SMITH, WYATT G., M.A. When to Move on to New Learning: Meta-cognition's Role on Updating with Incorporation. (2019)
Directed by Dr. Peter F. Delaney. 62pp.

Updating with incorporation occurs when people learn previously unstudied (*newer)* information with the intention of adding it to a body of previously studied *(older)* information. Testing information usually leads to better recall later, even when exposure time is matched equally with the restudy group. However, in a specific paradigm which has participants review older information immediately before learning newer information, testing produces costs - poorer recall of the newer items - which outweigh the benefits - better recall of the older items.

I propose that in this paradigm, testing the older information just before learning the newer information influences meta-cognitions and can encourage rehearsing older information at the cost of studying newer information. While my manipulations were successful, I did not find that increasing confidence about older learned information by manipulating test difficulty (Experiment 1 and 3) or word fluency (Experiment 2) diminished the costs and benefits. The costs and benefits are influenced by more than just participants' confidence. Post hoc evidence suggests participants only need moderate confidence while restudying older information to move on to studying newer information. Failure to replicate the costs and benefits with unassociated items hints at the importance of word association in producing this pattern of results.

WHEN TO MOVE ON TO NEW LEARNING: META-COGNITION'S

ROLE ON UPDATING WITH INCORPORATION

by

Wyatt G. Smith

A Thesis Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Greensboro
2019

Approved by

____
Committee Chair

To Ben, Margaret, and Kailey

APPROVAL PAGE

This thesis written by WYATT G. SMITH has been approved by the following committee of the Faulty of The Graduate School at the University of North Carolina at Greensboro.

Committee Chair ____________________________________
Peter Delaney

Committee Members ____________________________________
Dayna Touron

____
Chris Wahlheim

____
Date of Acceptance by Committee

____
Date of Final Oral Examination

ACKNOWLEDGMENTS

I would like to show my appreciation to Dr. Peter F. Delaney for advising me in my scientific research for the better part of a decade. I would also like to thank my other committee members: Dr. Dayna Touron and Dr. Chris Wahlheim as well as Dr. Jeni Pathman, for their time and effort.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

Imagine that a chemistry teacher asks students, “Do you remember when we talked about the properties of atoms? Now we are going to learn about the structure of molecules.” The teacher is purposefully referencing previously studied *(older)* information just before teaching previously unstudied *(newer)* information to help the students incorporate the pieces together. *Updating with incorporation* occurs when people learn newer information with the intention of adding it to a body of older information. Should the teacher actually take the opportunity to evaluate the knowledge of the students before moving on to the new learning? Do feelings about how well the older information has been learned affect how well the students can learn the newer information?

The purpose of this paper is to explore whether such feelings about older information impact the learning of newer information during updating with incorporation. The concept of the paper was inspired by work on updating with incorporation that shows testing can lead to poorer recall than restudying (Bettencourt & Delaney, 2017; Davis & Chan, 2015; Finn & Roediger, 2013). In this paper I intend to make salient how bizarre of a result this is by reviewing the broader literature on updating. Then I will report on the mechanisms that have been proposed to explain the costs and benefits in updating with incorporation. Finally, I will outline three ways that the currently-dominant hypothesis can be interpreted and propose methods to distinguish between them.

Reviewing and Updating

Reviewing older information can strengthen it. Since the 19th century laboratory research has found that people have a tendency to forget information if it is not reviewed (Ebbinghaus, 1885/1964). There are multiple ways to review information, however. After an initial study session, participants who review with a test typically have higher final recall than participants who review by restudying (Carpenter, 2012, Dunlosky, Rawson, Marsh, Nelson, & Willingham, 2013; Karpicke, Lehman, & Aue, 2014; Roediger & Karpicke, 2006; Roediger, Putnam, & Smith, 2011; Rowland 2014). Interpolating testing into study sessions has been shown to promote learning above that of a control study group (Chan, 2009, 2010; Chan, McDermott, & Roediger, 2006; Cranney, McKinnon, Morris, & Watts, 2009; Kornell, Hays, & Bjork, 2009; Hays, Kornell, & Bjork, 2013; Little, Bjork, Bjork, & Angello, 2012; Little & Bjork, 2016; Little, Storm, & Bjork, 2011; Richland, Kornell, & Kao, 2009; Rowland & Delosh, 2014). Adding some testing into a studying routine improves not only the retention of the tested items that are reviewed, but also the untested items. Thus, reviewing older information can act to preserve it with testing typically being more effective than studying.

There are occasions where the older information that we learned is no longer valid, and reviewing it in the future would be counterproductive. In order to have the most accurate and up-to-date memory, we must substitute newer information for older information *updating with replacement*. This can be achieved with directed forgetting

(Bjork, 1972, 1989, 2015; Bjork & Bjork 1996; MacLeod, 1998). In directed forgetting, we study an item but later learn it is inaccurate and should be forgotten. For example, participants in item-method directed forgetting experiments view a list of words. Half of the participants are instructed to remember them all (remember condition) the other half of the participants are instructed to remember some and forget others (forget condition). After the study phase ends, participants in both conditions are asked to recall all items. Recall of the items participants were instructed to remember is higher in the forget condition than in the remember condition, demonstrating the *benefits of directed forgetting*. However, recall of the items participants were instructed to forget is lower than the recall for matched items in the remember condition, demonstrating the costs of directed forgetting (Bjork & Woodward, 1973; Foster & Sahakyan, 2012; Lee, 2013; Sahakyan, Delaney, Foster & Abushanab, 2013; Sahakyan & Foster, 2009). By reviewing older information, we can effectively forget those pieces that are no longer up to date in order to reduce interference.

In contrast, updating with incorporation occurs when we learn that our information is not wrong, but incomplete. In updating with incorporation, newer information extends the older information, which is still preserved. Similar to how our chemistry teacher was connecting the ideas of atoms and molecules together under one topic, your representation of a person is updated as additional relevant information is added. When you first meet someone, your initial impression might only consist of a few pieces of information like her face and name. After several meetings, you might also learn her profession, hobbies, hometown, and so on. As each piece of information is added, we do not wish to forget any of the older information; we want to recall and integrate all these characteristics.

Several recent studies on updating with incorporation have found costs to receiving a test reminder as opposed to a restudy reminder just prior to updating (Bettencourt & Delaney, 2017; Davis & Chan, 2015; Finn & Roediger, 2013). This was surprising, since as I noted initially, tests usually help learning relative to restudy. The first of the studies was conducted by Finn and Roediger (2013), whose participants studied 20 face-name pairs. After a math distractor task, participants had an opportunity to review the face and name they had studied earlier. One third of the participants reviewed by having the face-name pair presented again (restudy condition). The remaining two thirds of participants were shown only the face and were asked to enter the associated name. A third of the participants saw the correct face and name after entering an answer (test with feedback condition). The final third of the participants were given no feedback (test without feedback condition). Immediately after reviewing each pair, participants learned a newer piece of information: the person's profession. They were instructed that the name, face, and profession were all to be associated with one another. Each of the 20 face name pairs were updated with a profession. Participants completed a math distractor task, and then were tested by having only the older face presented and had to enter the name and profession associated with that face. There was no difference in the pattern of results between the test with feedback and test without feedback conditions, but there was a difference between the test and restudy conditions. Similar to updating with replacement, the effect of tests can be expressed in terms of both costs and benefits

(Bettencourt & Delaney, 2017). The *benefit* observed in the test condition was the higher recall for the face in the test condition as compared to the restudy condition. The *cost* observed in the test condition was the lower recall for profession in the test condition as compared to the restudy condition. Strangely, the small benefits and large costs in updating with incorporation resulted in restudying being more effective for retention than testing (Finn & Roediger, 2013). Older information was best preserved in the test condition while newer information was best learned in the restudy condition. These three scenarios (preserving, replacing and incorporating) all share a similar goal: ensuring that the most accurate information is recollected later. It would seem sensible to think that prior research in testing (updating with preservation) and updating with replacement can be used to explain updating with incorporation. Based on the testing literature described above, one might expect testing should enhance the recall of not just the tested items but also the untested items. In the context of Finn and Roediger's experiments, that would mean participants who review with a test would not only recall more of the tested items (names), but also more of the untested items (professions). While participants in the test condition recalled more of the older items than the restudy condition, participants in the restudy condition recalled more professions than the test condition. This contradicts the prediction that testing benefits both the tested and untested items. Testing seemed to only benefit the older item. In sum, the extant literature on the testing effect does not adequately explain why costs and benefits occur in updating with incorporation.

Potential Mechanisms

In other updating tasks, a variety of mechanisms have been proposed to account for how updating occurs. A logical first step is to ask whether those same mechanisms can be used to explain the costs and benefits of updating with incorporation. Specifically, I will examine mechanisms derived from reconsolidation, reminding, directed forgetting, and selective rehearsal.

Reconsolidation occurs when previously consolidated memories are reactivated and must be consolidated again (see Hupbach, Gomez & Nadel, 2015). Consolidation is assumed to occur automatically after enough time has elapsed from the initial learning, and acts to protect well learned information from forgetting. However, if an item is reactivated, it becomes labile and requires reconsolidation for it to return to a consolidated state. In typical reconsolidation studies, participants memorize a sequence of items or actions on Day 1. On Day 2, some participants receive an explicit reminder of Day 1 just before learning a new sequence, *the reminder condition,* while others learn a new sequence without any explicit reminder, *the no reminder condition*. On Day 3, all participants are tested on both the sequence from Day 1 and the sequence from Day 2. When tested on the sequence from Day 1, the no reminder condition recalls more Day 1 items and has fewer intrusions from Day 2 than the reminder condition.

If we apply the reconsolidation hypothesis to updating with incorporation, participants who are not reminded of their earlier experience of learning should show less forgetting than those who are reminded. Successful retrieval on a test could be analogous to reminding. In the updating with incorporation paradigm, participants who recall the older item just before learning the newer item should reactivate the prior experience similar to participants in the reminder condition in a reconsolidation paradigm. Therefore, in the test condition, successfully recalling the older item should produce more interference than failing to retrieve it. Reconsolidation theory therefore predicts that successfully retrieving an older item should lead to poorer performance on a final test both for the older and newer item, because participants have reactivated the trace and subjected it to interference. This prediction contradicts results obtained by Finn and Roediger (2013) and Davis and Chan (2015): participants who successfully retrieved the older item were *more* likely to recall both the older and the newer item as compared those participants who did not successfully retrieve the older item. A prediction from this theory would be that successfully retrieving the older item would make updating more difficult, which does not seem to predict the pattern in results observed in updating with incorporation.

Finn and Roediger (2013) proposed an explanation based on the reminding literature (Jacoby & Wahlheim, 2013; Wahlheim & Jacoby, 2013). In typical reminding experiments, participants are tasked with learning multiple lists of related pairs (e.g. angel-saint in List 1 and angel-wings in List 2). Participants would report when the cue for an older pair in List 1 is repeated in List 2 but its response has changed. Compared to control items that are only studied once (i.e., no reminding is possible), items which participants detected change during List 1 and successfully recollected change at the final test showed facilitation – higher recall of both the older and newer item - while items which participants detected change during List 1 and failed to recollect change at the final test showed *competition* - lower recall of both the older and newer item.

If we apply the reminding literature to updating with incorporation, participants who successfully retrieve the older item just before learning the newer item should be more likely to successfully recall both the older and newer item on final recall. If retrieval is successful, the older information is recognized and the newer information can be properly updated. If retrieval is not successful, participants might create new associations between the initial words and not receive the benefits of updating. During the final test, competition between the two different learning contexts could be responsible for the low recall of the to-be-learned information (for a similar mechanism, see Jonker, Seli, & MacLeod, 2013). Even though the earlier pair is contained within the later trio, updating can only occur when participants recognize that they have seen the older items. Otherwise, one creates a new trace that competes with the original trace at the final test.

Finn and Roediger (2013) tested this theory by compiling data from Experiments 1 – 3 and found that when participants successfully recalled the name on the initial test, they did not show deficits in recall of the profession associated with that name during the final test. Finn and Roediger concluded that the test condition's poorer recall of the profession was due to competition that only occurred when participants did not recognize that they had seen the item earlier. This means that when participants remember the item from before, there does not seem to be any competition. Based on these analyses, Finn and Roediger proposed that reviewing a studied item and reviewing a successfully recalled tested item function similarly in whether the item can be updated with newer information because only when the item is retrieved can the item be updated. One prediction about updating with incorporation based on the work of Jacoby and Wahlheim might be that *successful retrieval allows for updating*.

Turning now to a selective rehearsal mechanism, Finn and Roediger (2013) had considered an alternative account whereby people in the test condition are more likely to selectively rehearse the older items at the expense of the newer items. This mechanism can produce both the costs and benefits, because study time is allocated to the older item and taken from the newer item. Participants in the test condition could have low recall of the newer information because they focused either on the feedback itself or on rehearsing the name as a consequence of the feedback. They dismissed this idea because participants had lower recall of the newer item than participants in the restudy condition in *both* the test with feedback and *without* feedback conditions. They concluded the costs and benefits of testing in updating with incorporation was not explained by participant's focus on the older target because it was not re-presented in the no feedback condition.

However, the selective rehearsal account was revived by Davis and Chan (2015) in the form of the *borrowed time hypothesis* which states that participants could still keep trying to retrieve the older information even without feedback. That in turn would stop them from studying the newer information. This explains why participants in both the test with and without feedback conditions do not learn the newer word: it is because they are so focused on the older word. The borrowed time hypothesis explains the costs and benefits of updating with incorporation in the same way that selective rehearsal accounts have been used to explain the costs and benefits of updating with *replacement* (Basden &

Basden, 1996; Basden, Basden & Garbano, 1993; Bjork, 1970, 1972; Hourihan & Taylor, 2006; Woodward & Bjork 1971). Participants can intentionally forget an item because they are busy rehearsing the to-be-remembered item. The to-be-forgotten item is inhibited later as a result of this selective rehearsal (Fawcett & Taylor, 2008, 2010; Hourihan & Taylor, 2006; Taylor, 2005; Zacks, Radvansky, & Hasher, 1996). Unsuccessful trials in the test condition of updating with incorporation could be analogous to trials in the forget condition of updating with replacement. Participants in both of these situations may engage in selective rehearsal rather than distributing their learning more equally, which explains why they incur the costs (lower recall of the ignored item), and gain the benefits (higher recall of the selected item).

Selective rehearsal can explain the costs and benefits observed in updating with incorporation. From the research on updating, participants might only engage in selective rehearsal when they are explicitly instructed or when they perform poorly on an explicit memory test. However, if the decision to engage in selective rehearsal is primarily dependent on the evaluations made by participants, an unanswered question is whether or not we can observe the costs and benefits in instances when we manipulate the evaluations made by participants by means other than testing.

Remaining Issues: Role of Meta Cognitions

Unlike updating with replacement (where participants are clearly told which item to ignore), in updating with incorporation participants are left to allot their learning however they decide. Why do participants in the test condition favor reviewing the older item more than the newer item when they are instructed to learn both?

Davis and Chan (2013) argued that borrowing time happens because testing an item "increases its perceived importance, its perceived probability of being tested again, or because being tested on an item highlights its difficulty" (Davis & Chan, 2013, p. 3). The reason why participants rehearse the older item in the test condition is because that item is more valuable as a result of being tested. Consequently, the test condition encourages borrowing. Later, Davis and Chan stated that if participants perform well on the test then there is no need to continue reviewing the older item because it is already learned. This means that when the older item is successfully retrieved, the newer item will be successfully incorporated. Similar to the conclusions drawn by Finn and Roediger (2013) and the reminding literature, successful retrieval allows for updating. However, a unique prediction added by the borrowed time hypothesis is that testing impairs new learning.

Perhaps it is not the test itself *per se* which encourages participants in the test condition to rehearse the tested item, but judgments that result from using a test as the method for review. Restudying produces higher confidence in retention than testing (Dunlosky & Metcalfe, 2009; Koriat, Lichtenstein, & Fischhoff, 1980; Roediger & Karpicke, 2006) which can explain why participants in the test condition are more likely to keep rehearsing the old item. However, participants choose not to rehearse information they believe they already know (Dunlosky & Hertzog, 1998; Metcalfe, 2002; Metcalfe & Kornell, 2003) which might explain why Finn and Roediger (2013) found that participants who do well on the test have performance similar to the restudy condition. Participants who are successful on the test and those in the restudy condition have the same kind of meta-cognitive judgments about the information they have reviewed. The method of review and test success might not have a direct effect on when participants engage in selective rehearsal. Instead meta-cognitions about the older item just before the newer item is shown could solely determine whether a participant will selectively rehearse the older item at the expense of the newer item.

Finally, an account not yet proposed is a selective rehearsal account that predicts that when participants feel like they do not know the older item at review, they are more likely to rehearse the older item and ignore the newer item. This account predicts that in order for participants to move on to studying the newer item, sufficient confidence in the retention of the older item is required. Participants will choose to continue rehearsing an older item that is not well retained before they are sufficiently confident to move on to learning the newer item. Only participants with high enough confidence in their retention of the older item by the end of feedback will move on to studying the newer item. I specifically predict that *meta-cognitions influence selective rehearsal* even without using a test.

Davis and Chan (2013) suggest that meta-cognitions about the test explain why the participants selectively rehearse the older item but I am unsure if that means any intervention where the participants believes the older item is more important will result in selective rehearsal. If this effect is the result of meta-cognitions and not testing, then why refer to the effect as *test* impaired new learning? Why not call it meta-cognition impaired new learning? Meta-cognitions could mediate the effects of both review method and test success. Restudying as a method of review often increases the confidence in retention of items while testing produces realistic confidence judgments about the retention of items (Dunlosky & Metcalfe, 2009; Koriat et al., 1980; Roediger & Karpicke, 2006). Review method might not directly impact borrowing but rather affect it indirectly because testing is more likely to result in low confidence as compared to restudying. Test success can also be potentially mediated by meta-cognitions. Participants can use their performance on a test as a diagnostic of their own memory. This can subsequently influence what items they select to study. Is there something unique that happens only during a test that produces the costs and benefits or can manipulations that alter meta-cognitive confidence produce similar results without a test?

The purpose of these studies is to test three different accounts about updating with incorporation. First, perhaps the costs and benefits occur because *successful retrieval allows for updating*. This account predicts that on trials where the older item is successfully retrieved by a participant in the test condition, the newer item is more likely to be incorporated. Second, perhaps the costs and benefits occur because *testing impairs new learning*. This account predicts that meta-cognitions specifically about a test are what encourage selective rehearsal. Third, perhaps the costs and benefits could occur because *meta-cognitions influence selective rehearsal*. This account predicts that when a participant does not feel confident about remembering the older item, they will selectively rehearse the older item. Because review method and test success can influence meta-cognitions, distinguishing these predictions requires the variation in meta-cognitions not to be the result of review method or test success.

CHAPTER II

OUTLINE OF PROCEDURES

Experiment 1

The aim of this first experiment was twofold: first, to replicate the costs and benefits from earlier studies; and second, to determine whether the costs and benefits are related to the difficulty of the test. A restudy condition was compared to two test conditions. The hard test condition was identical to the earlier test with feedback condition (Finn & Roediger, 2013), while the easy test condition was nearly identical but made slightly easier by providing the first and last letter of the target item. This manipulation would increase retrieval success which would also change evaluations made by the participant.

According to the successful retrieval allows for updating account, if the easy test does result in more test success than the hard test, participants should show relatively *less* costs and benefits than if they were taking a harder test but relatively *more* costs and benefits than if they were restudying.

The testing impairs new learning account predicts that participants in both the harder and easier test condition should be more likely to incur the cost and benefits than the restudy condition. However, because both conditions involve a test then there should be no difference in how much participants in the harder test condition and participants in the easier test condition incur the costs and benefits. The meta-cognitions influence selective rehearsal account has similar predictions to the successful retrieval allows for updating account. Participants in the easier test condition should have higher confidence in their retention of the older item than participants in the harder test condition but lower confidence than participants in the restudy condition. Subsequently participants in the easier condition should experience relatively smaller costs and benefits than the harder condition, but relatively more costs and benefits than the restudy condition.

Both the successful retrieval allows for updating account and the meta-cognitions influences selective rehearsal account predict an effect of the costs and benefits being proportional to the difficulty of the test, while the testing impairs new learning prediction hypothesizes that the costs and benefits should be equal between the two test conditions.

Method

Power Analyses. Both Finn and Roediger (2013) and Davis and Chan (2015) reported large effect sizes when comparing a test condition, which is identical to the relatively harder test condition, and a restudy condition. We hypothesize that the relatively easier test condition should fall between relatively harder test condition and the restudy only condition. For that reason we assume only a medium effect size between the relatively easier test condition and the relatively harder test condition as well as only a medium effect size between the relatively easier test condition and the restudy only condition. In order to detect a medium effect size difference using an independent sample *t*-test with an alpha of .05 and a beta of .80, I will need 64 participants in each condition (Cohen 1992) resulting in a total sample size of 192.

Participants. A total of 200 University of North Carolina at Greensboro students participated in this experiment for course credit. For not complying with instructions (clicking the mouse when asked not to, failing to pay attention to the screen, or not attempting the math distractor), 45 participants were excluded. Therefore, 155 participants are used in the analysis below with 47 in the study only condition (*restudy condition*), 55 in the relatively harder test condition (*hard test condition*) and 53 in the relatively easier test condition (*easy test condition*).

Materials. Participants studied 20 sets of English words arranged as trios (Dog-Spoon-Rose). All the words were randomly drawn from the Toronto Word Pool (Friendly et al., 1982). Concreteness and imagery of words ranged from medium (3.5/7) to high (7/7). Word frequency was not restricted. Words were no longer than two syllables and no longer than ten letters. Each trio was reviewed by the research team to make sure no obviously highly associated words are grouped together. For each participant the order of the word trios was random but within each trio the order was fixed. For the easy test condition, the first and last letter of the target word was shown next to the cue word. For the hard test condition, only the cue word was presented next to a blank space.

Procedure. Participants were randomly assigned to one of the three conditions: the *restudy condition, hard test condition,* and *easy test condition*. All participants were informed at the beginning that the words being studied would be tested and that words presented together were to be learned together. All participants were presented 20 word pairs, the initial two words in the trio, at a rate of 5s each and afterwards completed 60s of math problems.

Following the math distractor, participants in the *restudy condition* were presented each of the 20 pairs studied earlier again. After 5s of reviewing the pair, the third item was added to the other two for 5s. Participants in the *hard test condition* were presented only one word from the earlier studied pair and were prompted to enter the word they associated with it earlier. There was no limit on how long participants could take to enter their answer. Once entered, the correct response appeared for 2s and then was immediately followed by the third new item to be associated with the earlier two for 5s. Participants in the *easy test condition* were presented with the earlier studied pair, one completely filled in and one with only the first and last letter filled in. Participants in the easy test condition were given unlimited time to enter in the missing word, similar to the hard test condition.

Once the list was reviewed and the new word incorporated, all participants completed another 60s of math. The final test was the same for all conditions: the first word from every trio was shown in a random order -- the same word presented for the test groups during the review phase -- and participants were asked to enter both words that were associated in any order. Participants had unlimited time to respond.

Results

First, the total proportion correct on the initial test was compared. The easy test condition (M = .31, SE = .025) correctly recalled more items on the initial test that the hard test condition (M = .18, SE = .027), t (106) = 3.41, p = .001.

A series of one directional t-tests were used to compare differences in recall between the hard test condition, easy test condition, and restudy only condition.

Second, the total proportion correct of all items recalled was compared. The hard test condition (*M* = .25, *SE* = .025) and the easy test condition (*M* = .22, *SE* = .026) were not significantly different, *t* (106) = .96, *p* = .34. The hard test condition and the restudy condition (*M* = .20, *SE* = .026) were also not significantly different, *t* (100) = .53, *p* = .57, and neither were the easy test condition and the restudy condition, *t* (98) = .44, *p* = .66.

Third, the proportion correct of older items was compared to determine the benefits of testing. Figure 1 shows the proportion correct on the final test and associated standard errors as a function of condition (hard test, easy test and restudy) and item type (older or newer). There were no significant differences: for the hard test condition and the easy test condition, *t* (106) = .92, *p* = .36; the hard test condition and the restudy condition, *t* (100) = .24, *p* = .81; and the easy test condition and the restudy condition, *t* (98) = .69, *p* = .49.

Finally, the recall for the newer items was compared to determine the costs of testing. There were no significant differences: for the hard test condition and the easy test condition, *t* (106) = .87, *p* = .39; the hard test condition and the restudy condition, *t* (100) = .79, *p* = .43; and the easy test condition and the restudy condition, *t* (98) = .06, *p* = .95.

Discussion

The successful retrieval allows for updating account and the meta-cognitions influence selective rehearsal account both predict that when participants are successful on a test, they do not show the costs and benefits. This prediction was tested indirectly by conditionalizing recall on success (Davis & Chan, 2015; Finn & Roediger, 2013). In Experiment 1, I moderated the difficulty of the initial test which resulted in the easy test condition recalling more items on the initial test than the hard test condition. Success rates were directly manipulated by which condition the participant was assigned.

There were no significant differences in correct proportion recalled between the hard test, easier test or restudy condition. These results were inconsistent with prior experiments (Davis & Chan, 2015; Finn & Roediger, 2013) and I failed to observe the costs and benefits. The major difference between this experiment and prior experiments is that the items in this experiment had low association with each other. Highly associated items might be a pre-requisite to observe the costs and benefits in updating with incorporation.

Meta-cognitions could still be influencing whether a participant will engage in selective rehearsal and subsequently incur the costs and benefits of updating with incorporation. When items are unassociated with each other, participants have both lower confidence in their recall and lower recall performance of those items (Nelson, Bennet, & Leibert, 1997). If all the participants, even those restudying, had low confidence as a result of the low association of the items, then participants would rarely move on to studying the new item. Initial test performance for our easy test condition looked like the test condition from prior experiments (Davis & Chan, 2015; Finn & Roediger, 2013): around 30%. If our easy test condition and our restudy condition are having meta-cognitions similar to those in the test condition from prior experiments due to the low association between the items, then it is possible that everyone is deciding not to move onto learning the new item and that is why I failed to observe the costs and benefits.

In Experiment 2, I manipulated participants' meta-cognitive judgments without using a test. The costs and benefits could still be observed in materials that have low association given that sometimes participants can be overconfident due to the way materials are presented.

Experiment 2

The aim of Experiment 2 was to determine whether I could obtain similar costs and benefits as have been observed in previous updating with incorporation experiments without using a test. If the costs and benefits result from how participants decide to allot their study time, testing might not be the only way to produce the costs and benefits. I can influence what items they choose to study without using a test if I employ another method to manipulate meta-cognitions.

Judgments of learning (JOLs) are made while studying material and reflect how well a person feels the material has been learned (for a review, see Dunlosky & Metcalfe, 2009). The meta-cognitions influence selective rehearsal account predicts that lower JOLs of the older item during the restudy phase will inform the participant to hold off studying the newer information because the older information is not well enough retained. Testing as compared to a restudy control group results in lower JOLs (Roediger and Karpicke, 2006), meaning the participant feels the material is not well learned. However, items that are tested are actually more likely to be recalled on the final test, despite participants' lower JOLs. Because the test condition in Experiment 1 uses a test at review, participants are more likely to both make and have low JOLs of the material they are reviewing than participants who are in the restudy condition. These meta-cognitions, facilitated by testing, guide participants to dedicate more resources to the material they believe it is most advantageous to study at the expense of the nearby material.

To affect the JOLs of participants, an altered restudy condition manipulated the fluency of the words. Words that are presented in hard to read font size produce lower JOLs than their clear to read controls in a within-subjects design (Rhodes & Castel,2008). It should be noted that while the JOLs are affected by word fluency, recall remains largely unaffected.

According to both the successful retrieval allows for updating account and the testing impairs new learning account, if participants review by restudying then they should not incur the costs and benefits. According to the meta-cognitions influence selective rehearsal account, participants could still incur the costs and benefits if they feel like they do not know the material, such as when the items are presented in small font. Participants will not feel like they know those and will still be rehearsing them while the new item is presented. Situations when the items are presented in smaller font should be analogous to situations in which the participant did not perform well on the test. Both the testing impairs new learning prediction and the successful retrieval allows for updating prediction hypothesize no difference between the high and low fluency items because both are reviewed by restudying. Because the fluency of the items can impact meta-cognitions, similar to retrieval success (or failure), the meta-cognitions influence selective rehearsal account hypothesizes a difference between restudied items with the lower fluency items being more likely to result in the costs and benefits.

Method

Participants. A total of 115 University of North Carolina at Greensboro undergraduates at least 18 years old participated in this study for course credit. For not complying with instructions (clicking the mouse when asked not to, failing to pay attention to the screen, or not attempting the math distractor), 11 participants were excluded. Therefore, 104 participants were included in the following analyses.

Materials. The selection and creation of materials was identical to the restudy only condition from Experiment 1 with only one difference: the items participants viewed could have a large font size or small font size. Participants studied 20 sets of English words arranged as trios (Dog-Spoon-Rose). All the words used were randomly drawn from the Toronto Word Pool (Friendly et al. 1982). Concreteness and imagery of words ranged from medium (3.5/7) to high (7/7). Word frequency was not restricted. Word length was no more than two syllables and no more than ten letters. Each trio was reviewed by the research team to make sure no obviously highly associated words are grouped together. For each participant the order of the word trios were random but within each trio the order was fixed.

Design and Procedure. This experiment was all within-subjects. This was done because Rhodes and Castel (2008) used a within-subjects design when manipulating participants' JOLs with the fluency of the words. The procedure for this experiment was similar to the procedure for the restudy condition from Experiment 1. Participants were informed at the beginning that the words being studied would be tested and that words presented together were to be learned together. Participants were presented 20 word pairs, the initial two words in the trio, at a rate of 5s each and afterwards completed 60s of math problems. These words appeared in a random order for each participant and half the word pairs appeared 18pt font and the other half in 48pt font. Words were counterbalanced between two conditions so that every word that was viewed in 18 pt font in one condition was viewed in 48 pt font in the other and vice versa. As each word pair was shown, participants were asked to respond to the question “How well do you think you will remember this word pair for a later memory test?” on a 7 point Likert scale with 1 labeled as not very well, 4 labeled as somewhat well and 7 labeled as very well.

Following the distractor, each of the 20 word pairs that were studied earlier were reviewed one at a time with word pairs appearing in a random order for each participant. After 5s of reviewing the pair, the third item was added to the other two for 5s. Word pairs were presented in the same font size they had been shown earlier and the newly added third word conformed to the font size that the other members in the trio were presented in. Once the list had been reviewed and the new word incorporated, participants completed another 60s of math.

The final test was the same as for Experiment 1: the first word that was presented from every trio was shown in a random order and participants were asked to enter both words that were associated in any order. Participants had unlimited time to respond. After completing this procedure with 20 sets of trio, participants repeated the procedure with a new set of 20 trios.

Results

A series of one-tailed *t*-tests similar to those conducted in Experiment 1 were conducted in Experiment 2. First, differences in JOLs reported when viewing large font items (M = 4.21, SE = 1.77) was compared to the JOLs reported when viewing small font items (M = 4.07, SE = 1.79). JOLs for large font items were significantly higher than the JOLs for small font items, t (103) = 2.78, p = .006, suggesting the manipulation was successful. Second, total proportion correct of items recalled was compared. Small font items (M = .37, SE = .016) were recalled significantly more than large font items (M = .33, SE = .014), t (103) = 30.78, p < .001.

The recall rates of the older items and newer items are plotted in Figure 2. To evaluate the benefits, the proportion correct of older items recalled was compared. Small font items were recalled significantly more than large font items, t (103) = 14.90, p <.001. Finally, to evaluate the costs the proportion correct of newer items recalled was compared. Small font items were recalled significantly more than large font items, t (103) = 20.74, p < .001. The correct recall rates of the older items and newer items are plotted below in Figure 2.

After all planned analyses were conducted, a post hoc analysis of recall probability conditionalized upon the JOLs reported by participants was conducted. A set of two tailed *t*-tests comparing successive values of JOL (e.g., 1 vs. 2, 5 vs. 6) were used to determine significance. Results are plotted in Figure 3 and the *t*-tests are shown in Table 1. Participants' performance for the older and newer items both increased as JOLs increase.

Discussion

In Experiment 2, I replicated some of the results from the Rhodes and Castel (2008) experiment. Participants did rate lower fluency words as being less likely to be remembered on a later memory test. This manipulation is important because it mimics confidence ratings for participants in the test condition in updating with incorporation. If these participants have similar JOLs as participants in the test condition, they might make similar decisions and also produce the costs and benefits. However, words with a lower fluency were recalled more often overall on the final test than items with a higher fluency. This result was not found in the original studies which used single words, not pairs, and did not allow participants to view the items on multiple occasions.

Meta-cognitions could still be guiding participants' decisions. When participants viewed a large font item they showed higher JOLs. Overconfident participants might need more time to study the older item than they allot themselves. When participants move on to learning the newer item too early they should incur the costs. We observe the costs because the low fluency older items are better recalled on the final test then the high fluency older items. When participants move on to learning the newer item they should also incur the benefits. Why were the high fluency newer items recalled less often than the low fluency newer items? There might be two possible reasons.

The first is that overconfident participants are borrowing, but across trials rather than within. If a participant sees a trio of items in large easy to read font they might feel so confident in recalling those items later that they begin to think about earlier items that they are less confident in, low fluency items. This would explain why participants did better on average for the low fluency items than the high fluency items. The second reason is that overconfident participants are laboring in vain. When participants view the low fluency items they might have more accurate JOLs and thus when they move on to learning the newer item it is because it is advantageous. When participants are viewing the high fluency items, they might always feel like moving on to the newer item even though they are not likely to learn the item and should really allocate more study time to the older item.

The post hoc analysis which conditionalized recall on the final test as a function of the JOLs participants reported did yield significant results. The question of this paper is when *should* people move on to new learning. This analysis has given insight into when people *do* move on to new learning. As reported JOLs increase, so does the probability of recalling the older item on the final test. There were significant differences in recall for JOLs at the bottom of the scale (2 and 3) the middle of the scale (3 and 4) and the top of the scale (6 and7). This shows participants were relatively accurate in their meta-cognitions despite the manipulation of fluency which produced better recall for items with lower JOLs. Even though participants did not make a JOL on the newer item, I used the previous JOL for the initial pair it will be incorporated with to predict how well the newer item would be recalled. There is a similar trend in that as JOLs increase, the probability of recalling the newer item increases, but the only significant difference for the newer items was at the middle of the Likert scale ("I feel somewhat confident"). Participants who are not confident at all may not even look at the newer item. As confidence increases, so does the likelihood of the participant choosing to study the newer item. The middle of the scale is the threshold where most participants will choose to study the new item. This might mean that when we are teaching our students newer material that is related to older material, they do not have to be highly confident in the older information in order to study the newer information; they only need to be somewhat confident.

Experiment 3

In Experiment 3, Experiment 1 was replicated with the following changes: it was made a within-subjects design and the final test was single item recognition. As noted before in the discussion for Experiment 1, the low association between the words might have made it too difficult for participants to do well on a cued recall test. Because recognition tests are easier than recall tests, I continued to use words with low association to avoid any ceiling effects.

The predictions from Experiment 1 are similar to this experiment: both the successful retrieval allows for updating account and the meta-cognitions influence selective rehearsal account predict that when a participant reviews an older item with an easier test, the newer item is more likely to be remembered as compared to when the item was reviewed with a harder test. However, the newer item added after an easier test should be remembered less often than the newer item added after restudying. The testing impairs new learning account would not predict any differences between newer items that were added after reviewing with an easier test and newer items that were added reviewing with a harder test.

Method

Participants. A total of 83 University of North Carolina at Greensboro undergraduates participated in this experiment for course credit. For not complying with instructions (clicking the mouse when asked not to, failing to pay attention to the screen, or not attempting the math distractor), 7 participants were excluded. Therefore, 76 participants were included in the analyses.

Materials. Selection and creation of words was similar to Experiment 1, except instead of studying 20 sets of trios participants studied 21. The reason for this is because the experiment is within-subjects and the number of items needed to be divisible by the number of conditions. All the words were randomly drawn from the Toronto Word Pool (Friendly et al., 1982). Concreteness and imagery of words ranged from medium (3.5/7) to high (7/7). Word frequency was not restricted. Words were no longer than two syllables and no longer than ten letters. Each trio was reviewed by the research team to make sure no obviously highly associated words are grouped together. For each participant the order of the word trios was random, but the order of the three words within the trio was the same for everyone.

Design and Procedure. This experiment used a within-subjects design. This was done to limit individual differences adding noise to the results (for successful within-subjects replication of the costs and benefits, see Bettencourt & Delaney, 2017). Similar to Experiments 1 and 2, all participants were informed at the beginning that the words being studied would be tested and that words presented together were to be learned together. All participants were presented 21 word pairs - the initial two words in the trio - at a rate of 5s each and afterwards completed 60s of math problems.

Following the distractor, participants reviewed all 21 pairs they had seen earlier: 7 of the items by restudying them, 7 of the items by taking an easier test, and 7 of the items by taking a harder test. If the pair was restudied, participants saw the cue and target word presented for 5s. Afterwards, the third item was added to the other two so all three could be seen together for 5s. If the pair was reviewed with a harder test, only the cue word from the earlier studied pair was shown and participants were prompted to enter the target word they associated with it earlier. There was no limit on how long participants could take to enter their answer. Once entered, the correct response appeared for 2s and then was immediately followed by the third new item to be associated with the earlier two for 5s. If the pair was reviewed with an easier test, the cue word was completely filled in and the target was partially filled in with the first and last letter. There was no limit on how long participants could take to enter their answer. Once entered, the correct response appeared for 2s and then was immediately followed by the third new item to be associated with the earlier two for 5s.

When the list had been reviewed and the new word incorporated, all participants completed another 60s of math. The final test was a single item old/new recognition test. In a random order the older target, the newer target and a matching number of foils were presented and participants were asked to respond whether they saw the item during the experiment (old) or whether they had not seen that item during the experiment (new). Participants had unlimited time to respond.

Results

A series of one-tailed *t*-tests similar to those conducted in Experiment 1 were conducted in Experiment 3. Because this experiment was all within-subjects, correct rejection is not specific to one item type or review method. Participants correctly rejected items they had not reviewed at any time earlier 86.8% of the time.

First, the total proportion correctly recalled on the initial test was compared. The easy test items (M = .21, SE = .047) were recalled more often on the initial test that the hard test items (M = .14, SE = .039), t (75) = 3.99, p < .001.

Second, the total proportion of all items correctly recognized as old was compared. The hard test items (M = .74, SE = .016) were correctly recognized more often than the restudy items (M = .67, SE = .019), t (75) = 3.72, p < .001. The easy test items (M = .72, SE = .015) were also correctly recognized more often than the restudy items, (75) = 2.66, p = .001. There was no significant difference between the hard test items and the easy test items, t (75) = 1.05, p = .298.

The recognition rates of the older and newer items are plotted below in Figure 4. Third, the proportion of older items correctly recognized as old was compared to determine the benefits of testing. The hard test items were correctly recognized more often than the restudy items, t (75) = 5.43, p < .001. The easy test items were correctly recognized more often than both the restudy items, t (75) = 7.32, p < .001, and the hard test items, t (75) = 2.00, p = .049.

Finally, the proportion of newer items correctly recognized was compared to determine the costs of testing. The hard test items were not correctly recognized significantly more than the restudy items, t (75) = 1.07, p = .29, and the restudy items were not correctly recognized significantly more than the easy test items, t (75) = 1.91, p = .061. However, the hard test items were correctly recognized more often than the easy test items, t (75) = 2.78, p = .007.

Discussion

I failed to find costs or benefits when comparing the hard test items and the restudy items. It seems that testing benefited the retention of the older items and did not impair learning of the newer items. Observing the benefits but not the costs between the restudy items and hard test items show how fragile this pattern is. In Experiment 1, I failed to observe the costs and benefits in a cued recall task when words had low association to each other. In Experiment 3, I failed to observe the costs and benefits even though I used a recognition task to make it easier for participants to perform well. It would seem the pre-existing associations between items plays a key role in observing this effect.

While cumulatively there was no significant difference between the number of hard test items correctly recognized and number of easy test items correctly recognized, there was a significant difference between the individual recognition rates of older and newer items. When reviewing with an easier test, participants correctly recognized the older item more often. When reviewing with a harder test, participants correctly recognized the newer item more often. It would appear there are costs and benefits of reviewing with either a harder test or easier test even when the words have low association.

The results of Experiment 3 contradict the predictions of all three accounts. The successful retrieval allows for updating account predicted the items participants got right to be updated. The easier test items had more correct responses on the interpolated test, thus they should not show the costs and benefits to the same degree as when participants are less successful on the interpolated test. The testing impairs new learning account would not have predicted any costs and benefits as both conditions are tests. The meta-cognitions influence selective rehearsal account predicted participants who reviewed with a harder test to have lower confidence in recalling the older item when all three items are presented and thus would ignore the newer item so they could selectively rehearse the older item. Vice versa, participants who reviewed with an easier test should have higher confidence in recalling the older item when all three items are presented and thus be more likely to move on to learning the newer item.

In my experiment, I found that the costs and benefits were not produced with unrelated materials between a test condition and a restudy condition using cued recall. However, I did find costs and benefits with unrelated materials between two test conditions using a recognition test (at least with the easy test). This observation might be due to the nature of reminding and the method of testing. Participants review either by restudying or by being tested. Even though the final recall test is single item recognition, participants who review with a cued recall test still should improve their performance for the final recognition test (Macleod, 1975). Unlike cued recall, recognition tests do not require full recollection and participants can rely only on familiarity. Participants had high familiarity for items reviewed with an easy test but low familiarity for items that were added to the easy test items which produced costs and benefits between the two test conditions.

What could cause this strange pattern where reviewing with an easier test results in relatively greater costs and benefits than reviewing with a harder test? As I pointed out earlier in Experiment 2, participants might be borrowing across trials because of the within-subjects design. If participants review an item they believe is easy, they might forego studying that item in favor of reviewing some more difficult item that they remember from before. This could explain the benefits, better recognition on the older item when reviewing with an easier test item, as well as the costs, worse recognition on the newer item when reviewing with an easier test.

CHAPTER III

DISCUSSION

The aim of this research paper was to explain the pattern of costs and benefits in updating with incorporation. After reviewing the relevant literature, there appeared to be three plausible accounts.

The first was the successful retrieval allows for participants to update account. When participants fail to recall the older information, they are less likely to remember the newer information on a later memory test. This is potentially because when participants fail to retrieve the previous context in which they learned the earlier pair, competition between the two learning contexts can cause interference. During the test, the cue might remind participants of the first time they saw the cue which included only the older target. Finn and Roediger (2013) and Davis and Chan (2015) ran conditional analyses to test this idea. They found that when participants do well on the test at review, they are more likely to remember the newer item on the final test. This is similar to reminding studies (Wahlheim & Jacoby, 2013) where participants who reported detecting a change when viewing a new list were and recollected the change at final recall more likely to remember both the older and newer items on the final test as compared to those participants who detected change but failed to recollect it on the final test. Because of this prior work, I thought successful retrieval should predict when participants will successfully recall the older and newer items on the final test.

However, the results of Experiment 1 and Experiment 3 contradict predictions from the successful retrieval allows for updating account. In both experiments, I altered the difficulty of the test during the review phase so that retrieval rates could be directly manipulated. This meant we could test whether increased rates of retrieval would produce better learning of the newer item without having to use a conditional analysis. In Experiment 1, despite participants in the easy test condition doing better on the interpolated test during review, the manipulation yielded no significant differences between conditions on the final test. In Experiment 3, the relative difficulty manipulation was again successful, but participants who reviewed with an easier test had lower recall rates on the final test of the newer item than when participants reviewed with a harder test. This means that when participants did well on a test at review because of a direct manipulation, they were less likely to learn the newer item. The results of these experiments seem to suggest that increasing successful retrieval does not enhance learning of the newer item.

Davis and Chan (2015) introduced the idea that something unique about taking a test causes both the costs and benefits. The null results we observed in Experiment 1 could support this account. Participants who took an easier test and those who took a harder test had similar patterns in recall. I also did not observe the costs and benefits in Experiment 2 which would also support the testing impairs new learning account as it suggests there is something special about testing that causes this effect. In Experiment 3, I observed benefits from reviewing with a test but failed to observe any costs when compared to a restudy condition. This would suggest that testing did not impair new learning. Overall, I failed to find strong support the testing impairs new learning account.

However, the only time I observed costs and benefits was between a relatively more difficult and a relatively easier test condition which would support the idea that there is something unique about testing that is necessary for this pattern to emerge.

Two of the ideas proposed by the borrowed time hypothesis (Davis & Chan, 2015): participants in the test condition are borrowing time and successful performance on a test reduces participants from borrowing, are difficult to reconcile together. The purported reason why participants will borrow time is because testing highlights importance, likelihood of being tested again, or difficulty. At the same time, they suggest that when participants perform well on the test, they will decide not to borrow because they feel it is already learned. How does successful retrieval of a tested item negate all the reasons why testing should impair new learning? Does successfully recalling an item on a test not highlight its importance? Did participants not think they would be tested on the item again because they got it right? There is not much evidence to conclude good performance on a test negates how the test impacts participants feelings of importance (Dirkx, Thoma, Kester, & Kirschner, 2014; Wilford, Chan, & Tuhn, 2014), likelihood of being tested (Weinstein, Gilmore, Szpunar, & McDermott, 2014), or difficulty (Kang, 2010). Independently, these mechanisms for explaining the costs and benefits put forth by the borrowed time hypothesis are logical, yet their relationship to one another is unclear.

The third account I wanted to test in this paper was the meta-cognitions influence selective rehearsal account. If participants do not feel confident about the older information, whether it was because of a test or any other reason, they are going to choose to selectively rehearse the older information until they feel confident enough to move on. In Experiments 1 and 3, the meta-cognitions influence selective rehearsal account predicted that if participants performed better on a test, they would be less likely to selectively rehearse the older item as compared to a harder test condition. I successfully altered the difficulty of the test, which directly manipulated the retrieval rates for the interpolated test at review. However, I failed to observe the expected direction of the costs and benefits in Experiment 1 and 3, which contradicts predictions made by the meta-cognitions influence selective rehearsal account.

In Experiment 2, the meta-cognitions influence selective rehearsal account would have predicted that when participants saw words in an easy to read font they would do well on the newer item and when participants saw words in a hard to read font they would do well on the older item. The hard to read font manipulation was intended to have participants generate lower JOLs (Rhodes & Castel, 2008) similar to a test condition but without the test. The manipulation was successful and participants who reviewed with hard to read words had lower JOLs than participants who reviewed with easy to read words. However, I observed that items presented in hard to read font were better remembered than items presented in easy to read font for both the older and newer items. Failure to find costs and benefits in Experiment 2 contradicts predictions made by the meta-cognitions influence selective rehearsal account.

Selective rehearsal guided by meta-cognitions might still explain the results I observed. No costs or benefits in Experiment 1 could have been due to the low association between the items. Because participants in all three conditions had low confidence as a result of the low word association, none of them might have moved on to learning the newer item. If no one moved on to learning the newer item, we would not expect to observe any costs and benefits due to selective rehearsal. The meta-cognitions influence selective rehearsal account might still predict the results of Experiment 1 if we assume everyone selectively rehearsed the older item because of the low association between all three items.

The only experiment in this study where we observed costs and benefits was in Experiment 3 when the final test was recognition. The easy test condition had the highest performance on the older items and the lowest performance on the newer items. This supports the idea that updating with incorporation can incur the cost and benefits, but why would the easier test condition have greater costs and benefits than a harder test? One possibility is that participants selectively rehearsed some items from another trial. If the participants believed the items that are present should not be studied, because they were well-learned due to their high rate of retrieval, then they might choose to selectively rehearse some other trio of items that they thought were more difficult. Participants who reviewed with an easier test did not use their time when all three items in the trio are present to study the newer item and thus had the lowest recall of the newer items. Participants borrowing across trials might also explain why I failed to observe the expected costs and benefits in Experiment 2. If participants believe the items that are present should not be studied, because they are so well learned due to the easy to read font, then they might choose to selectively rehearse some other trio of items that were presented in a more difficult to read font. The meta-cognitions influence selective rehearsal account could still explain the results of Experiment 2 and 3 if we assume participants might be borrowing across trials.

A future experiment where participants rehearse aloud could reveal if they are borrowing across trials. Rehearsing aloud experiments involve an aloud condition in which participants say aloud each word they see (Rundus & Atkinson, 1970). In addition, they might be instructed to study only the word present, rehearse it with earlier words (Delaney &Verkoeijen, 2009) or create a story with all the words presented (Delaney, Spirgel & Toppino, 2012). Asking participants to rehearse aloud ensures that they are rehearsing as instructed. A future experiment on updating would instruct participants to either study only the present words or encourage participants to rehearse the present words with earlier words. I predict that participants who are instructed to rehearse only the present words should not show the costs and benefits. I also predict that participants who are told to rehearse the present words with earlier words should show similar costs and benefits as I observed in Experiment 2 and 3. Rehearse aloud instructions in future experiments might uncover whether participants are borrowing across trials.

It would be worthwhile to investigate how material must be associated in order to observe the costs and benefits. Even though I replicated much of the design and procedure from Finn and Roediger (2013) and Davis and Chan (2015), I was not able to replicate the costs and benefits that they observed. This might have been due to low association between the words I used. In a future experiment, I would like to manipulate the association strength between all three items in the trio so that I can replicate earlier conditions where all three items are highly associated and the first two items are highly. Though I would want to know about all the other possible conditions (e.g., association strength between cue and Target 1 is low, association strength between cue and Target 2 is low, and association strength between Targets 1 and 2 is high). This experiment should clarify what materials are likely to produce the costs and benefits.

Delineating the meta-cognitions experienced by participants during recall could clarify the phenomenological mechanisms involved in updating. In a recent paper for instance, reminders were shown to potentially impair memory, facilitate memory or have no effect on memory based on the recollection by participants (Wahlheim, Smith, & Delaney, 2017). By asking questions about the nature of participant's recollection (e.g., scale of confidence, remember/know judgment, etc.) we might understand how participants update their memories. We might expect participants who *remember* an item has been updated to perform different than participants who *know* the item has been updated. By asking participants about their meta-cognitions, we can more precisely portray the mechanism associated with updating by apprising how participants make judgments about updated information.

Meta-cognitions can also measure retrieval more accurately than only success on a test. Most of us at some point have experienced the tip of the tongue phenomenon (TOT) where your meta-memory judgment (I know this item) does not match up with your memory performance (failure to retrieve the item) (for a review, see Schwartz & Cleary, 2016). Would you expect a student who failed the test by guessing and a student who failed a test but experienced a TOT state to review the item in the same way? While test success could be a proxy for how confident participants are in their retention of older information, in future studies I would like to take a more accurate probe by asking participants to report how likely they feel they will remember the item for a later memory test. Tests typically only reflect retrieval as being successful or unsuccessful. Meta-cognitive judgments however, can reflect the process of retrieval on a dimensional scale. The effort it takes to recall the item should inform the participant how likely they are to recall it in the future. Recording meta-cognitions in future experiments can reveal the state of retrieval better than using test success.

REFERENCES

Basden, B. H., & Basden, D. R. (1996). Directed forgetting: Further comparisons of the item and list methods. *Memory, 4*(6), 633-653. doi:10.1080/741941000

Basden, B. H., Basden, D. R., & Gargano, G. J. (1993). Directed forgetting in implicit and explicit memory tests: A comparison of methods. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(3), 603-616. doi:10.1037/0278-7393.19.3.603

Bettencourt, K., & Delaney, P. F. (2017). *A two factor account of test impaired new learning:Separating the costs and benefits.* Manuscript submitted for publication.

Bjork, R. A. (1970). Positive forgetting: the noninterference of items intentionally forgotten. *Journal of Verbal Learning and Verbal Behavior, 9*, 255-268.

Bjork, R. A. (1972). Theoretical implications of directed forgetting. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 217-235). Washington, D.C.: Winston.

Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 309-330). Hillsdale, NJ: Erlbaum.

Bjork, R. A. (2015). Forgetting as a friend of learning. In D. S. Lindsay, C. M. Kelley, A. P. Yonelinas, & H. L. Roediger III, (Eds.), *Remembering: Attributions, processes, and control in human memory: Essays in honor of Larry Jacoby* (pp. 15-28). New York, NY, US: Psychology Press.

Bjork, R. A., & Woodward, A. E. (1973). Directed forgetting of individual words in free recall. *Journal of Experimental Psychology,* 99(1), 22-27. doi:10.1037/h0034757

Bjork, E. L., & Bjork, R. A. (1996). Continuing influences of to-be-forgotten information. *Consciousness and Cognition*, *5*, 176-196.

Carpenter, S. K. (2012). Testing enhances the transfer of learning. *Current Directions in Psychological Science, 21*(5), 279-283. doi:10.1177/0963721412452728

Chan, J. 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*(2), 153-170. doi:10.1016/j.jml.2009.04.004

Chan, J. K. (2010). Long-term effects of testing on the recall of nontestedmaterials. *Memory, 18*(1), 49-57. doi:10.1080/09658210903405737

Chan, J. K., McDermott, K. B., & Roediger, H. I. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, 135(4), 553-571. doi:10.1037/0096-3445.135.4.553

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159. doi:10.1037/0033-2909.112.1.155

Cranney, J., Ahn, M., McKinnon, R., Morris, S., & Watts, K. (2009). The testing effect, collaborative learning, and retrieval-induced facilitation in a classroom setting. *European Journal of Cognitive Psychology, 21*(6), 919-940. doi:10.1080/09541440802413505

Davis, S. D., & Chan, J. K. (2015). Studying on borrowed time: How does testing impair new learning?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(6), 1741-1754. doi:10.1037/xlm0000126

Delaney, P. F., Spirgel, A. S., & Toppino, T. C. (2012). A deeper analysis of the spacing effect after 'deep' encoding. *Memory & Cognition, 40*(7), 1003-1015. doi:10.3758/s13421-012-0207-3

Delaney, P. F., & Verkoeijen, P. L. (2009). Rehearsal strategies can enlarge or diminish the spacing effect: Pure versus mixed lists and encoding strategy. *Journal Of Experimental Psychology: Learning, Memory, And Cognition, 35*(5), 1148-1161. doi:10.1037/a0016380

Dirkx, K. H., Thoma, G., Kester, L., & Kirschner, P. A. (2015). Answering questions after initial study guides attention during restudy. *Instructional Science, 43*(1), 59-71. doi:10.1007/s11251-014-9330-9

Dunlosky, J., & Hertzog, C. (1998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker, J. Dunlosky, A. C. Graesser, D. J. Hacker, J. Dunlosky, A. C. Graesser (Eds.) , *Metacognition in educational theory and practice* (pp. 249-275). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Thousand Oaks, CA, US: Sage Publications, Inc.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4-58. doi:10.1177/1529100612453266

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology.* Oxford, England: Dover.

Fawcett, J. M., & Taylor, T. L. (2008). Forgetting is effortful: Evidence from reaction time probes in an item-method directed forgetting task. *Memory & Cognition, 36*(6), 1168-1181. doi:10.3758/MC.36.6.1168

Fawcett, J. M., & Taylor, T. L. (2010). Directed forgetting shares mechanisms with attentional withdrawal but not with stop-signal inhibition. *Memory & Cognition, 38*(6), 797-808. doi:10.3758/MC.38.6.797

Finn, B., & Roediger, H. L. III (2013). Interfering effects of retrieval in learning new information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(6), 1665-1681. doi:10.1037/a0032377

Foster, N. L., & Sahakyan, L. (2012). Metacognition influences item-method directed forgetting. *Journal of Experimental Psychology: Learning, Memory, And Cognition*, *38*(5), 1309-1324. doi:10.1037/a0027868

Friendly, M., Franklin, P. E., Hoffman, D., & Rubin, D. C. (1982). The Toronto Word Pool: Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080 words. *Behavior Research Methods & Instrumentation, 14*(4), 375-399. doi:10.3758/BF03203275

Hays, M. J., Kornell, N., & Bjork, R. A. (2013). When and why a failed test potentiates the effectiveness of subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(1), 290-296. doi:10.1037/a0028468

Hourihan, K. L., & Taylor, T. L. (2006). Cease remembering: Control processes in directed forgetting. *Journal of Experimental Psychology: Human Perception and Performance, 32*(6), 1354-1365. doi:10.1037/0096-1523.32.6.1354

Hupbach, A., Gomez, R., & Nadel, L. (2015). Memory reconsolidation. In D. R. Addis, M. Barense & A. Duarte (Eds.) , *The Wiley handbook on the cognitive neuroscience of memory* (pp. 244-264). Wiley-Blackwell. doi:10.1002/9781118332634.ch12

Jacoby, L. L., & Wahlheim, C. N. (2013). On the importance of looking back: The role of recursive remindings in recency judgments and cued recall. *Memory & Cognition, 41*(5), 625-637. doi:10.3758/s13421-013-0298-5

Jonker, T. R., Seli, P., & MacLeod, C. M. (2015). Retrieval-induced forgetting and context. *Current Directions in Psychological Science, 24*(4), 273-278. doi:10.1177/0963721415573203

Kang, S. K. (2010). Enhancing visuospatial learning: The benefit of retrieval practice. *Memory & Cognition*, *38*(8), 1009-1017. doi:10.3758/MC.38.8.1009

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In B. H. Ross (Eds.) , *The psychology of learning and motivation, Vol.61* (pp. 237-284). San Diego, CA, US: Elsevier Academic Press.

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 187-194. doi:10.1037/0278-7393.31.2.187

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, *6(*2), 107-118. doi:10.1037/0278-7393.6.2.107

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 989-998. doi:10.1037/a0015729

Kornell, N., & Son, L. K. (2009). Learners' choices and beliefs about self-testing. *Memory*, *17*(5), 493-501. doi:10.1080/09658210902832915

Lee, Y. (2013). Costs and benefits in item-method directed forgetting: Differential effects of encoding and retrieval. *Journal of General Psychology*, *140*(3), 159-173. doi:10.1080/00221309.2012.750591

Little, J. L., & Bjork, E. L. (2016). Multiple-choice pretesting potentiates learning of related information. *Memory & Cognition*, doi:10.3758/s13421-016-0621-z

Little, J. L., Bjork, E. L., Bjork, R. A., & Angello, G. (2012). Multiple-choice tests exonerated, at least of some charges: Fostering test-induced learning and avoiding test-induced forgetting. *Psychological Science*, *23*(11), 1337-1344. doi:10.1177/0956797612443370

Little, J. L., Storm, B. C., & Bjork, E. L. (2011). The costs and benefits of testing text materials. *Memory, 19*(4), 346-359. doi:10.1080/09658211.2011.569725

MacLeod, C. M. (1975). Long-term recognition and recall following directed forgetting. *Journal Of Experimental Psychology: Human Learning And Memory, 1*(3), 271-279. doi:10.1037/0278-7393.1.3.271

MacLeod, C. M. (1998). Directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.) , *Intentional forgetting: Interdisciplinary approaches* (pp. 1-57). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning?. *Journal of Experimental Psychology: General*, *131*(3), 349-363. doi:10.1037/0096-3445.131.3.349

Metcalfe, J., & Kornell, N. (2003). The dynamics of learning and allocation of study time to a region of proximal learning. *Journal of Experimental Psychology: General, 132*(4), 530-542. doi:10.1037/0096-3445.132.4.530

Nelson, D. L., Bennett, D. J., & Leibert, T. W. (1997). One step is not enough: Making better useof association norms to predict cued recall. *Memory & Cognition, 25*(6), 785-796. doi:10.3758/BF03211322

Price, J., & Murray, R. G. (2012). The region of proximal learning heuristic and adult age differences in self-regulated learning. *Psychology and Aging, 27*(4), 1120-1129. doi:10.1037/a0029860

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General, 137*(4), 615-625. doi:10.1037/a0013684

Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning?.*Journal of Experimental Psychology: Applied, 15*(3), 243-257. doi:10.1037/a0016496

Roediger, H. L. III., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249-255. doi:10.1111/j.1467-9280.2006.01693.x

Roediger, H. L. III., Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In J. P. Mestre & B. H. Ross (Eds.) , *The psychology of learning and motivation: Cognition in education, Vol. 55* (pp. 1-36). San Diego, CA, US: Elsevier Academic Press. doi:10.1016/B978-0-12-387691-1.00001-6

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140*(6), 1432-1463. doi:10.1037/a0037559

Rowland, C. A., & DeLosh, E. L. (2014). Benefits of testing for nontested information: Retrieval-induced facilitation of episodically bound material. *Psychonomic Bulletin & Review, 21*(6), 1516-1523. doi:10.3758/s13423-014-0625-2

Rundus, D., & Atkinson, R. C. (1970). Rehearsal processes in free recall: A procedure for direct observation. *Journal of Verbal Learning & Verbal Behavior, 9*(1), 99-105. doi:10.1016/S0022-5371(70)80015-9

Sahakyan, L., & Delaney, P. F. (2010). Item-specific encoding produces an additional benefit of directed forgetting: Evidence from intrusion errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(5), 1346-1354. doi:10.1037/a0020121

Sahakyan, L., Delaney, P. F., Foster, N. L., & Abushanab, B. (2013). List-method directed forgetting in cognitive and clinical research: A theoretical and methodological review. In B. H. Ross (Ed.), *The psychology of learning and motivation* (pp. 131–190). Amsterdam, the Netherlands: Elsevier

Sahakyan, L., Delaney, P. F., & Kelley, C. M. (2004). Self-evaluation as a moderating factor of strategy change in directed forgetting benefits. *Psychonomic Bulletin & Review, 11*(1), 131-136. doi:10.3758/BF03206472

Sahakyan, L., & Foster, N. L. (2009). Intentional forgetting of actions: Comparison of list-method and item-method directed forgetting. *Journal of Memory and Language, 61*(1), 134-152. doi:10.1016/j.jml.2009.02.006

Schwartz, B. L., & Cleary, A. M. (2016). Tip-of-the-tongue states, déjà vu experiences, and other odd metacognitive experiences. In J. Dunlosky & S. K. Tauber (Eds.) , *The Oxford handbook of metamemory* (pp. 95-108). New York, NY, US: Oxford University Press.

Soderstrom, N. C., & Bjork, R. A. (2014). Testing facilitates the regulation of subsequent study time. *Journal of Memory and Language*, 7399-115. doi:10.1016/j.jml.2014.03.003

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 204-221. doi:10.1037/0278-7393.26.1.204

Taylor, T. L. (2005). Inhibition of return following instructions to remember and forget. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 58A*(4), 613-629. doi:10.1080/02724980443000115

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(4), 1024-1037. doi:10.1037/0278-7393.25.4.1024

Wahlheim, C. N., & Jacoby, L. L. (2013). Remembering change: The critical role of recursive remindings in proactive effects of memory. *Memory & Cognition, 41*(1), 1-15. doi:10.3758/s13421-012-0246-9

Wahlheim, C.N., Smith, W.G., & Delaney, P.F. (2017) *Reminders impair episodic memory updating: A memory-for-change perspective.* Manuscript submitted for publication.

Weinstein, Y., Gilmore, A. W., Szpunar, K. K., & McDermott, K. B. (2014). The role of test expectancy in the build-up of proactive interference in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(4), 1039-1048. doi:10.1037/a0036164

Wilford, M. M., Chan, J. K., & Tuhn, S. J. (2014). Retrieval enhances eyewitness suggestibility to misinformation in free and cued recall. *Journal of Experimental Psychology: Applied*, *20*(1), 81-93. doi:10.1037/xap0000001

Winne, P. H. (1997). Experimenting to bootstrap self-regulated learning. *Journal of Educational Psychology, 89*(3), 397-410. doi:10.1037/0022-0663.89.3.397

Woodward, A. E., & Bjork, R. A. (1971). Forgetting and remembering in free recall: Intentional and unintentional. *Journal of Experimental Psychology*, *89*, 109-116.

Zacks, R. T., Radvansky, G., & Hasher, L. (1996). Studies of directed forgetting in older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(1), 143-156. doi:10.1037/0278-7393.22.1.143

APPENDIX A

TABLES

Table 1. Probability of Recall on Final Test for Experiment 2 as a Function of Judgments of Learning and Item Type.

Older	**1 vs 2**	**2 vs 3**	**3 vs 4**	**4 vs 5**	**5 vs 6**	**6 vs 7**
df	138	177	199	196	186	154
p-value	.814	.016	<.001	.211	.067	.007
t-value	0.23	2.42	4.16	1.25	1.84	2.72
Newer	**1 vs 2**	**2 vs 3**	**3 vs 4**	**4 vs 5**	**5 vs 6**	**6 vs 7**
df	138	177	199	196	186	154
p-value	.083	.065	.002	.290	.387	0.426
t-value	1.75	1.87	3.12	1.06	0.87	0.80

APPENDIX B

FIGURES

Figure 1. Probability of Recall on Final Test for Experiment 1 as a Function of Condition and Item Type.

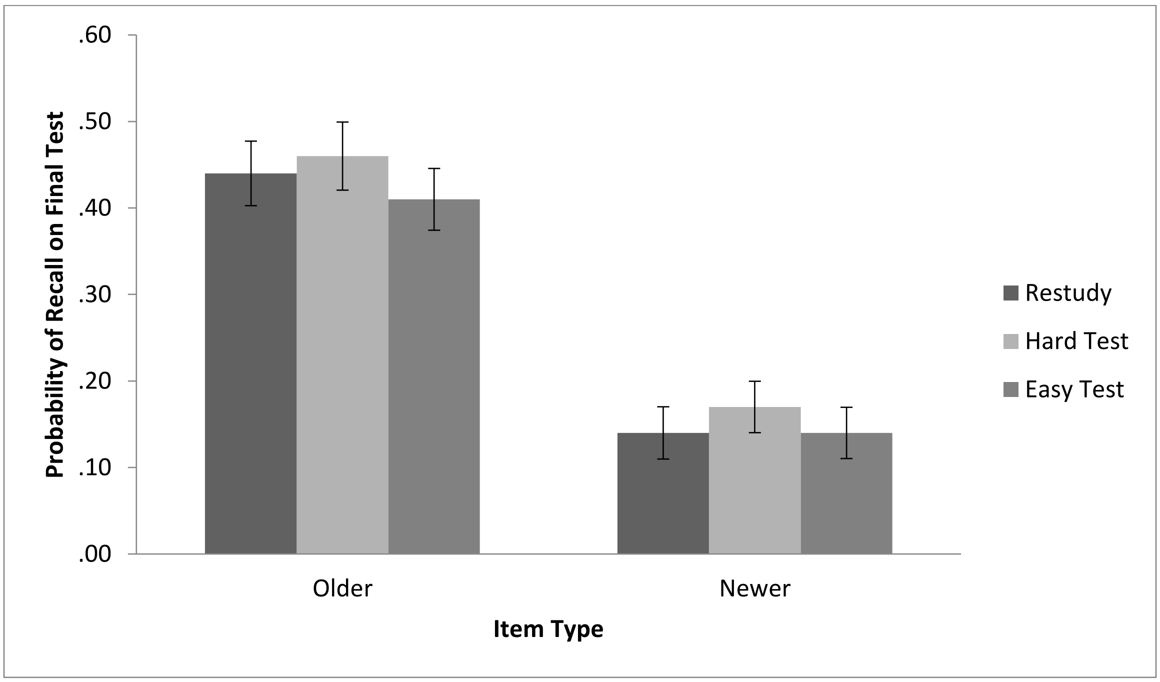

Figure 2. Probability of Recall as a Function of Word Fluency and Item Type

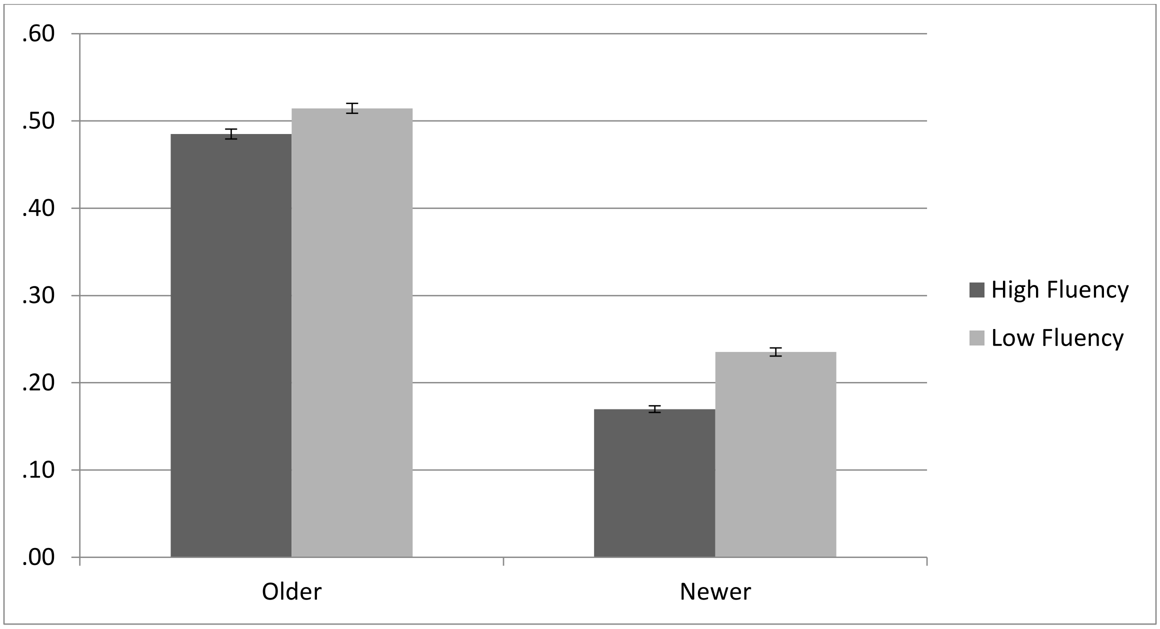

Figure 3. Probability of Recall on Final Test as a Function of Judgments of Learning and Item Type.

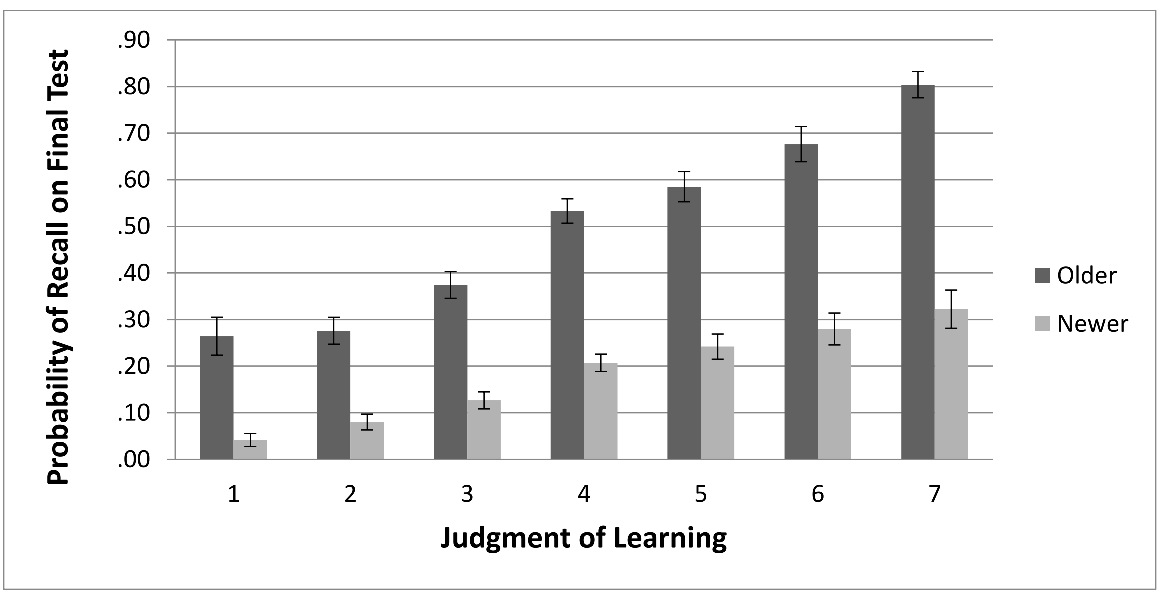

Figure 4. Probability of Correct Recognition on Final Test as a Function of Review Method and Item Type.