-induced forgetting with category-exemplar word pairs and a final old-new recognition test. As expected, the researchers found a testing effect. Explicitly, Aslan et al. found a greater d' for Rp+ exemplars than Nrp exemplars. Interestingly, Aslan and Bäuml reported no significant correlation between the magnitude of the testing effect and the magnitude of retrieval-induced forgetting (the difference between d' for Nrp and Rp- exemplars), $r = -.08$, $p > .20$. This null effect violates the blocking account, which would lead one to predict a positive relationship between the testing effect and retrieval-induced forgetting because the higher the memory strength of a Rp+ exemplar, the more it should interfere at final test. However, as Raaijmakers and Jakab (2013) noted, a null effect does not mean that there is no relationship.

Storm and Nestojko (2010) found retrieval-induced forgetting when Rp+ exemplar facilitation was impossible. The researchers asked participants to learn a set of category-exemplar word pairs (e.g., *Dog* – Lab). After studying the word pairs, participants were given a semantic generation task. That is, participants were given some studied categories with a two-letter stem and asked to generate an unstudied exemplar. For example, if participants were given "*Dog* – Co____," participants might supply the word "Collie." Notably, semantic generation is similar to retrieval practice with one difference. For retrieval practice, participants are asked to retrieve an exemplar from the study phase, which would directly strengthen the recalled exemplar. For semantic generation, participants are asked to generate any nonstudied exemplar that fits the category-plus-stem. For half of the category-plus-stems, it was possible to generate an appropriate exemplar (e.g., *Fruit* – Le____), whereas for the other half of the materials, it was impossible to supply an appropriate exemplar (e.g., *Fruit* – Ow____). On a final category-plus-one-letter stem cued recall task, participants demonstrated retrieval-induced forgetting in both the possible and impossible semantic generation tasks.

According to the inhibition account, studied exemplars should interfere when a participant attempts to produce an unstudied exemplar during semantic generation regardless of whether it is possible to produce a suitable exemplar. In order to bypass this interference, participants must inhibit the studied exemplars. However, the blocking account is *not* consistent with this finding, because one would predict Rp- exemplar impairment only when other items in the studied list were strengthened. If Rp+ exemplars cannot be strengthened (because it is not possible to generate an exemplar that fits a given cue), no retrieval-induced forgetting should occur. In sum, strength independence (see also Erdman & Chan, 2013; Román, Soriano, Gómez-Ariza, & Bajo, 2009; Shivde & Anderson, 2001) provides inhibition advocates with strong evidence for their theory.

Interference dependence

A third assumption of the inhibition account is interference dependence, which states that items must interfere with another exemplar's retrieval in order to be inhibited. In accordance with this notion, Anderson et al. (1994) found retrieval-induced forgetting for high taxonomic frequency exemplars (e.g., Orange) but not for low taxonomic frequency exemplars (e.g., Kiwi). The authors attributed this difference to an interference dependence property of retrieval-induced forgetting. Specifically, at retrieval practice, exemplars that interfere with one's ability to recall the correct exemplar should be inhibited. Noncompetitors therefore should not be affected. For example, if one was given the cue "*Fruit* – Or____," one may need to inhibit the exemplar "Banana" (a high taxonomic frequency exemplar – a likely competitor) but not "Kiwi" (a low taxonomic frequency exemplar – an unlikely competitor). On the other hand, according to the blocking account, retrieval-induced forgetting occurs whenever some studied exemplars are preferentially strengthened, so even low frequency unpracticed items should be blocked by the practiced items. Although other researchers have found retrieval-induced forgetting with high taxonomic frequency exemplars but not low frequency exemplars (Bäuml, 1998; Shivde & Anderson, 2001), this finding's reliability has been under serious scrutiny. For example, Williams and Zacks (2001) failed to find any difference in the magnitude of retrieval-induced forgetting for high and low taxonomic frequency exemplars. Likewise, Jakab & Raaijmakers (2009) manipulated the strength of the Rp- exemplars through the number of presentations and the position of the exemplar within a category, and they failed to find a difference in the magnitude of retrieval-induced forgetting for stronger exemplars compared to weaker exemplars.

Spitzer and Bäuml (2009) have also challenged this assumption, because they found retrieval-induced forgetting in a category recognition task. Participants learned words in various colors (e.g., the word "Blizzard" presented in red), and they were asked to remember both the words and a word's color category. In the retrieval practice phase, participants were given two-letter stem cued recall, which was in a word's correct color. That is, participants might be given "Bl____" in the color red and were asked to recall "Blizzard." After a distractor task, a color recognition test was given. Specifically, instead of recognizing the identity of the word, participants had to recognize its color. Spitzer and Bäuml found that recognition performance of the Rp- exemplars was worse than that of the Nrp exemplars. This finding provided support for the blocking account. Inhibition account advocates have stated that shared information between the Rp- and Rp+ exemplars (i.e., the color in this case) is not subject to inhibition because that information should not interfere with retrieval (Anderson, Green, & McCulloch, 2000).

Retrieval specificity

Another assumption of the inhibition account is retrieval specificity. To be precise, inhibition is required only when one needs to retrieve a competing item;

therefore, restudy of the category-exemplar pair or category retrieval practice (e.g., showing a participant "*W*____ - Blizzard" in order to cue "Blizzard") is not sufficient to create retrieval-induced forgetting. For example, Anderson, Bjork, and Bjork (2000) demonstrated that noncompetitive retrieval practice (e.g., Blizzard – *W*____) did not result in retrieval-induced forgetting. Conversely, the researchers found retrieval-induced forgetting with competitive retrieval practice, where participants were given a category and asked to recall an exemplar (see also Hulbert, Shivde, & Anderson, 2011).

Raaijmakers and Jakab (2012) have challenged the retrieval specificity assumption of the inhibition account, because they found retrieval-induced forgetting with noncompetitive retrieval practice by simply increasing the difficulty of the retrieval practice trials. Specifically, Raaijmakers and Jakab used categories based on physical properties and low taxonomic frequency exemplars. By making the noncompetitive retrieval practice more difficult, Raaijmakers and Jakab argued that they improved the strength of the association between the category-exemplar pair, which is crucial to creating blocking at final test. However, to combat problems with unsuccessful retrieval practice trials, the researchers administered feedback. Notably, feedback is typically not administered (e.g., displaying "*Weather* – Blizzard" after participants are given an opportunity to recall "Blizzard" in retrieval practice) in retrieval-induced forgetting experiments. However, if participants perform poorly during retrieval practice, feedback would allow participants to strengthen all of the Rp+ exemplars. When feedback is not given, receiving the memorial benefits of testing is contingent on correct recall. If participants are unable to recall a cued exemplar, the cued exemplar will not be better remembered. However, feedback can mitigate this problem by allowing participants to improve their memory (through a feedback trial) even with unsuccessful retrieval practice. This strengthening of the Rp+ exemplar pair is crucial in order to find retrieval-induced forgetting according to the blocking account. Therefore, Raaijmakers and Jakab allowed for a higher likelihood of retrieval-induced forgetting from blocking.

Individual differences

Aside from assumptions of the inhibition account, individual difference studies have contributed evidence to the inhibition versus blocking account debate. Firstly, Aslan and Bäuml (2012) found that participants with higher working memory capacity demonstrated a higher magnitude of retrieval-induced forgetting than participants with lower working memory capacity. Engle (2002) suggested that working memory capacity is a measure of cognitive control. Furthermore, some inhibition account proponents have suggested that retrieval-induced forgetting might be a result of executive control (Levy & Anderson, 2008). However, some researchers have had difficulty finding a relationship between working memory capacity and the magnitude of retrieval-induced forgetting (Bell, 2005; Erdman, 2011).

Although the above review presented inhibition and blocking as mutually exclusive possibilities (and indeed they are often treated as such), Aslan and Bäuml (2010) suggested that multiple processes may contribute to retrieval-induced forgetting. Some cognitive development theorists have argued that poor performance by children in some cognitive tasks is a result of poor inhibitory control (Bjorklund & Harnishfeger, 1990). Aslan and Bäuml examined performance by kindergartners, second graders, and adults on two retrieval-induced forgetting tasks that differed only on the final test – one employed a recognition final test and the other employed a recall final test. Aslan and

Bäuml found that all three groups demonstrated retrieval-induced forgetting in the recall task. However, in the recognition task, only second graders and adults demonstrated retrieval-induced forgetting. The authors suggested that the recall final test might allow blocking to contribute more to retrieval-induced forgetting, while the recognition task was less susceptible to blocking effects – leaving inhibition as the primary contributor to retrieval-induced forgetting. Since the kindergartners may lack inhibitory control, they did not demonstrate retrieval-induced forgetting in the recognition task. On the other hand, by second grade, children might have developed the necessary inhibitory control. Therefore, second graders (and adults) may have inhibited the Rp- exemplars for the recognition final test.

Furthermore, Aslan and Bäuml (2012) found that older adults over 75 years old did not demonstrate retrieval-induced forgetting with a recognition final test, but older adults between the ages of 70 and 75 demonstrated retrieval induced forgetting. Consistent with this idea, Lustig, Hasher, and Zacks (2007) posited that older adults may have deficient inhibition relative to younger adults. Aside from age of participants, Storm and Levy (2012) offered that the situation (e.g., the type of retrieval practice or final test) and personal characteristics (e.g., age or working memory capacity) may dictate whether blocking or inhibition is primarily driving a specific retrieval-induced forgetting effect. That is, reflecting on much of the evidence that I have just discussed, Storm and Levy argued that there is evidence for both the interference and blocking accounts.

Finally, aside from the blocking and inhibition accounts, there is another minor account: strategy disruption. In three experiments, Dodd, Castel, and Roberts (2006) manipulated which items participants were cued to recall during retrieval practice and demonstrated that retrieval-induced forgetting disappears when Rp+ items receive retrieval practice in their studied order. For example, retrieval-induced forgetting was not found in groups that received retrieval practice over the first five studied items or the last five study items in their encoding order. Furthermore, Dodd et al. did not find retrieval-induced forgetting when participants were cued to retrieve every other studied item (e.g., studied items 1, 3, 5, 7…) during retrieval practice. Critically, retrieval-induced forgetting occurred only when the retrieval practice order was randomized (which is the procedure used in all studies demonstrating retrieval-induced forgetting). That is, when retrieval practice does not interfere with the natural order that one may attempt to retrieve the exemplars (their serially studied order), retrieval-induced forgetting does not occur. Therefore, aside from the two major accounts of retrieval-induced forgetting (blocking and inhibition), strategy disruption may be a notable contributing factor. I now provide a review for the test-potentiated learning literature.

Empirical review of test-potentiated learning

Test-potentiated learning refers to increased learning of new information after taking a test. In this dissertation, I use the term “test-potentiated learning” to indicate the effect of testing on the learning of new information, “testing effect” to indicate the influence of testing on previously studied information, and “interim testing” to indicate the methodological procedure of providing tests after learning a subset of word pairs (before one learns another subset of material). Several researchers have found that testing could facilitate learning of later information. For example, Darley and Murdock (1971) asked participants to study 10 lists of unrelated nouns. For half of the lists, participants took an initial free recall test over the lists. For the other half of the lists, participants performed a distractor task instead. Therefore, each participant studied five lists followed by an immediate test and five lists followed by a distractor task. Participants did not perform more than three distractor or three retrieval practice lists in a row. Importantly, Darley and Murdock noted that 81% of the intrusions on the initial tests were nouns from the nontested lists. Consequently, taking an immediate test reduced intrusions from that list on future tests of other lists compared to performing a distractor task.

Tulving and Watkins (1974) examined interference with an A-B, A-C paradigm. For example, participants might have studied “Cave – Face” (A-B). Later, they might have studied “Cave – Goose” (A-C). Tulving and Watkins manipulated whether a recall test was administered only after the A-B list, only after the A-C list, after both lists, or neither list. Participants were then given an opportunity to recall all of the targets (e.g., B & C) when given the cue (e.g., A; see Barnes & Underwood, 1959). Remarkably, when a test was not administered after the A-B list, participants recalled fewer “C” words on both the immediate test and the final test. Tulving and Watkins attributed this detriment to impoverished learning of the A-C lists due to proactive interference when no interim test was taken.

A study by Szpunar et al. (2008) has revived interest in examining the effects of interim testing on memory performance. Szpunar et al. suggested that interim testing may influence later learning by reducing proactive interference. Studies on proactive interference have amassed an extensive literature (see Postman, 1971; Roediger, Weinstein & Agarwal, 2010). Proactive interference occurs when material that was presented before the to-be-remembered information impedes one's ability to learn new information. In their study, Szpunar et al. (2008) asked participants to learn five lists of words. Half of the participants received a free recall test immediately after studying each list (the *interim test* group), whereas the other participants completed a math distractor task instead of the recall test (the *control* group). Participants repeated this study-test cycle for four lists. After the fifth list was presented, all participants completed a free recall test over this fifth list. Finally, all participants performed a cumulative free recall test over all the lists.

Szpunar et al. (2008) found that participants in the interim test group recalled more List 5 items and produced fewer intrusions from prior lists than the control group. Furthermore, the researchers found that the interim test group recalled more studied words on the cumulative final test than the control group. In another experiment, Szpunar et al. demonstrated this finding using category-exemplar word pairs and with an interim study condition rather than an interim math condition. The researchers attributed this benefit of interim testing to a reduction of proactive interference. Participant's List 5 recall rate in the interim testing condition was similar to their List 1 recall rate, and participant's List 5 recall rate in the interim math condition was much lower than their List 1 recall rate. Therefore, the researchers argued that testing inoculated participants against the negative effects of proactive interference. Furthermore, this effect has been extended to the learning of face-name pairs (Weinstein, McDermott, & Szpunar, 2011), prose (Wissman, Rawson, & Pyc, 2011), and online lectures (Szpunar, Khan, & Schacter, 2013).

Pastötter, Shichker, Nidernhuber, and Bäuml (2011) examined the neural correlates of test-potentiated learning using electroencephalogram (EEG). Pastötter et al. followed Szpunar et al.'s design with two major additions. Firstly, Pastötter et al. recorded EEG data while participants learned Lists 1-5. Secondly, in addition to interim testing and interim math, the researchers added a restudy condition, a semantic generation condition, and a 2-back task condition. In the restudy condition, instead of taking an interim recall test, participants restudied the words in the current list, one at a time, in a new random order. In the semantic generation condition, participants generated words from a given semantic category instead of taking an interim test. In the 2-back condition, participants were asked to note whether an item was presented two trials earlier in a sequence instead of taking an interim recall test.

Interestingly, participants in the testing, semantic generation, and 2-back conditions remembered more List 5 items than participants in the distractor or restudy conditions. Furthermore, participants in the testing and generation conditions produced much lower intrusion rates than those in the 2-back, restudy, or distractor conditions. Pastötter et al. (2008) also found that alpha power, which is thought to indicate attention failure (Palva & Palva, 2007), increased across multiple lists without some sort of interpolated retrieval (i.e., the testing, generation, or 2-back conditions). Specifically, they suggested that encoding becomes less effective across lists without interpolated retrieval, and they argued that the retrieval triggered an internal context change, thus "resetting" encoding operations for each study list. Pastötter et al. noted that their results were consistent with Szpunar et al. (2008). Previous research has focused on how interim testing over the entire previously learned list can benefit new learning. Because this literature has not examined how testing a subset of previously learned information will affect new learning, I chose to examine such in the current dissertation.

Overview of the present experiments

In this dissertation, I investigated whether completing interim tests on a subset of the previously learned list can alleviate proactive interference, and I investigated how the presentation order of Rp- exemplars and Rp+ exemplars influenced retrieval-induced forgetting. Participants studied four lists of category-exemplar word pairs and performed retrieval practice over some of the exemplars after each list, studied category-exemplar word pairs and performed interim math after each list, or learned all category-exemplar pairs before performing retrieval practice over some of the word pairs. After a distractor task, participants performed a category cued recall test (Experiment 1) or a category-plus-one-letter stem cued recall test (Experiments 2 and 3).

The current dissertation contributes evidence towards two separate literatures: the retrieval-induced forgetting literature and the test-potentiated learning literature. Regarding the retrieval-induced forgetting literature, the current dissertation compared predictions that would result from the inhibition and blocking accounts. I examined what predominantly contributes to retrieval-induced forgetting (blocking or inhibition). Furthermore, Szpunar et al. (2008) suggested that interim testing before studying new information reduced proactive interference. However, Szpuner et al. provided an interim test over all of the previously studied information. I examined whether only testing a subset of the information reduced proactive interference. More specifically, I examined how interim testing affected retrieval-induced forgetting.

Aside from the theoretical implications, the current study offered information relevant to educational practice. When students test themselves over information prior to an exam, they are unlikely to test themselves over all the material. Similarly, instructors are unable to test all information in exams. In these types of situations, students and educators can choose whether to give multiple smaller tests or a large test. The current dissertation will compare how memory is affected by both formats of testing.

In this dissertation, I examined the effects of breaking up learning and retrieval practice into multiple blocks versus taking a single cumulative retrieval practice test on later memory performance of both the Rp+ and Rp- items. That is, participants either completed retrieval practice under high or low retrieval competition. Moreover, participants performed retrieval practice or mental arithmetic after studying each block of category-exemplar pairs. The level of retrieval competition was manipulated by either presenting Rp- exemplars before retrieval practice (a high competition condition) or after retrieval practice (a low competition condition).

On a theoretical level, if blocking plays a larger role in retrieval-induced forgetting than inhibition, presentation order should not affect retrieval-induced forgetting because blocking occurs at the final test, after all the exemplars have been presented, and retrieval-induced forgetting should be observed in the cumulative retrieval practice, high competition, and low competition conditions. Because the inhibition account assumes that impairment occurs at retrieval practice, the order of presentation is important. In the high competition condition, the Rp- exemplars may compete for retrieval and thus be suppressed during retrieval practice. However, in the low competition condition, the Rp- exemplars have yet to be presented when participants perform retrieval practice on the Rp+ exemplars. Therefore, the Rp- exemplars are unlikely to compete during retrieval practice, and participants in the low competition condition should not demonstrate retrieval-induced forgetting.

Furthermore, there were two control conditions – the cumulative retrieval practice condition and interim math condition. In the cumulative retrieval practice condition, participants studied a random presentation of all word pairs and received one set of retrieval practice. In the interim math condition, participants studied the word pairs in separate blocks of trials, but they answered math problems instead of performing retrieval practice at the end of each study block. After a 20-minute distractor period, participants took a final test on their memory of the exemplars. All three experiments used the same design except for the final test format. Experiment 1 employed a category cued recall final test. Experiments 2 and 3 employed a category-plus-stem cued recall final test.

The cumulative retrieval practice condition was included to establish a significant retrieval-induced forgetting effect with the current materials and subject population based on the typical procedure. The interim math condition was included as a baseline condition to examine whether performing interim tests on the studied materials (as in the high competition and low competition conditions) would reduce proactive interference for new learning.

Finally, the type of final test is important as the final test can limit contributions from each of the potential sources of retrieval-induced forgetting (i.e., blocking or inhibition). A category cued recall test allows for both inhibition and blocking to contribute to a retrieval-induced forgetting effect. However, a category-plus-stem cued recall test limits some of the contributions from blocking. Although a category-plus-stem final test cannot eliminate all blocking influences, it does afford one the ability to manipulate output order, thereby reducing the influence of output interference on final test performance. For example, one can test the Rp- exemplars prior to their Rp+ counterparts, thus reducing the likelihood that earlier output of the Rp+ exemplars would block retrieval of the Rp- exemplars. Nonetheless, evidence for or against either of the primary accounts of retrieval-induced forgetting will depend on whether participants in the high competition, low competition, neither, or both conditions demonstrate retrieval-induced forgetting.

CHAPTER 2. EXPERIMENT 1

There were three goals to Experiment 1. Firstly, I investigated how retrieval-induced forgetting was influenced by interim retrieval practice rather than cumulative retrieval practice. Again, the interim retrieval practice conditions broke up the learning phase into four blocks, each included a study and retrieval practice cycle, whereas the cumulative retrieval practice condition included a single block of study trials followed by a single block of retrieval practice trials (i.e., the cumulative retrieval practice condition). Secondly, I examined whether learning order affected retrieval-induced forgetting. Specifically, according to the inhibition account, retrieval-induced forgetting should not occur when Rp- exemplars are presented after the Rp+ exemplars have completed retrieval practice (the low competition condition). However, according to the blocking account, one would predict retrieval-induced forgetting to occur in the low competition condition.

Finally, I examined if performing interim testing on only a subset of the studied exemplars would increase the number of exemplars learned in future blocks (specifically Nrp exemplars in this case to avoid any potential effects of testing or retrieval-induced forgetting) compared to the interim math condition. That is, I compared the Nrp exemplar recall probabilities in the interim math condition with those in the high competition and low competition conditions. If interim testing reduces proactive interference, participants in the high competition and low competition condition should experience less proactive interference than participants in the interim math condition. Therefore, participants in the interim math condition should have a lower final test Nrp exemplar recall probabilities in later blocks relative to earlier blocks, but participants in the high competition and low competition conditions should demonstrate an equal final test Nrp exemplar recall probabilities for later blocks and earlier blocks. That is, in the interim math condition, I expected participant's recall rate to drop from Block 1 to Block 4 (e.g., participants might recall 75% of the Nrp exemplars in Block 1 but 50% of the Nrp exemplars from Block 4), which would suggest that the buildup of proactive interference across the encoding blocks had a negative influence on learning during the latter encoding blocks. Conversely, interim testing should eliminate the buildup of proactive interference over the encoding blocks. However, I predicted that participants in the high competition and low competition conditions would maintain the same recall rate across blocks. The specific question addressed in Experiment 1 was how will interim testing affect memory for the Rp-, Rp+, and Nrp exemplars in a category cued recall final test?

Method

Design

Experiment 1 had a 4 (learning condition: high competition, low competition, cumulative retrieval practice, interim math) X 3 (retrieval practice status: Rp+, Rp-, Nrp) mixed design. Learning condition was manipulated between-subjects. The learning condition variable refers to the order in which materials were studied and whether interim testing was performed during the learning phase. In the high competition condition, during the first two blocks of trials, participants studied category-exemplar pairs and then took an interim test on the filler items instead of the target items after each block of study trials. During the last two blocks of trials, participants took an interim test on the target (Rp+) items. In the low competition condition, the target items were tested in the first two blocks, and the filler items were tested in the last two blocks. In the interim math condition, participants answered math problems instead of performing retrieval practice after each block of study trials. In the cumulative retrieval practice condition, participants studied the Rp- exemplars, the Rp+ exemplars, the Nrp exemplars, and the filler exemplars presented in a random order. After subjects studied all of the category-exemplar pairs, they completed the retrieval practice phase. Retrieval practice status was manipulated within-subjects. Filler items were included so that retrieval practice occurred after all blocks of study trials in the high competition and the low competition conditions.

Participants

One hundred twenty-eight Iowa State University students participated in Experiment 1 in return for research credit. This resulted in thirty-two participants per learning condition.

Materials

The target items consisted of 48 category-exemplar pairs taken from the strong exemplars from Experiment 3 of Anderson et al. (1994). Anderson et al. initially collected these category-exemplar pairs from Battig and Montague (1969). The word pairs consisted of eight categories with six exemplars each. The category-exemplar pairs had an average taxonomic frequency of .34 (*SD* = .24). The set of 24 filler category-exemplar pairs consisted of eight categories with three exemplars apiece. Filler exemplars were selected from the categories in the Battig and Montague norms that did not have share exemplars with target categories. The filler exemplars had an average taxonomic frequency of .31 (*SD* = .19). The target category-exemplar pairs and filler pairs can be found in the Appendix A and B, respectively. The computerized operation span task (OSPAN) served as a distractor task (Unsworth, Heitz, Schrock, & Engle, 2005).

Presentation order was determined as follows. I first randomized the order of the categorized lists. I then randomized the order of the items within each list. The first four categories were presented in Block 1 and Block 3. The second four categories were presented in Block 2 and Block 4. Furthermore, each exemplar was presented equally often as part of a control and tested category. When the exemplars were presented within a tested category, half of the time the exemplar was presented as a Rp- exemplar and the other half of the time it was as a Rp+ exemplar.

In the interim math condition, encoding trials were presented in the same way as in the high competition and low competition conditions. However, because no interim memory test was administered, there was no difference in item types.

Procedure

Experiment 1 was divided into three phases: learning, distractor, and final test. In the learning phase, participants were asked to study category-exemplar pairs in one of four ways (manipulated between-subjects). In three of the four conditions (interim math, high competition, and low competition), participants were asked to study the category-exemplar pairs in four blocks. For these three conditions, participants studied a set of category-exemplar pairs and then completed a set of retrieval practice trials (high competition and low competition conditions) or math problems (interim math condition) in each block. In the cumulative retrieval practice condition, participants studied all the category-exemplar pairs in one block and then performed one set of retrieval practice trials. Figure 1 provides an outline of the type of items (e.g., Rp-, filler, Rp+, or Nrp) that

Figure 1. Outline of the learning conditions.

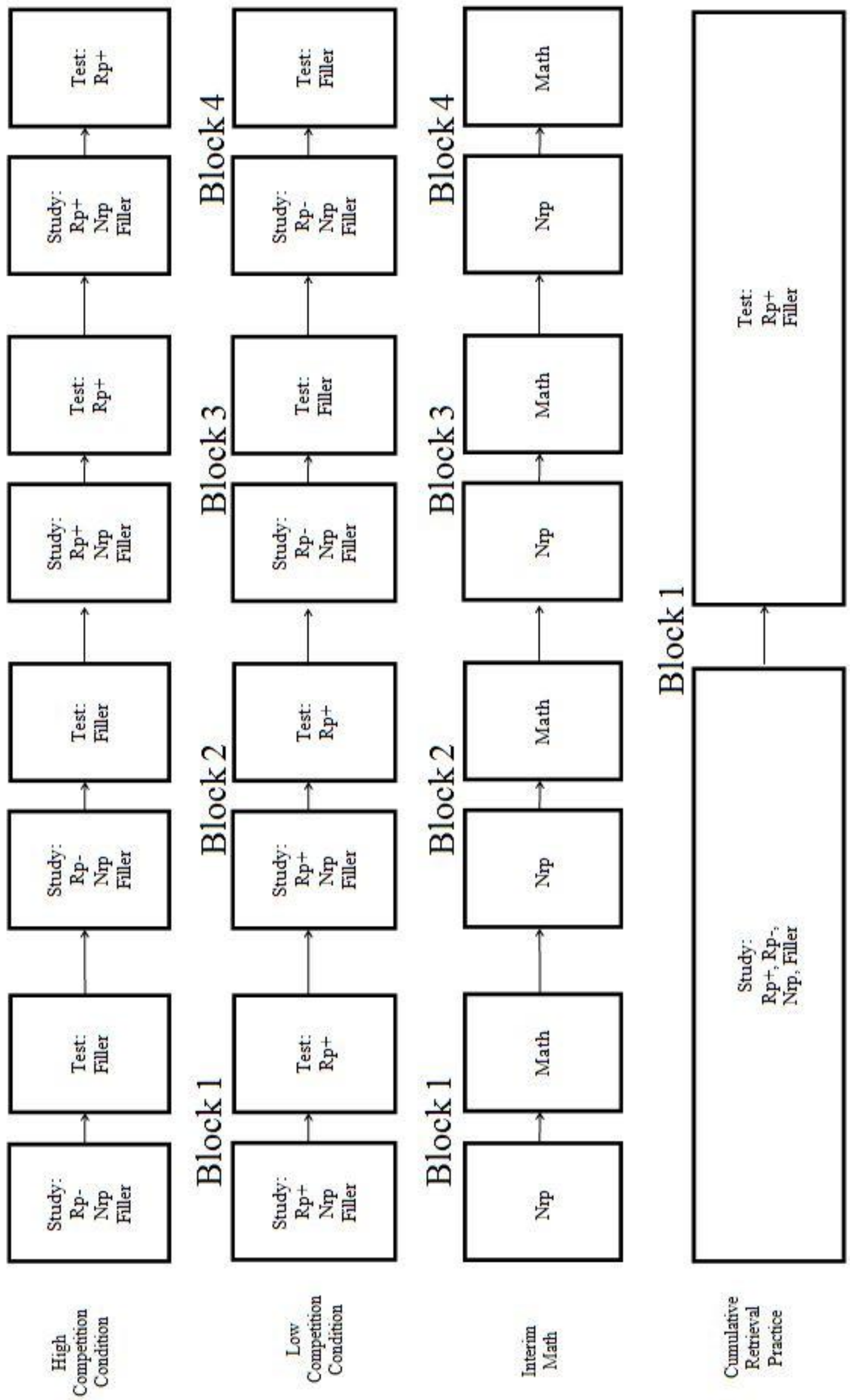

Figure 2. Sample exemplars in each learning condition

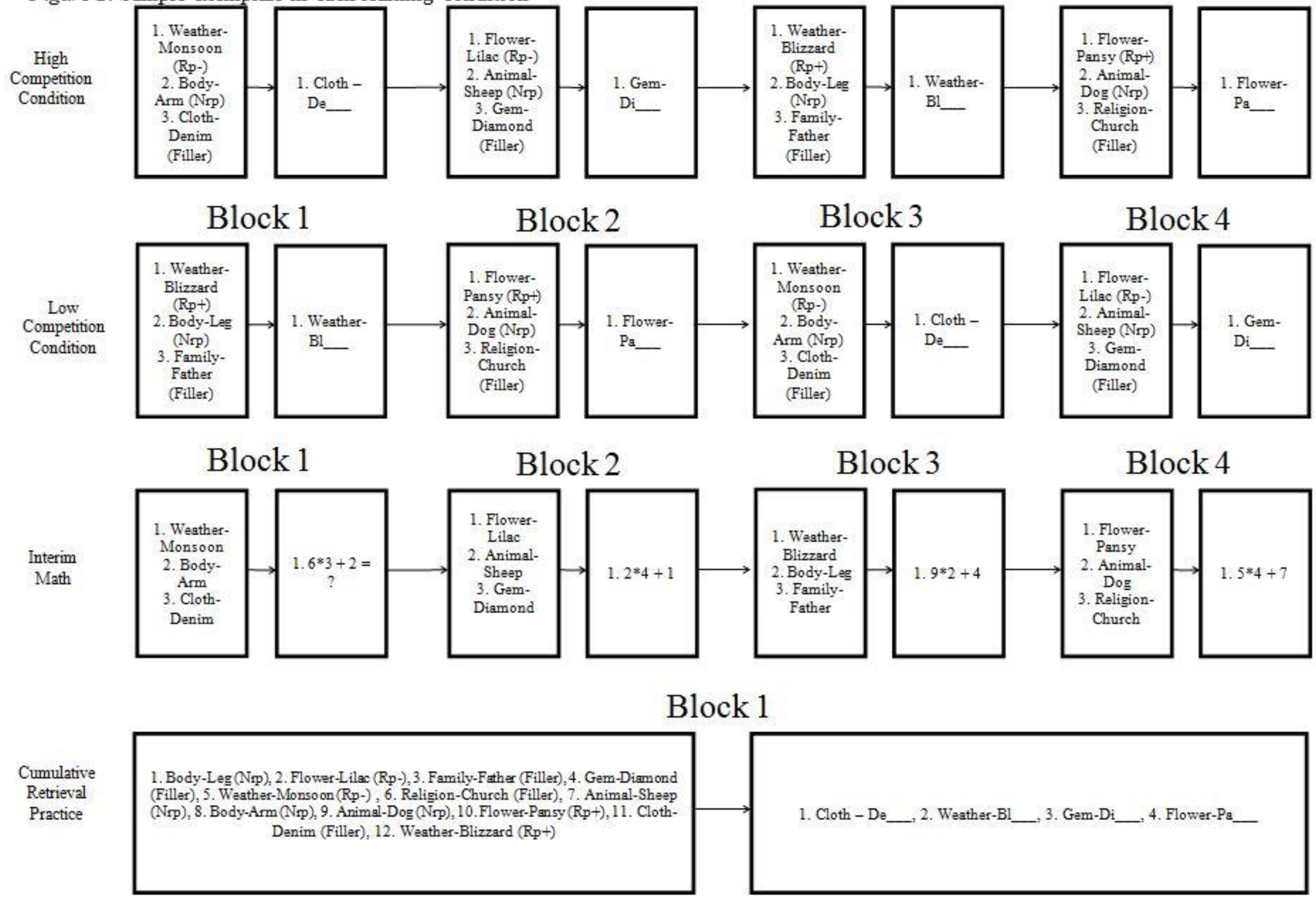

Table 1

Sample Exemplars for each Retrieval Practice Status and Filler Exemplars

Rp+	Rp-	Nrp	Filler
Weather-Blizzard	Weather-Monsoon	Body-Arm	Cloth-Denim
Flower-Pansy	Flower-Lilac	Body-Leg	Gem-Diamond
		Animal- Sheep	Family-Father
		Animal-Dog	Religion-Church

were presented in each block of each condition. Figure 2 provides specific examples of category-exemplar pairs that might have been presented in each study or test trial, and Table 1 defines each exemplar's retrieval practice status or filler status.

The *high competition* group had four study/test blocks. During each study trial, a category-exemplar pair was presented for 5 s, with a 500 ms interstimulus interval (ISI). In Blocks 1 and 2, participants studied six Rp- exemplars, six Nrp exemplars, and six filler exemplars. In Blocks 3 and 4, participants studied six Rp+ exemplars, six Nrp exemplars, and six filler exemplars. Each block provided a new set of exemplars for participants to study.

After participants studied a block of category-exemplar pairs, they completed category-plus-one-letter stem retrieval practice over either the filler exemplars (Blocks 1 and 2) or the Rp+ exemplars (Blocks 3 and 4). A 500 ms ISI separated each retrieval practice trial. Like previous studies on retrieval-induced forgetting, each exemplar received retrieval practice three times (Anderson et al., 1994; Camp et al., 2007; Jakab & Raaijmakers, 2009; Smith & Hunt, 2000). Explicitly, each test block had six unique test trials. Participants completed the six unique test trials one time in a random order except that items in the same categories did not appear on adjacent trials. After participants completed the test trials once, they completed the test trials two more times (both times in a fresh random order). Therefore, each test block had 18 test trials.

In Block 1 of the high competition condition, participants might study "*Weather* – Monsoon, *Body* – Arm, and *Cloth* – Denim." After studying all Block 1 exemplars, filler exemplars that were studied in Block 1 were tested (e.g., *Cloth* – Denim). These filler exemplars were included so that an interim test occurred for all blocks of trials while preserving the retrieval practice status of the target (Rp-) exemplars in Blocks 1 and 2. In Block 2, no exemplars shared categories with any studied exemplars from Block 1. New Rp- exemplars (e.g., *Flower* – Lilac), Nrp exemplars (e.g., *Animal* – Sheep), and filler exemplars (e.g., *Gem* – Diamond) were studied. Afterwards, the filler exemplars were tested. Blocks 3 and 4 were identical to Blocks 1 and 2 with two exceptions. Firstly, Rp+ exemplars were studied instead of Rp- exemplars. Secondly, the Rp+ exemplars received retrieval practice instead of the filler exemplars.

The *low competition* condition was identical to the high competition condition, except that the Rp- exemplars were presented in Blocks 3 and 4, whereas the Rp+ exemplars were presented in Blocks 1 and 2.

In the *interim math* condition, similar to the high competition and low competition conditions, participants studied the category-exemplar pairs in four blocks. However, instead of completing an interim memory test at the end of each study block, participants completed mental arithmetic problems (e.g., 502 – 37 = ?). Similar to the high competition and low competition conditions, each block had six unique math problems, which were presented three times. Each interim math trial was presented for 5 s.

The *cumulative retrieval practice* group only had one block of study trials and one block of retrieval practice trials. All Rp+, Rp-, Nrp, and filler exemplars were presented, one at a time, in a random order during the study block. After studying all category-exemplar pairs, participants performed retrieval practice three times over all of the Rp+ category-exemplar pairs and 12 filler pairs. Thus, the retrieval practice phase contained 72 trials, with four blocks of 24 unique trials each. Each block of retrieval practice trials contained 12 Rp+ and 12 filler exemplars presented in a different random order.

After participants performed one of the four learning conditions, they completed the OSPAN task, and they played the videogame Tetris until 20 min had elapsed between the learning phase and the final test. Finally, participants were given a category cued recall final test. Participants were presented with a studied category name (e.g., Weather) and had 30 s to type in all exemplars (e.g., Monsoon, Blizzard) that were studied under that category. The order of the categories was randomized. A 500 ms ISI separated each test trial.

Results and discussion

Statistical outcomes are reported with an alpha level of .05 unless otherwise noted. Partial eta squared (η_p^2) or eta squared (η^2) indicates effect size for analysis of variance (ANOVA), and Cohen's *d* indicates effect size for t-tests. I present data from retrieval practice followed by those from the final test.

Retrieval practice

Initial test results are displayed in Table 2. I examined the initial test accuracy for the cumulative retrieval practice condition separately from the high competition and low competition conditions. The cumulative retrieval practice condition had only one block while the high competition and low competition conditions had four blocks, and presentation order of the Rp+ and Rp- exemplars was controlled in the high competition and low competition conditions but not in the cumulative retrieval practice condition. Therefore, I thought comparisons between the cumulative retrieval practice condition and the high competition/low competition conditions was inappropriate. Furthermore, the

Table 2

Mean Proportion Correct During Retrieval Practice for Experiment 1

	Block 1			Block 2			Block 3			Block 4		
Retrieval Practice Trial	1	2	3	1	2	3	1	2	3	1	2	3
High Competition	.70	.76	.81	.91	.94	.96	.86	.90	.92	.85	.91	.92
	(.17)	(.20)	(.17)	(.12)	(.10)	(.08)	(.15)	(.13)	(.14)	(.18)	(.13)	(.13)
Low Competition	.85	.88	.91	.91	.93	.94	.79	.85	.83	.93	.96	.95
	(.15)	(.14)	(.10)	(.11)	(.10)	(.10)	(.15)	(.15)	(.14)	(.09)	(.07)	(.08)
Cumulative Retrieval Practice	.79	.81	.83									
	(.12)	(.12)	(.11)									

Note: Standard deviations are displayed in parentheses.

interim math condition did not have retrieval practice, so no retrieval practice data is shown. Because I expect improvement from the first to the third retrieval practice trial, I will focus on the first and third retrieval practice trials' accuracy.

In the cumulative retrieval practice condition, participants recalled more exemplars on their third retrieval practice trial than their first, $t(31) = 2.58$, $p = .02$, $d = .35$, a hypermnesia effect (Erdelyi & Becker, 1974). Similar to the cumulative retrieval practice condition, I focused on the recall on the third retrieval practice trials compared to the first in the high competition and low competition conditions. Block and learning condition were confounded with the specific type of exemplar recalled (Rp+ or filler). Furthermore, Rp+ exemplars may be easier than filler exemplars to learn. Though there is little difference between the taxonomic frequencies of the category-exemplar pairs (.34 for the target category-exemplar pairs and .31 for the filler pairs), there may have been a difference with participant's familiarity with each set of categories. Though there is not a detectable difference between target exemplar's ($M = 169.23$, $SD = 231.9$) and filler exemplar's ($M = 290.26$, $SD = 318.64$) word frequency (Thorndike & Lorge, 1944), $t(26) = 1.49$, $p = .15$, word frequency norms do not take into consideration different meanings of the same word. Because all exemplars were placed in the context of a specific category, participants are likely to think about a specific representation of the word, which is not addressed by such norms. Furthermore, filler exemplars were taken from categories in which Battig and Montague's (1969) participants could not generate sufficient exemplars to have six unique exemplars (also, there could not be shared exemplars between target and filler categories) in which none of the exemplars shared the first two letters. These smaller categories could have impacted retrieval practice performance. Nonetheless, participants in the high competition and low competition conditions recalled more exemplars on the third retrieval practice trial (*M* = .90, *SD* = .09) than the first retrieval practice trial (*M* = .84, *SD* = .10), $t(63) = 9.07$, $p < .01$, $d = .63$.

Final test. Final test results for Experiment 1, which are broken down by retrieval practice status and learning condition, are displayed in Figure 3. Because it is important in the current dissertation to verify whether a retrieval-induced forgetting effect occurred in each level of the learning condition variable, I report comparisons at each level of this variable to examine whether there is a significant retrieval-induced forgetting effect.

A 3 (learning condition: high competition, low competition, cumulative retrieval practice) X 3 (retrieval practice status: Rp+, Rp-, Nrp) mixed ANOVA revealed a main effect of retrieval practice status, $F(2, 186) = 295.66$, $p < .01$, $\eta_p^2 = .76$. The data demonstrated a testing effect as Rp+ exemplars (*M* = .73, *SD* = .18) were recalled far more often than Nrp exemplars (*M* = .40, *SD* = .13), $t(95) = 18.98$, $p < .01$, $d = 2.10$. Moreover, the data demonstrated retrieval-induced forgetting as Nrp exemplars were recalled more often than Rp- exemplars (*M* = .29, *SD* = .18), $t(95) = 6.98$, $p < .01$, $d = .70$. Moreover, significant retrieval-induced forgetting was found in the cumulative retrieval practice condition, $t(31) = 4.94$, $p < .01$, $d = .80$, the high competition condition, $t(31) = 3.02$, $p < .01$, $d = .55$, and the low competition condition, $t(31) = 4.26$, $p < .01$, $d = .66$. There was not a main effect of learning condition, $F < 1$, $p = .88$, or a learning condition by retrieval practice status interaction, $F < 1$, $p = .49$.

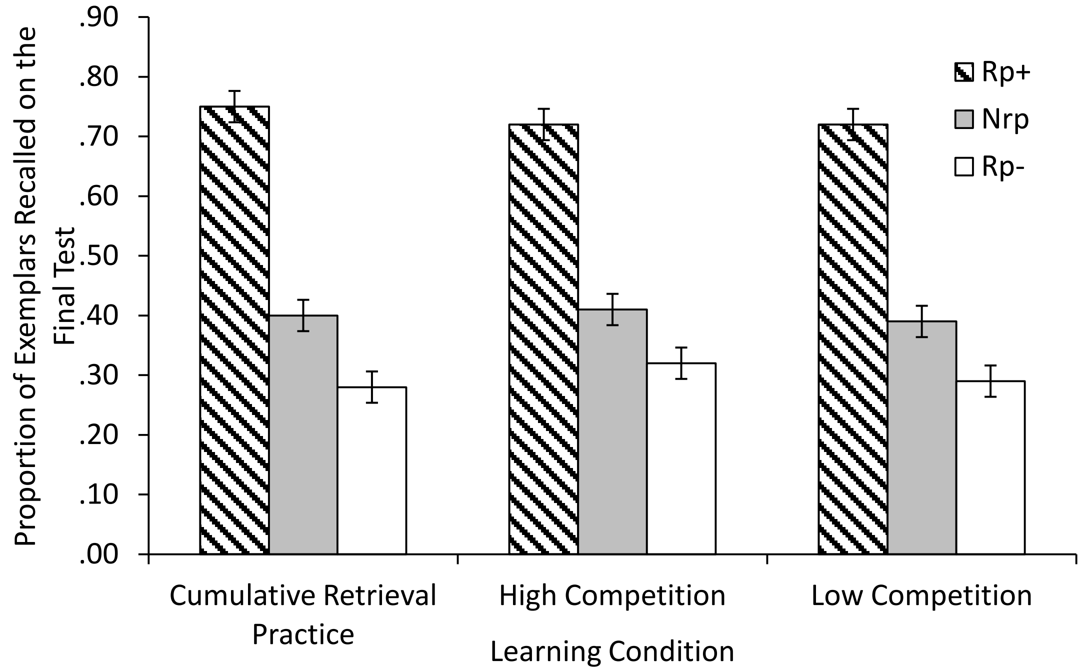

Figure 3. Proportion of exemplars recalled on the final test as a function of retrieval practice status and learning condition in Experiment 1. Error bars reflect the within-subjects 95% confidence interval.

Experiment 1's results are consistent with the blocking account. Explicitly, retrieval-induced forgetting occurred in both the high competition and low competition conditions. That is, the blocking account assumes that retrieval-induced forgetting is the result of blocking at final test. Consequently, with a category cued recall test, participants may have recalled the better learned Rp+ exemplars before they attempted to recall Rp- exemplars, which may have blocked Rp- exemplar recall (see Anderson et al., 1994).

This analysis also provided a test for the competition dependence assumption of the inhibition account of retrieval-induced forgetting. According to the competition dependence assumption, one would expect Rp- exemplar impairment (compared to Nrp exemplars) in the high competition condition but not low competition condition. However, retrieval-induced forgetting was found in both the high competition and low competition conditions. I further discuss the implications and possible limitations of these findings following presentation of the results regarding test-potentiated learning.

Final test results, broken down by block and learning condition, for Experiment 1 are displayed in Figure 4. To investigate whether retrieval practice released participants from proactive interference between lists, I examined Nrp exemplar recall probability across all four blocks for the high competition, low competition, and interim math conditions. Specifically, if testing reduced proactive interference, I expected participants in the high competition and low competition conditions to demonstrate a higher recall rate than the interim math condition for words presented in the later blocks (e.g., Blocks 3 and 4). A 3 (learning condition: high competition, low competition, interim math) X 4 (block: 1, 2, 3, 4) ANOVA revealed a main effect of block, $F(3, 279) = 6.00$, $p < .01$, $\eta_p^2 = .06$. Participants did not reliably recall more Nrp exemplars from Block 1 ($M = .39$, $SD = .21$) than from Block 2 ($M = .44$, $SD = .23$), $t(95) = 1.87$, $p = .07$, but they recalled fewer exemplars from Block 3 ($M = .32$, $SD = .21$) than from Block 2, $t(95) = 4.71$, $p < .01$, $d = .54$. Participants recalled more exemplars from Block 4 ($M = .37$, $SD = .24$) than Block 3, $t(95) = 1.99$, $p = .05$, $d = .22$. Though the reason for this pattern is unclear, aside from an inconclusive drop in recall probability of Nrp exemplars in the Block 3, participants performance across blocks did not seem to decline.

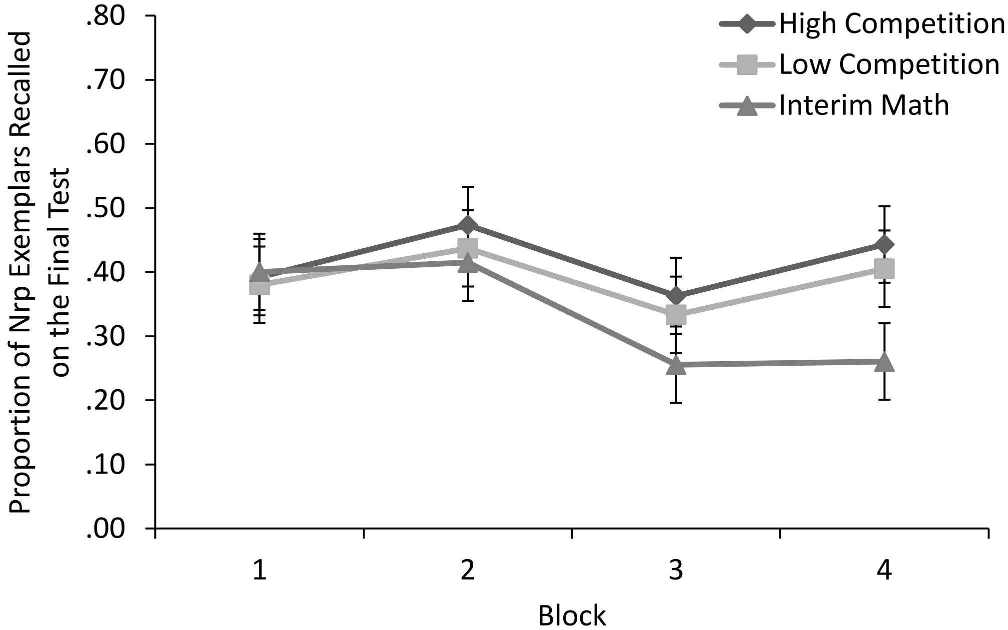

Figure 4. Proportion of Nrp exemplars recalled on the final test as a function of learning condition and block in Experiment 1. Error bars reflect the within-subjects 95% confidence interval.

There was also a main effect of learning condition, $F(2, 93) = 3.65$, $p = .03$, $\eta_p^2 = .07$. Notably, participants in the high competition condition ($M = .42$, $SD = .11$) recalled more Nrp exemplars than participants in the interim math condition ($M = .33$, $SD = .29$), $t(62) = 2.85$, $p < .01$, $d = .41$. There was no reliable difference between the interim math condition or the high competition condition and the low competition condition, $t(62) = 1.67$, $p = .10$, and $t < 1$, $p = .38$, respectively. Because participants in the high competition condition and low competition condition recalled more Nrp exemplars than those in the interim math condition, there is some evidence that retrieval practice over a subset of the Rp+ exemplars reduces proactive interference. However, one would predict a significant interaction between block and learning condition, but there was no evidence for such an interaction, $F(6, 279) = 1.46$, $p = .19$. I predicted that the interim math condition should have a buildup of proactive interference but not in the low competition and high competition conditions. Nonetheless, there was some evidence in that the final test data trended in this direction for all three conditions, but there may not have been enough power to detect the predicted interaction. That is, if one examines Figure 4, one can see that participant's recall rate in Block 1 is similar to that of Block 4 in the high competition and low competition conditions, but participant's recall rate in Block 4 appears lower than that of Block 1 for participants in the interim math condition.

In Experiment 1, I found a reliable retrieval-induced forgetting effect across the cumulative retrieval practice, high competition, and low competition conditions. These results could be predicted from the blocking account of retrieval-induced forgetting. Because I found this evidence of blocking with a category cued recall final test (a test that does not control for output interference - a source of associative blocking), I wanted to examine in Experiment 2 whether I could find evidence of inhibition with a test that can control for output interference, a category-plus-stem cued recall test. I also found some evidence of a reduction of proactive interference during initial learning, but the results were not quite as I predicted. That is, I did not find the learning condition by block interaction that I had predicted, so in Experiment 2, I also examined whether the pattern of Nrp exemplar recall would be replicated.

CHAPTER 3. EXPERIMENT 2

Blocking and inhibition may both contribute to retrieval-induced forgetting. Because output order is not controlled in a category cued recall test, blocking may play a large part in the retrieval-induced forgetting effect found in Experiment 1, which can mask the effects of inhibition on performance (Anderson, 2003). Experiment 2 controlled output order with a category-plus-stem cued recall final test. In a category cued recall test, participants are free to recall the exemplars in any order (e.g., they may recall the Rp+ exemplars before they recalled the Rp- exemplars). However, in the category-plus-stem cued recall test in Experiment 2, participants were cued to recall the Rp- exemplars prior to the Rp+ exemplars. Therefore, the Rp- exemplars may be less affected by output interference. This procedure thus reduces one source of associative interference (see Anderson et al., 1994).

Along these lines, the first three Nrp exemplars that were tested in a category were the baseline for the Rp- exemplars, and the second three Nrp exemplars that were tested in a category were the baseline for the tested exemplars.[6] For example, assume that a participant studied Blizzard, Drought, and Lightning (Rp+ exemplars); Sunny, Humidity, and Monsoon (Rp- exemplars); and Leg, Arm, Ear, Hand, Toe, and Mouth (Nrp exemplars). When exemplars on in the “Weather” category were tested on the final test, “Sunny, Humidity, and Monsoon” (Rp- exemplars) would be tested prior to “Blizzard, Drought, and Lightning” (Rp+ exemplars). Furthermore, assume that the Nrp

---

[6] Control exemplars for the Rp- exemplars are called “Nrp-,” and the control exemplars for the tested exemplars are called “Nrp+.”

exemplars were tested in the following order: Leg, Arm, Ear, Hand, Toe, and Mouth. Leg, Arm, and Ear were the first three Nrp exemplars from the "Body" category that were tested. Therefore, Leg, Arm, and Ear would be the Nrp exemplars that were compared to the Rp- exemplars. On the other hand, Hand, Toe, and Mouth would be the Nrp exemplars that were compared to the Rp+ exemplars. Furthermore, the delay between the initial and final tests was reduced to 10 min (from 20 min in Experiment 1). Aside from a category-plus-one letter stem cued recall final test and shorter distractor period, Experiment 2 was identical to Experiment 1.

Method

Design

Experiment 2 had a 4 (learning condition: high competition, low competition, cumulative retrieval practice, and interim math) X 3 (retrieval practice status: Rp+, Rp-, Nrp) mixed design. Experiment 2 was identical to Experiment 1 with two exceptions. Firstly, instead of a category cued recall final test, Experiment 2 used a category-plus-one letter stem cued recall test. Secondly, Experiment 2 had a 10-min period between retrieval practice and final test.

Participants

One hundred twenty-eight Iowa State University students participated in Experiment 2 in return for research credit. This resulted in thirty-two participants per learning condition.

Materials

The materials of Experiment 2 were identical to those of Experiment 1.

Procedure

The procedure of Experiment 2 was identical to that of Experiment 1 with two exceptions. Firstly, Experiment 2 used a category-plus-one letter stem cued recall final test (e.g., *Weather* – B____). Participants were presented with a category name and the first letter stem of a studied exemplar, and they were allowed 7 s to type in the studied exemplar. Every exemplar for each category was presented consecutively. The order of the categories was randomized. Within each category, the order of the exemplars was randomized except that the Rp- exemplars were probed before the Rp+ exemplars. This was done to minimize the contribution of output interference in producing retrieval-induced forgetting (Anderson et al., 1994). A 500 ms ISI separated each test trial. Secondly, Experiment 2 had a 10-minute delay between the learning phase and final test phase. During the 10-minute delay, participants played Tetris.

Results and discussion

Retrieval practice

Retrieval practice results are presented in Table 3. In the cumulative retrieval practice condition, participants recalled more exemplars on their third retrieval practice trial than their first, $t(31) = 4.76$, $p < .01$, $d = .43$. Participants in the high competition and low competition conditions also recalled more exemplars on the third retrieval practice trial ($M = .90$, $SD = .08$) than the first retrieval practice trial ($M = .86$, $SD = .10$), $t(63) = 5.29$, $p < .01$, $d = .44$.

Final test

Final test results for Experiment 2, which are broken down by retrieval practice status and learning condition, are displayed in Figure 5. Because the Rp+ exemplars and

Table 3

Mean Proportion Correct During Retrieval Practice for Experiment 2

	Block 1			Block 2			Block 3			Block 4		
Retrieval Practice Trial	1	2	3	1	2	3	1	2	3	1	2	3
High Competition	.77	.79	.82	.91	.92	.96	.82	.84	.86	.86	.88	.91
	(.19)	(.17)	(.18)	(.13)	(.14)	(.09)	(.17)	(.19)	(.17)	(.13)	(.15)	(.11)
Low Competition	.91	.93	.93	.92	.93	.94	.85	.85	.86	.87	.93	.94
	(.14)	(.13)	(.13)	(.10)	(.09)	(.08)	(.18)	(.19)	(.17)	(.16)	(.11)	(.12)
Cumulative Retrieval Practice	.75	.78	.81									
	(.12)	(.12)	(.12)									

Note: Standard deviations are displayed in parentheses.

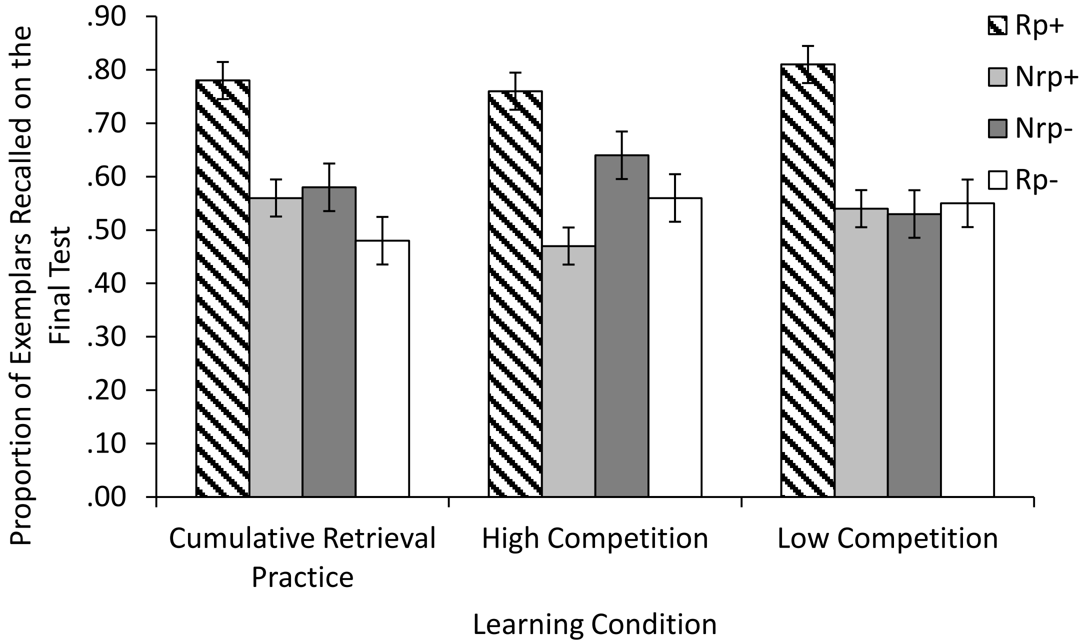

Figure 5. Proportion of exemplars recalled on the final test as a function of retrieval practice status and learning condition in Experiment 2. Error bars reflect the within-subjects 95% confidence interval.

Rp- exemplars had separate Nrp exemplars, their data were analyzed separately. To examine the influence of retrieval practice on the Rp+ exemplars, a 3 (learning condition: high competition, low competition, cumulative retrieval practice) X 2 (retrieval practice status: Rp+, Nrp+) mixed ANOVA was conducted. This ANOVA revealed a main effect of retrieval practice status, $F(1, 93) = 223.97$, $p < .01$, $\eta_p^2 = .71$. That is, there was a reliable testing effect as participants recalled Rp+ exemplars ($M = .78$, $SD = .16$) at a higher rate than Nrp+ exemplars ($M = .52$, $SD = .17$). There was no reliable main effect of learning condition, $F(2,93) = 2.03$, $p = .14$, or retrieval practice status by learning condition interaction, $F(2, 93) = 1.69$, $p = .19$.

To examine the influence of retrieval practice on the Rp- exemplars, I again examined potential retrieval-induced forgetting effects within each learning condition. Within the cumulative retrieval practice condition, participants recalled Nrp- exemplars (*M* = .58, *SD* = .17) at a higher rate than Rp- exemplars (*M* = .48, *SD* = .16), $t(31) = 2.89$, $p < .01$, $d = .61$. Within the high competition condition, participants recalled Nrp- exemplars (*M* = .64, *SD* = .23) at a higher rate than Rp- exemplars (*M* = .56, *SD* = .16), $t(31) = 2.24$, $p = .03$, $d = .40$. Within the low competition condition, participant's recall rate between Rp- exemplars and the Nrp- exemplars did not differ reliably, $t < 1$, $p = .61$. In other words, in the current experiment, I found a retrieval-induced forgetting effect in the cumulative retrieval practice condition and in the high competition condition. However, I did not find evidence of retrieval-induced forgetting in the low competition condition.

Rp- exemplars were presented prior to retrieval practice in the high competition condition, so, according to the inhibition account, the Rp- exemplars should compete for retrieval and thus be suppressed. However, in the low competition condition, the Rp- exemplars had yet to be presented when participants perform retrieval practice. Therefore, the Rp- exemplars were unlikely to compete during retrieval practice and consequently should not be the target of inhibition. Data in the current experiment is consistent with this interpretation as retrieval-induced forgetting was found in the high competition but not in the low competition condition.

The different results from Experiment 1 and Experiment 2 may be attributed to the different final tests. In Experiment 1, a category cued recall final test was employed, which did not control for output interference, whereas Experiment 2 employed a category-plus-stem cued recall test, which controlled for output interference. Output interference refers to intrusion of previously recalled exemplars when one tries to recall non-recalled, studied exemplars. In a category cued recall test, it is difficult (if not impossible) to differentiate the contribution of blocking from output interference in a retrieval-induced forgetting effect. Specifically, output interference refers to impairment from previously recalled (i.e., output) exemplars, while retrieval-induced forgetting resulting from blocking refers to impairment from Rp+ exemplar learning. For example, in the case of a category cued recall test, a participant will likely recall the better learned exemplars (the Rp+ exemplars) prior to Rp- exemplars. Therefore, the Rp+ exemplars may have intruded because Rp+ exemplars were better learned and were recently recalled.

Final test results for Experiment 2, broken down by block and learning condition are displayed in Figure 6. To examine the effect of retrieval practice on proactive interference, I again examined Nrp exemplar final test recall and conducted a 4 (block: 1, 2, 3, 4) X 3 (learning condition: high competition, low competition, interim math) mixed ANOVA. This ANOVA revealed a main effect of block, $F(1, 93) = 16.68$, $p < .01$, $\eta_p^2 = .15$, and a block by condition interaction, $F(2, 93) = 6.56$, $p < .01$, $\eta_p^2 = 6.56$. Unlike Experiment 1, the block by learning condition interaction was primarily driven by a lower recall rate for Nrp exemplars in Block 1 in the low competition condition ($M = .52$, $SD = .23$) than in the interim math condition ($M = .64$, $SD = .17$), $t(62) = 2.48$, $p = .02$, $d = .59$. When examining recall rates across blocks between the interim math and low competition conditions and the interim math and high competition conditions, no other reliable

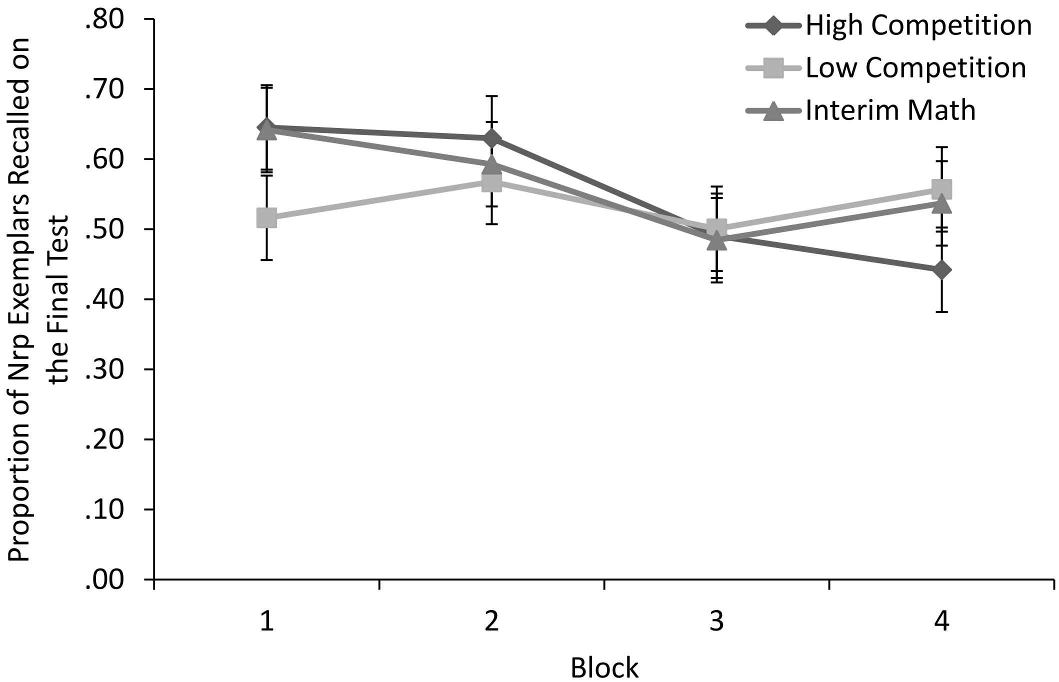

Figure 6. Proportion of Nrp exemplars recalled on the final test as a function of learning condition and block in Experiment 2. Error bars reflect the within-subjects 95% confidence interval.

differences were found, *t*s < 1.6, *p*s > .12. There was no main effect of learning condition, $F < 1$, $p = .76$.

Experiment 2 provided some evidence of the inhibition account as retrieval-induced forgetting was found in the high competition and cumulative retrieval practice conditions but not in the low competition condition. However, Experiment 2 changed both the final test format and distractor task. Though not predicted, there was a chance that the change in distractor task contributed to finding retrieval-induced forgetting in all three conditions of Experiment 1 but not Experiment 2. Therefore, in Experiment 3, I sought to remove distractor task as a potential explanation for the Experiment 2 data. Experiment 3 was the same as Experiment 2, except Experiment 3 had a 20-min distractor period that included both OSPAN and Tetris (similar to Experiment 1). In regards to the test-potentiated learning data, Experiment 2 also failed to find my predicted pattern of Nrp exemplar recall on the final test. That is, I expected to find progressively worse recall across blocks for Nrp exemplars in the interim math condition but not in the interim testing conditions. However, I found evidence of this pattern in both the interim math and high competition conditions but not in the low competition condition. I am unsure why I found this pattern. Therefore, I planned to again examine the effect of interim testing (compared interim math) on participant's ability to learn new information in order to, hopefully, find a consistent pattern.

CHAPTER 4. EXPERIMENT 3

In Experiment 2, I used a category-plus-stem cued recall final test and a 10-min distractor task that included only Tetris, but in Experiment 1 I used a category cued recall test and a 20-min distractor task that included Tetris and the OSPAN task. Therefore, the elimination of the retrieval-induced forgetting effect in the low competition condition may not have resulted from inhibition, but the change in delay and distractor task between the learning and final test phases. Experiment 3 sought to eliminate delay and distractor task as a potential explanation as to why retrieval-induced forgetting was found in the low competition condition of Experiment 1 but not Experiment 2. To that end, Experiment 3 was identical to Experiment 2 except that Experiment 3 had a 20-min delay between retrieval practice and final test that included the OSPAN task and Tetris (similar to Experiment 1).

Method

Design

Experiment 3 had a 4 (learning condition: high competition, low competition, cumulative retrieval practice, and interim math) X 3 (retrieval practice status: Rp+, Rp-, Nrp) mixed design.

Participants

One hundred twenty-eight Iowa State University students participated in Experiment 3 in return for research credit. This resulted in thirty-two participants per learning condition.

Materials

The materials of Experiment 3 were identical to those of Experiment 2.

Procedure

The procedure of Experiment 3 was identical to Experiment 2 with one exception: Experiment 3 had a 20-min delay between the learning phase and final test. During the distractor phase, participants performed the OSPAN task and then played Tetris for a total of 20 min.

Results and discussion

Retrieval practice

Retrieval practice results are displayed in Table 4. In the cumulative retrieval practice condition, participants recalled more exemplars on their third retrieval practice trial than their first, $t(31) = 4.27$, $p < .01$, $d = .28$. Participants in the high competition and low competition conditions also recalled more exemplars on the third retrieval practice trial ($M = .91$, $SD = .13$) than the first retrieval practice trial ($M = .87$, $SD = .12$), $t(63) = 2.92$, $p = .01$, $d = .32$.

Final test

Final test results for Experiment 3, broken down by retrieval practice status and learning condition, are displayed in Figure 7. A 3 (learning condition: high competition, low competition, cumulative retrieval practice) X 3 (retrieval practice status: Rp+, Nrp+) mixed ANOVA revealed a main effect of retrieval practice status, $F(1, 93) = 188.74$, $p < .01$, $\eta_p^2 = .67$. That is, a reliable testing effect was found as participants recalled more Rp+ exemplars ($M = .79$, $SD = .18$) than Nrp+ exemplars ($M = .51$, $SD = .18$). There was no reliable main effect of learning condition, $F(2, 93) = 2.13$, $p = .13$, or learning condition by retrieval practice status interaction, $F < 1$.

Table 4

Mean Proportion Correct During Retrieval Practice for Experiment 3

	Block 1			Block 2			Block 3			Block 4		
Retrieval Practice Trial	1	2	3	1	2	3	1	2	3	1	2	3
High Competition	.79	.84	.87	.94	.96	.96	.84	.92	.94	.88	.90	.91
	(.19)	(.20)	(.17)	(.14)	(.13)	(.11)	(.17)	(.17)	(.13)	(.16)	(.17)	(.16)
Low Competition	.88	.91	.91	.86	.89	.90	.83	.85	.86	.86	.91	.92
	(.21)	(.15)	(.20)	(.20)	(.22)	(.21)	(.14)	(.19)	(.19)	(.15)	(.13)	(.15)
Cumulative Retrieval Practice	.73	.76	.77									
	(.17)	(.17)	(.16)									

Note: Standard deviations are displayed in parentheses.

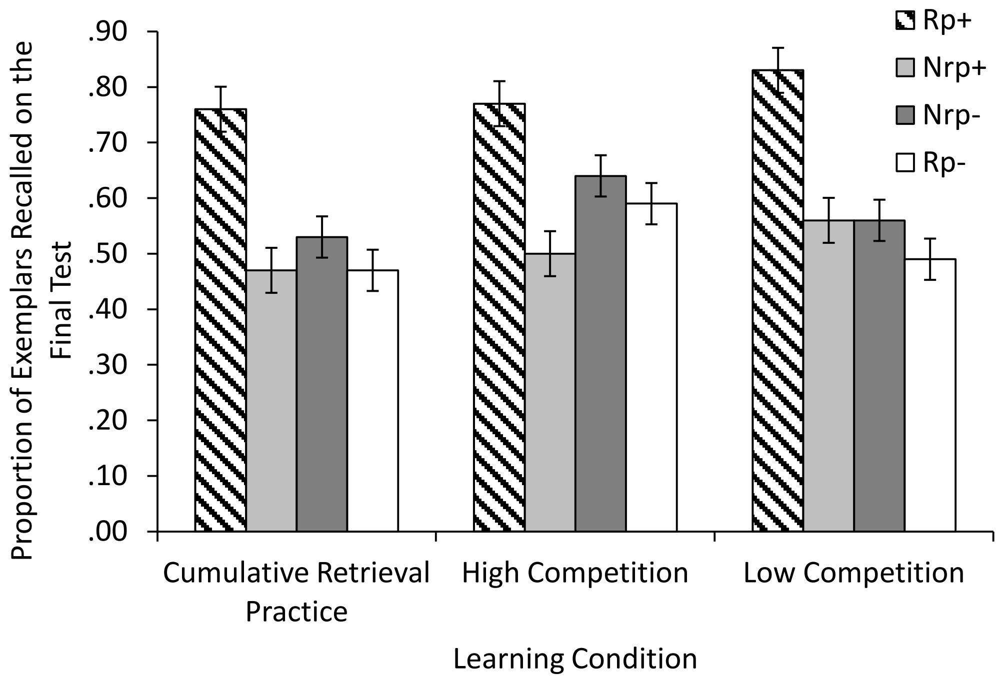

Figure 7. Proportion of exemplars recalled on the final test as a function of retrieval practice status and learning condition in Experiment 3. Error bars reflect the within-subjects 95% confidence interval.

To examine the influence of retrieval practice on the Rp- exemplars, I again examined potential retrieval-induced forgetting effects within each learning condition. Within the cumulative retrieval practice condition, participants did not recall Nrp- exemplars (M = .53, SD = .15) at a reliably higher rate than Rp- exemplars (M = .47, SD = .21), $t(31)$ = 1.66, p = .11. Within the high competition condition, participants did not recall Nrp- exemplars (M = .64, SD = .20) at a reliably higher rate than Rp- exemplars (M = .59, SD = .16), $t(31)$ = 1.57, p = .13. Within the low competition condition, participant's did recall Nrp- exemplars (M = .56, SD = .17) at a higher rate than Rp- exemplars (M = .49, SD = .15), t = 2.23, p = .03, d = .41. In other words, in the current experiment, I found a significant retrieval-induced forgetting effect in the low competition condition but not the cumulative retrieval practice or high competition condition.

Curiously, this does not replicate the results of Experiments 1 or 2. However, because participants in the retrieval-induced forgetting condition of Experiment 3 failed to demonstrate a retrieval-induced forgetting effect, the results of Experiment 2 may be more reliable. That is, the results of Experiment 3 not only do not replicate the current dissertation's work but much other literature as no retrieval-induced forgetting was found in the cumulative retrieval practice condition (see Anderson, 2003).

Final test results for Experiment 3, broken down by block and learning condition, are displayed in Figure 8. In order to examine the effect of retrieval practice on proactive interference, I again examined Nrp exemplar recall on the final test based on block. A 4 (block: 1, 2, 3, 4) X 3 (learning condition: high competition, low competition, interim math) ANOVA revealed a main effect of block, $F(3, 279) = 7.07$, $p < .01$, $\eta_p^2 = .07$, and a block by learning condition interaction, $F(6, 279) = 2.59$, $p = .02$, $\eta_p^2 = .05$. The block by learning condition interaction was primarily driven by a lower Nrp exemplar recall rate in the third block of the interim math condition ($M = .40$, $SD = .25$) than the low competition condition ($M = .55$, $SD = .20$), $t(62) = 2.67$, $p = .01$, $d = .66$, and a lower first block Nrp exemplar recall rate in the low competition condition ($M = .50$, $SD = .22$) than in the high competition condition ($M = .64$, $SD = .24$), $t(62) = 2.44$, $p = .02$, $d = .44$. No other differences were reliable, *t*s < 2, *p*s > .05. It is unclear why these results did not replicate Experiments 1 and 2.

These results are difficult to interpret. Although participants in the low competition condition appeared to show no recall deficit due to proactive interference,

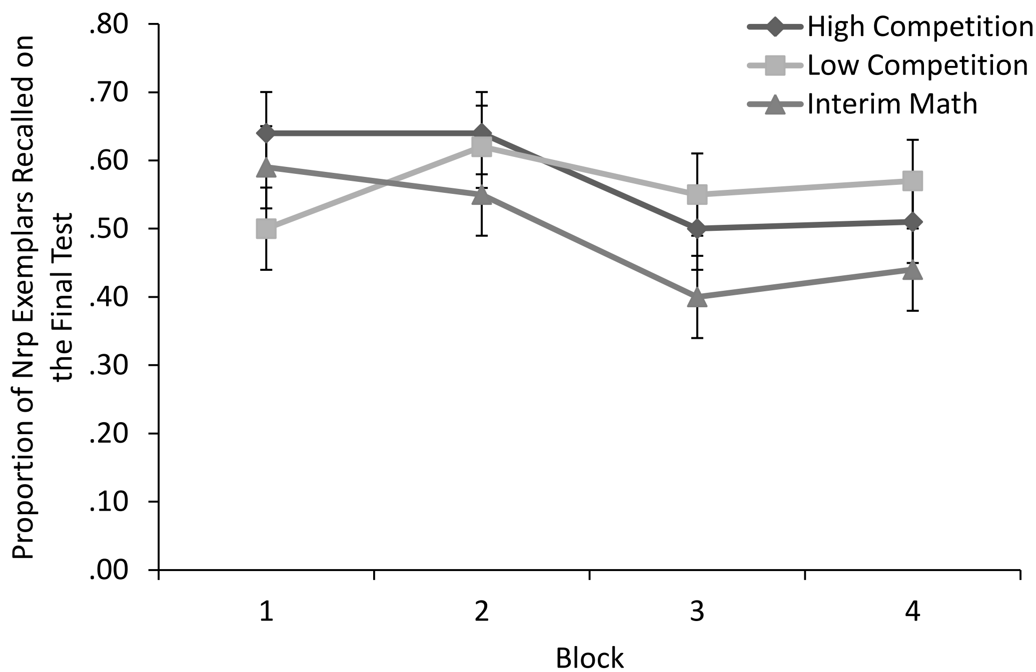

Figure 8. Proportion of Nrp exemplars recalled on the final test as a function of learning condition and block in Experiment 3. Error bars reflect the within-subjects 95% confidence interval.

they had a lower Block 1 recall rate than participants in the other conditions (.52 recall rate versus approximately .64 for the other conditions). Nonetheless, participants in the low competition condition may not have experienced proactive interference because participant's recall rate does not decline across blocks. However, it is difficult to compare the low competition condition to the high competition or interim math conditions because the low competition has a lower baseline (as seen in the Block 1 recall rates). Because of this lower baseline in the low competition condition, the current experiment should be replicated to ensure the data did not occur by chance.

CHAPTER 5. GENERAL DISCUSSION

Overview of the current dissertation

In three experiments, I examined how retrieval-induced forgetting was affected when participants used one learning block to learn exemplars versus multiple smaller blocks. Furthermore, I examined how testing a subset of the previously studied exemplars, within these smaller blocks, affected future learning. Experiment 1 used a category cued recall test and found a reliable retrieval-induced forgetting effect in the cumulative retrieval practice, high competition, and low competition conditions. Notably, this would be predicted from the blocking account of retrieval-induced forgetting. However, in Experiment 2, when participants were given a 10-min distractor task and a category-plus-stem cued recall final test, retrieval-induced forgetting was demonstrated by participants in the cumulative retrieval practice condition and in the high competition condition, but participants did not demonstrate retrieval-induced forgetting in the low competition condition. The results of Experiment 2 are consistent with the inhibition account of retrieval-induced forgetting. In Experiment 3, participants in the cumulative retrieval practice did not demonstrate retrieval-induced forgetting. Therefore, the results of Experiment 3 did not replicate prior research (Anderson, 2003). With regards to the test-potentiated learning effect, there was some evidence in Experiment 1 that, with interim math, participants recalled fewer Nrp exemplars in later blocks (compared to the low competition and high competition condition), but this trend was not reliable in Experiment 1 and was not replicated in Experiments 2 and 3.

The blocking and inhibition accounts

Though the current dissertation is not the first to suggest that retrieval-induced forgetting may stem from a combination of inhibition and blocking effects (see Storm & Levy, 2012), it provides evidence both for the inhibition and blocking accounts. That is, in Experiment 1, with a category-cued recall test, I found evidence for the blocking account. In Experiment 2, with a category-plus-stem cued recall test, I found evidence for the inhibition account. However, the results from Experiment 3 were inconclusive as no established accounts of retrieval-induced forgetting would lead one to predict a significant retrieval-induced forgetting effect in only the low competition condition. It is unclear why this discrepancy occurred.

Because the results of Experiment 2 were not replicated in Experiment 3, there is a possibility that the nature of the tasks in the distractor period or length of the distractor period may have contributed to the results that were found in Experiment 2. If length of the distractor period is the reason for the discrepancy, there is little literature to suggest such a pattern. Alternatively, it must be acknowledged that the data from Experiment 3 (and by extension Experiment 2) might have occurred due to unexplained sampling error. It is presently unclear whether the results from Experiment 2 or Experiment 3 more accurately represent “the truth.” As a result, the first order of business for future research on this topic is to replicate Experiment 3. Here I highlight the differences between Experiment 2 and Experiment 3 in an attempt to identify the possible reasons for discrepancies between the data from these experiments – assuming that these discrepancies are “real.”

In the current dissertation, Experiment 2 had a 10-min distractor period, and Experiment 3 had a 20-min distractor period. Notably, participants in Experiment 2 demonstrated a reliable retrieval-induced forgetting effect only in the high competition and cumulative retrieval practice conditions, but participants in Experiment 3 demonstrated retrieval-induced forgetting only in the low competition condition. Many researchers have used a distractor period that is shorter than 20 min between retrieval practice and final test (see Aslan & Bäuml, 2010; Hulbert et al., 2011; Potts et al., 2011; Storm & Nestojko, 2012). Because many researchers have found retrieval-induced forgetting and (more recently) seem to be adopting this shorter length, this shorter delay may more reliably create retrieval-induced forgetting (for studies that demonstrate retrieval-induced forgetting with a 20-min delay, see Anderson et al., 1994; Chan, 2009). Though Chan (2009) experimentally manipulated short (e.g., 20 min) versus long delays (e.g., 24 hr), to my knowledge, no studies have experimentally manipulated the influence of smaller differences in delay on the magnitude of retrieval-induced forgetting effects (e.g., 10-min versus 20-min versus 60-min distractor period delays). Because there is the possibility that shorter delays lead to more reliable effects (or somehow differentially create predominantly blocking-based or inhibition-based retrieval-induced forgetting effects), distractor delay should be more closely examined. Nonetheless, one must acknowledge that there were other differences aside from distractor delay between Experiments 2 and 3, such as the participants and the administration of the OSPAN task prior to final test.

I initially predicted that if the retrieval-induced forgetting effect was driven by inhibition, I should find evidence for retrieval-induced forgetting in the high competition but not in the low competition condition. That is, I predicted that the Rp- exemplars must be presented prior to retrieval practice of the Rp+ exemplars in order for Rp- exemplars to compete for recall. Though the high competition and cumulative retrieval practice conditions in Experiment 2 provided some evidence of this pattern, Experiment 3 did not demonstrate this predicted pattern. Proponents of the inhibition account (Anderson, 2003; Anderson & Levy, 2009) define competitors as any response that needs to be inhibited in order to allow for weaker, more contextually appropriate responses to be recalled. I suspect the inconsistency in obtaining retrieval-induced forgetting effect in the high competition and low competition conditions in Experiments 2 and 3 could have resulted from issues with exemplar competitiveness.

According to the inhibition account, for retrieval-induced forgetting to occur, Rp- and Rp+ exemplars must compete for access, so exemplars must be similar enough or probed from a shared cue. If competitiveness was solely based on the number of shared characteristics between two exemplars, one would have expected retrieval-induced forgetting in the low competition condition in Experiment 2. Notably, Anderson, Green, and McCulloch (2000) argued that participants that study Rp- exemplars that have a high amount of similarity (e.g., a large number of shared characteristics) with other Rp- exemplars can produce retrieval-induced forgetting, but exemplars that share many characteristics with Rp+ exemplars will not produce retrieval-induced forgetting (see also Goodmon & Anderson, 2011). Nevertheless, this brings up an interesting conundrum. Exemplars must be similar enough to compete for recall, but not too similar as to be "integrated." Though this may need to be more clearly defined, I also posit that there may be other factors that dictate what a competitor is, and future researchers may want to examine what these factors are. Other factors that may influence similarity could include the amount of time that occurred between the presentations of two exemplars within the same category. In fact, some researchers have argued that the amount of interference that results from the memory of other information can be predicted solely by the amount of time that has occurred between the learning of items (Brown & Lewandowsky, 2010). Therefore, in the current dissertation, interim testing may have reduced the amount of interference from Rp- exemplars (compared to cumulative retrieval practice) simply by having a longer delay between presentation of the Rp- and Rp+ exemplars. Specifically, Rp+ and Rp- exemplars were initially studied within the same block in the cumulative retrieval practice conditions but different blocks in the interim testing conditions.

Other than the amount of time that occurred between presentations of exemplars, Pastötter et al. (2008) also suggested that retrieval may trigger an internal context change. If participants are changing their internal context with every set of retrieval practice, participants may be able to selectively recall exemplars based by searching their memory for exemplars that most closely match their current internal context. Therefore, exemplars from prior blocks would not be competitors (or less likely to be competitors), and one would not expect to find retrieval-induced forgetting within interim testing. Interestingly, Pastötter et al. were not the first to suggest that people may be able to limit a memory search to a specific source. Jacoby et al. (2005) have also found evidence of this source-constrained retrieval.

Future researchers should further examine what creates competition among exemplars. Exploring this area may help reduce the number of nuisance variables in retrieval-induced forgetting experiments and allow for cleaner data. In the current dissertation, learning condition may be confounded with degree of internal context change, and both variables may have contributed to the degree that two exemplars are competitors. For example, in the high competition condition (and similar issues could occur in the low competition condition), participants studied Rp- exemplars in Blocks 1 and 2, but they would not receive retrieval practice over the Rp+ exemplars until Blocks 3 and 4. Therefore, participants may have shifted their internal context after learning the Rp- exemplars and prior to learning the Rp+ exemplars. In the cumulative retrieval practice condition, participants would not have the opportunity to shift their internal context prior to Rp+ exemplar retrieval practice because this condition only has one learning block. Therefore, researchers may want to experimentally examine degree of internal context change – perhaps by manipulating the degree of internal context change between Rp+ exemplar learning blocks and Rp- exemplar learning blocks (e.g., asking participants to perform semantic generation as well as retrieval practice after a learning block, a larger internal context change condition, versus only performing retrieval practice, a smaller internal context change condition). This may help researchers better understand if the degree of internal context change substantially contributed to the learning condition variable in the current dissertation.

Furthermore, inhibition account proponents may want to examine whether the delay between learning Rp- and Rp+ exemplars, internal context change, or other variables could influence exemplar competitiveness, especially if they wish to better understand how inhibition-based retrieval induced forgetting affects our memory of everyday events, such as where we parked our car, or intrusive memories, which we may

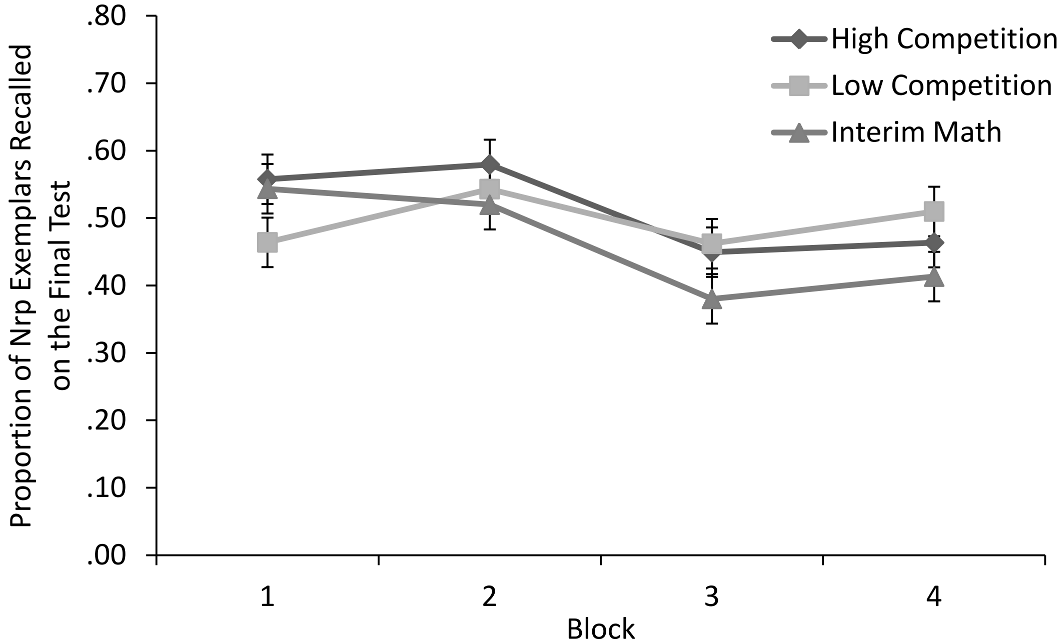

Figure 9. Proportion of Nrp exemplars recall final test as a function of learning condition and block across Experiments 1, 2, and 3. Error bars reflect the within-subjects 95% confidence interval.

want to forget, such as Hurricane Katrina (for further explanation, see Anderson & Levy, 2009).

Test-potentiated learning

I initially predicted that retrieval practice recall rates would decline with interim math, but retrieval practice recall rates would not decline with interim testing (e.g., the high competition and low competition conditions). However, I did not find reliable evidence of such a pattern. Because I suspected a lack of power for my inconsistent results, I collapsed my data across all three experiments to look for patterns. As shown in Figure 9, participants may have experienced less proactive interference in Blocks 3 and 4 in the high competition and low competition conditions compared to the interim math condition. In fact, participants recalled more Block 3 Nrp exemplars, $t(190) = 2.40$, $p = .02$, $d = .34$, and more Block 4 Nrp exemplars in the low competition condition than in the interim math condition, $t(190) = 2.61$, $p = .01$, $d = .38$. Participants in the high competition condition recalled more Nrp exemplars in Block 3 than participants in the interim math condition, $t(190) = 2.00$, $p = .05$, $d = .29$, but the difference between these two conditions was not reliable in Block 4, $t(190) = 1.43$, $p = .16$. This provides some evidence that interim testing a subset of previously learned information may lead to test-potentiated learning. Furthermore, the overall pattern (better Block 3 and Block 4 Nrp exemplar recall in the interim test conditions compared to the interim math condition) is consistent with my initial predictions.

However, this pattern still does not match what I initially predicted because the pattern was unreliable in the high competition condition, and participants in the low competition condition did inexplicably worse in Block 1 than participants in the high competition and interim math conditions. These issues may have resulted from a few different sources. Firstly, there may be an unknown problem with the current study's materials. Though there was no reliable difference between the word frequency or taxonomic frequency of the target and filler exemplar, there may be other differences between the two sets (or I may not have had enough power to detect differences between the target and filler pairs). For instance, exemplars from target pairs were taken from categories in which Battig and Montagues' (1944) participants generated more exemplars compared to the filler categories. Therefore, this systematic difference may have caused a systematic difference in how participants elected to study the materials or the ease of learning the materials. Therefore, future research should attempt to extend the current dissertation's findings with Szpunar et al.'s (2008) materials. Secondly, the current dissertation's data may not be as reliable as previous studies due to the number of tested items in each block. Szpunar et al. (2008) asked participants to perform free recall initial tests over each block, which had 18 items. The current dissertation based the test-potentiated learning data on only six Nrp exemplars per block.

In order to find a clearer pattern, future researchers may way want to follow up on the current dissertation in a couple ways. Firstly, future researchers may want to replicate the current experiments with more exemplars in each block. Secondly, researchers should examine the effect of interim testing when participants learn a set of unrelated exemplars. In the current dissertation, participants learned many filler, Rp+, and Rp- exemplars. These exemplars (because their potential contamination from retrieval-induced forgetting) was not examined, which excluded half of the target exemplars and all of the filler exemplars. Furthermore, Szpunar et al. (2008) focused on how initial test performance was improved in later blocks for interim testing conditions compared to earlier blocks. With the current dissertation, initial test performance was not examined because of potential differences (both in characteristics of the exemplars and the exposure to related exemplars in other blocks) between filler and Rp+ exemplars, so I examined participant's final test recall of Nrp exemplars. However, if one were to use all unrelated exemplars, one could examine how participant's performance is affected during initial test as well as final test for all exemplars.

Concluding remarks

The current dissertation was a first attempt to examine a few questions that have, to the best of my knowledge, been unanswered in the literature. Firstly, does breaking up a learning block into multiple learning blocks affect retrieval-induced forgetting? Secondly, does testing a subset of a block of information still reduce proactive interference of information that was studied in a later block? The answers to these questions are inconclusive from the current dissertation. When participants were given a category cued final test, participants demonstrated retrieval-induced forgetting in the high competition, low competition, and cumulative retrieval-induced forgetting conditions. However, when participants were given a category-plus-stem cued recall final test, the results were mixed. In Experiment 2, when participants were given a 10-min distractor task, they demonstrated retrieval-induced forgetting in only the high competition and cumulative retrieval-induced forgetting conditions, but when participants were given a 20-min distractor task, they demonstrated retrieval-induced forgetting only in the low competition condition. I conclude that the retrieval-induced forgetting literature may benefit from further exploring how competitors are defined to reduce the amount of nuisance variables in experiments and produce cleaner data.

The results from the test-potentiated learning analyses were also unclear. However, if one looks at the data collapsed across all three experiments, one may notice that the overall data showed that participants in interim testing conditions tended to recall more Nrp exemplars in Blocks 3 and 4 than participants in the interim math condition. However, some of these differences were unreliable, and there was a depressed baseline recall in the low competition condition. In the future, researchers may benefit from using more exemplars per block and using unrelated exemplars. Researchers should continue to develop their understanding of test-potentiated learning and retrieval-induced forgetting, for greater understanding of these areas of cognitive psychology may allow for more efficient learning with students.

APPENDIX A. EXPERIMENTAL CATEGORY-EXEMPLAR PAIRS

Category	Exemplar	Taxonomic Frequency	Thorndike-Lorge Word Frequency
Weapon	Pistol	.92	119
	Tank	.08	84
	Sword	.25	91
	Club	.25	999
	Rifle	.37	181
	Bomb	.28	137
Metal	Iron	.80	454
	Nickel	.15	39
	Gold	.61	712
	Silver	.57	334
	Aluminum	.59	40
	Brass	.22	100
Profession	Farmer	.06	519
	Dentist	.25	61
	Nurse	.11	714
	Plumber	.09	28
	Engineer	.25	218
	Accountant	.08	18
Fruit	Strawberry	.13	121
	Lemon	.30	301
	Orange	.88	351
	Tomato	.14	166
	Banana	.64	-
	Pineapple	.22	235
Insect	Hornet	.07	9
	Fly	.76	634
	Roach	.28	14
	Beetle	.36	14
	Mosquito	.51	34
	Grasshopper	.23	20
Fish	Guppy	.12	-
	Bluegill	.14	-
	Trout	.49	29
	Herring	.36	66
	Catfish	.32	32

	Flounder	.11	34
Drink	Ale	.11	28
	Rum	.38	30
	Vodka	.61	-
	Whiskey	.73	97
	Bourbon	.50	-
	Gin	.70	51
Tree	Elm	.48	48
	Redwood	.16	11
	Dogwood	.19	11
	Birch	.30	34
	Hickory	.11	19
	Spruce	.17	40

APPENDIX B. FILLER CATEGORY-EXEMPLAR PAIRS

Category	Exemplar	Taxonomic Frequency	Thorndike-Lorge Word Frequency
Toy	Wagon	.10	325
	Jacks	.09	-
	Puzzle	.09	278
Weather	Lightning	.16	101
	Typhoon	.13	-
	Blizzard	.10	27
Spice	Oregano	.17	-
	Mustard	.14	51
	Nutmeg	.10	21
Body	Nose	.64	478
	Finger	.63	858
	Ear	.59	595
Bird	Crow	.34	43
	Bluebird	.31	-
	Parakeet	.26	-
Flower	Orchid	.31	44
	Lily	.24	164
	Pansy	.24	39
Clothing	Skirt	.59	297
	Coat	.59	896
	Hat	.45	976
Animal	Tiger	.46	103
	Elephant	.41	144
	Pig	.32	75

REFERENCES

Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology, 22,* 861-876. DOI: 10.1002/acp.1391

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language, 49*, 415-445. DOI: 10.1016/j.jml.2003.08.006

Anderson, M. C. & Bell, T. (2001). Forgetting our facts: The role of inhibitory processes in the loss of propositional knowledge. *Journal of Experimental Psychology: General, 130,* 544-570. DOI: 10.1037/0096-3445.130.3.544

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1063-1087. DOI: 10.1037/0278-7393.20.5.1063

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (2000). Retrieval-induced forgetting: Evidence for a recall-specific mechanism. *Psychonomic Bulletin & Review, 7* (3), 522-530. DOI: 10.3758/s13421-011-0131-y

Anderson, M. C. & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature, 410*, 366-369. DOI: 10.1038/35066572

Anderson, M. C., Green, C., & McCulloch, K. C. (2000). Similarity and inhibition in long-term memory: Evidence for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1141-1159. DOI: 10.1037/0278-7393.26.5.1141

Anderson, M. C. & Levy, B. J. (2009). Suppressing unwanted memories. *Current Directions in Psychological Science, 18* (4), 189-194. DOI: 10.1111/j.1467-8721.2009.01634.x

Anderson, M. C. & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review, 102*, 68-100. DOI: 10.1037/0033-295X.102.1.68

Aslan, A. & Bäuml, K.-H. T. (2012). Retrieval-induced forgetting in old and very old age. *Psychology and Aging, 27*, 1027-1032. DOI: 10.1037/a0028379

Aslan, A. & Bäuml, K.-H. T. (2010). Retrieval-induced forgetting in young children. *Psychonomic Bulletin & Review*, *17*, 704-709. DOI: 10.3758/PBR.17.5.704

Aslan, A. & Bäuml, K.-H. T. (2011). Individual differences in working memory capacity predict retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 264-269. DOI: 10.1037/a0021324

Aslan, A., Bäuml, K.-H. T., & Pastötter, B. (2007). No inhibitory deficit in older adults' episodic memory. *Psychological Science, 18,* 72-78. DOI: 10.1111/j.1467-9280.2007.01851.x

Barnes, J. M. & Underwood, B. J. (1959). "Fate" of first-list associations in transfer theory. *Journal of Experimental Psychology, 58,* 97-105. DOI: 10.1037/h0047507

Battig, W. F. & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology, 80,* 1-46. DOI: 10.1037/h0027577

Bäuml, K.-H. T. (1998). Strong items get suppressed, weak items do not: The role of item strength in output interference. *Psychonomic Bulletin & Review, 5,* 459-463. DOI: 10.3758/BF03208822

Bäuml, K.-H. T., & Spitzer, B. (2009). Retrieval-induced forgetting in a category recognition task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 286-291. DOI: 10.1037/a0014363

Bell, T. A. (2005). *Individual differences in memory inhibition* (Unpublished dissertation). University of Oregon, United States.

Bjorklund, D. F. & Harnishfeger, K. K. (1990). The resources construct in cognitive development: Diverse sources of evidence and a theory of inefficient inhibition. *Developmental Review, 10,* 48-71. DOI: 10.1016/0273-2297(90)90004-N

Brown, G., & Lewandowsky, S. (2010). Forgetting in memory models: Arguments against trace decay and consolidation failure. In S. D. Sala (Ed.), *Forgetting: Current issues in psychology* (pp. 253-284), New York, NY: Psychology Press.

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 918-928. DOI: 10.1037/0278-7393.34.4.918

Butler, A. C. & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition, 36,* 604-616. DOI: 10.3758/MC.36.3.604

Camp, G., Pecher, D., & Schmidt, H. G. (2007). No retrieval-induced forgetting using item-specific independent cues: Evidence against a general inhibitory account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33,* 950-958. DOI: 10.1037/0278-7393.33.5.950

Camp, G., Pecher, D., & Schmidt, H. G., Zeelenberg, R. (2009). Are independent probes truly independent? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 934-942. DOI: 10.1037/a0015536

Carpenter, S. K., & Pashler, H. (2007). Testing beyond words: Using tests to enhance visuospatial map learning. *Psychonomic Bulletin & Review, 14* (3), 474-478. DOI: 10.3758/BF03194092

Carpenter, S. K., Pashler, H., & Vul E. (2006). What types of learning are enhanced by a cued recall test? *Psychonomic Bulletin & Review, 13,* 826-830. DOI: 10.3758/BF03194004

Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, *36* (2), 438-448. DOI: 10.3758/MC.36.2.438

Carroll, M., Campbell-Ratcliffe, J., Murnane, H., & Perfect, T. (2007). Retrieval-induced forgetting in educational contexts: Monitoring, expertise, text integration, and test format. *European Journal of Cognitive Psychology, 19*, 580-606. DOI: 10.1080/09541440701326071

Chan, J. C. K. (2009). When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language, 61,* 153-170. DOI: 10.1016/j.jml.2009.04.004

Chan, J. C. K., McDermott, K. B., & Roediger, H. L. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General, 135,* 553-571. DOI: 10.1037/0096-3445.135.4.553

Darley, C. F. & Murdock, B. B. (1971). Effects of prior free recall testing on final recall and recognition. *Journal of Experimental Psychology, 91*, 66-73. DOI: 10.1037/h0031836

Dodd, M. D., Castel, A. D., & Roberts, K. E. (2006). A strategy disruption component to retrieval-induced forgetting. *Memory & Cognition, 24*(1), 102-111. DOI: 10.3758/BF03193390

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*, 19-23. DOI: 10.1111/1467-8721.00160

Erdelyi, M. H. (2010). The ups and downs of memory. *American Psychologist, 65,* 623-633. DOI: 10.1037/a0020440

Erdelyi, M. H. & Becker, J. (1974). Hypermnesia for pictures: Incremental memory for pictures but not words in multiple recall trials. *Cognitive Psychology, 6*, 159-171. DOI: 10.1016/0010-0285(74)90008-5

Erdman, M. R. (2011). *The influence of corrective feedback on retrieval-induced forgetting* (Unpublished thesis). Iowa State University, United States.

Erdman, M. R. & Chan, J. C. K. (2013). Providing corrective feedback during retrieval practice does not increase retrieval-induced forgetting. *Journal of Cognitive Psychology*.

Garcia-Bajos, E., Migueles, M. & Anderson, M. C. (2009). Script knowledge modulates retrieval-induced forgetting for eyewitness events. *Memory, 17* (1), 92-103. DOI: 10.1080/09658210802572454

Glover, J. A. (1989). The “testing” phenomenon: Not gone but nearly forgotten. *Journal of Educational Research, 81,* 392-399. DOI: 10.1037/0022-0663.81.3.392

Gómez-Ariza, C. J., Lechuga, M. T., Pelegrina, S., & Bajo, M. T. (2005). Retrieval-induced forgetting in recall and recognition of thematically unrelated sentences. *Memory & Cognition, 33*, 1431-1141. DOI: 10.3758/BF03193376

Harris, C. B., Sutton, J., & Barnier, A. J. (2010). Autobiographical forgetting, social forgetting, and situated forgetting: Forgetting in context. In S. D. Sala (Ed.), *Forgetting: Current issues in psychology* (pp. 253-284), New York, NY: Psychology Press.

Hicks, J. L. & Starns, J. J. (2004). Retrieval-induced forgetting occurs in tests of item recognition. *Psychonomic Bulletin & Review, 11* (1), 125-130. DOI: 10.3758/BF03206471

Hulbert, J. C., Shivde, G., & Anderson, M. C. (2011). Evidence against associative blocking as a cause of cue-independent retrieval-induced forgetting. *Experimental Psychology, 59,* 11-21. DOI: 10.1027/1618-3169/a000120

Huddleston, E. & Anderson, M. C. (2012). Reassessing critiques of the independent probe method for studying inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1408-1418. DOI: 10.1037/a0027092

Izawa, C. (1971). The test trial potentiating model. *Journal of Mathematical Psychology, 8,* 200-224. DOI: 10.1016/0022-2496(71)90012-5

Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychonomic Bulletin & Review, 12* (5), 852-857. DOI: 10.3758/BF03196776

Jakab, E. & Raaijmakers, J. G. W. (2009). The role of item strength in retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 607-617. DOI: 10.1037/a0015264

Jonker, T. R., Seli, P., MacLeod, C. M. (2012). Less we forget: Retrieval cues and release from retrieval-induced forgetting. *Memory & Cognition, 40,* 1236-1245. DOI: 10.3758/s13421-012-0224-2

Koriat, A. & Goldsmith, M. (1996). Monitoring and control processes in strategic regulation of memory accuracy. *Psychological Review, 103,* 490-517. DOI: 10.1037/0033-295X.103.3.490

Koutstaal, W., Schacter, D. L., Johnson, M. K. & Galluccio, L. (1999). Facilitation and impairment of even memory produced by photograph review. *Memory & Cognition, 27* (3), 478-493. DOI: 10.3758/BF03211542

Little, J., Storm, B. C., & Bjork, E. L. (2011). The costs and benefits of testing text materials. *Memory, 19,* 346-359. DOI: 10.1080/09658211.2011.569725

Leeming, F. C. (2002). The exam-a-day procedure improves performance in psychology classes. *Teaching of Psychology, 28,* 210-212. DOI: 10.1207/S15328023TOP2903_06

Levy, B. J. & Anderson, M. C. (2008). Individual differences in the suppression of unwanted memories: The executive deficit hypothesis. *Acta Psychologica, 127,* 623-635. DOI: 10.1016/j.actpsy.2007.12.004

Lustig, C., Hasher, L., & Zacks, R. T. (2007). Inhibitory deficit theory: Recent developments in a “new view.” In D. S. Gorfein & C. M. MacLeod (Eds.), The place of inhibition in cognition (pp. 145-162), Washington, DC: American Psychological Association.

MacLeod, M. D. (2002). Retrieval-induced forgetting in eyewitness memory: Forgetting as a consequence of remembering. *Applied Cognitive Psychology, 16,* 135-149. DOI: 10.1002/acp.782

Macrae, C. N., & MacLeod, M. D. (1999). On recollections lost: When practice makes imperfect. *Journal of Personality and Social Psychology, 77*, 463-473. DOI: 10.1037/0022-3514.77.3.463

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology, 19,* 494-513. DOI: 10.1080/09541440701326154

Nungester, R. J. & Duchastel, P. C. (1982). Testing versus review: Effects on retention. *Journal of Educational Psychology, 74*, 18-22. DOI: 10.1037/0022-0663.74.1.18

Palva, S., & Palva, J. M. (2007). New vistas for α-frequency band oscilliations. *Trends in Neurosciences, 30*, 150-158. DOI: 10.1016/j.tins.2007.02.001

Pashler, H. Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate the learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 3-8. DOI: 10.1037/0278-7393.31.1.3

Pastötter, B., Schicker, S., Niedernhuber, J. & Bäuml, K.-H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 287-297. DOI: 10.1037/a0021801

Postman, L. (1971). Transfer, interference, and forgetting. In L. King & L. A. Riggs (Eds.), *Experimental Psychology* (pp. 1019-1132), New York: Holt, Rinehart, & Winston.

Potts, R., Law, R., Golding, J. F., & Groome, D. (2011). The reliability of retrieval-induced forgetting. *European Psychologist, 17,* 1-10. DOI: 10.1027/1016-9040/a000040

Raaijmakers, J. G. W. & Jakab, E. (2012). Retrieval-induced forgetting without competition: Testing the retrieval specificity assumption of the inhibition theory. *Memory & Cognition, 40*, 19-27. DOI: 10.3758/s13421-011-0131-y

Raaijmakers, J. G. W. & Jakab, E. (2013). Rethinking inhibition theory: On the problematic status of the inhibition theory for forgetting. *Journal of Memory and Language, 68*, 98-122. DOI: 10.1016/j.jml.2012.10.002

Roediger, H. L. & Karpicke, J. D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249-255. DOI: 10.1111/j.1467-9280.2006.01693.x

Roediger, H. L. & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1,* 181-210. DOI: 10.1111/j.1745-6916.2006.00012.x

Roediger, H. L., Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In J. Mestre & B. Ross (Eds.), Psychology of learning and motivation: Cognition in education (pp. 1-36), Oxford: Elsevier.

Roediger, H. L., Weinstein, Y., & Agarwal, P. K. (2010). Forgetting: Preliminary considerations. In S. D. Sala (Ed.), *Forgetting: Current issues in psychology* (pp. 1-22), New York, NY: Psychology Press.

Román, P., Soriano, M. F., Gómez-Ariza, C. J., & Bajo, M. T. (2009). Retrieval-induced forgetting and executive control. *Psychological Science, 20* (9), 1053-1058. DOI: 10.1111/j.1467-9280.2009.02415.x

Schacter, D. L. (2001). *The seven sins of memory: How the mind forgets and remembers*. New York, NY: Houghton Mifflin Company.

Shivde, G. & Anderson, M. C. (2001). The role in meaning selection: Insights from retrieval-induced forgetting. In D. S. Gorfein (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 175-190). Washington, D.C.: American Psychological Association.

Smith, R. E., & Hunt, R. R. (2000). The influence of distinctive processing on retrieval-induced forgetting. *Memory & Cognition*, *28* (4), 503-508. DOI: 10.3758/BF03201240

Spitzer, H. F. (1939). Studies in retention. *The Journal of Educational Psychology, 30,* 641-656. DOI: 10.1037/h0063404

Spitzer, B. J., & Bäuml, K.-H. T. (2009). Retrieval-induced forgetting in a category recognition task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 286-291. DOI: 10.1037/a0014363

Storm, B. C., Bjork, E. L., & Bjork, R. A. (2008). Accelerated relearning after retrieval-induced forgetting: The benefit of being forgotten. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 230-236. DOI: 10.1037/0278-7393.34.1.230

Storm, B. C. & Levy, B. J. (2012). A progress report on the inhibitory account of retrieval-induced forgetting. *Memory & Cognition, 40*, 827-843. DOI: 10.3758/s13421-012-0211-7

Storm, B. C. & Nestojko, J. F. (2010). Successful inhibition, unsuccessful retrieval: Manipulating time and success during retrieval practice. *Memory, 18* (2), 99-114. DOI: 10.1080/09658210903107853

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences, 110* (16), 6313-6317. DOI: 10.1073/pnas.1221764110

Szpunar, K. K., McDermott, K. B., & Roediger, H. L. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory and Cognition, 34*, 1392-1399. DOI: 10.1037/a0013082

Thorndike, E. L. & Lorge, I. (1944). *The teacher's word book of 30,000 words*. New York: Teachers College, Columbia University.

Toppino, T. C. & Cohen, M. S. (2009). The testing effect and the retention interval: Questions and answers. *Experimental Psychology, 56*, 252-257. DOI: 10.1027/1618-3169.56.4.252

Tulving, E. & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352-373. DOI: 10.1037/h0020071

Tulving, E. & Watkins, M. J. (1974). On negative transfer: Effects of testing one list on recall of another. *Journal of Verbal Learning and Verbal Behavior, 13*, 181-193. DOI: 10.1016/S0022-5371(74)80043-5

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*, 498-505. DOI: 10.3758/BF03192720

Verde, M. F. (2012). Retrieval-induced forgetting and inhibition: A critical review. *Psychology of Learning and Motivation, 56*, 47-80. DOI: 10.1016/B978-0-12-394393-4.00002-9

Verde, M. F. & Perfect, T. J. (2011). Retrieval-induced forgetting is absent under time pressure. *Psychonomic Bulletin & Review, 18*, 1166-1171. DOI: 10.3758/s13423-011-0143-4

Weinstein, Y., McDermott, K. B., & Szpunar, K. K. (2011). Testing protects against proactive interference in face-name learning. *Psychonomic Bulletin & Review, 18*, 518-523. DOI: 10.3758/s13423-011-0085-x

Williams, C. C. & Zacks, R. T. (2001). Is retrieval-induced forgetting an inhibitory process? *American Journal of Psychology, 114*, 329-354. DOI: 10.2307/1423685

Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2011). The interim test effect: Testing prior material can facilitate the learning of new material. *Psychonomic Bulletin & Review, 18*, 1140-1147. DOI: 10.3758/s13423-011-0140-7

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Langauge, 46*, 441-517. DOI: 10.1006/jmla.2002.2864