THE EFFECTS OF FRAMING PRACTICE TESTS AS RESTUDY ON FINAL RECALL

A Master's Thesis Presented to the Faculty of

The Department of Psychology

Villanova University

In Partial Fulfillment

Of the Requirements of the Degree of

Master of Science in

Psychology

By

Taylor M. Curley

March, 2016

Under the Direction of

Thomas C. Toppino, Ph.D. (*Chairperson*)

Irene P. Kan, Ph.D.

Michael F. Brown, Ph.D.

Table of Contents

List of Tables

List of Figures

Abstract

Previous research has shown that learners do not make study decisions that reflect an appreciation for the benefits of practice test feedback unless the directions explicitly state its restudy potential (Heslin et al., 2014). The current study examined whether final recall for word pairs is better when practice tests are framed as preceding restudy opportunities (emphasizing restudy potential) relative to when they are framed simply as practice tests followed by feedback. Individuals who received feedback had higher final recall accuracy than those who did not receive feedback at all. Final recall performance, however, only differed between the two framing conditions for items that were not retrieved during practice tests such that participants in the test framing condition performed better than those in the restudy framing condition. The replicability of this finding and why it occurs remain to be established by future research.

Introduction

Learning is a fundamental process in human cognition. Each day, individuals come across information, both important and trivial, that must be encoded and stored in memory so that it can be retrieved when needed. This process can occur incidentally, but individuals are often aware of their learning and allocate resources differentially based on factors related to the individual and the items that need to be remembered, including the importance of the information, their past learning experiences, and the likelihood that the information will be remembered later. This paper will begin by exploring the strategies that people use when learning, including the amount of time apportioned to successive study opportunities and how the items are studied during these additional trials. The efficacy of these strategies with regard to memory performance will also be considered, particularly for items that are viewed again as a practice test. The main focus of this paper, however, will be on the different ways that memory items are processed in the context of the situations that learners encounter when studying. More specifically, this research will examine memory performance for word-pair stimuli when the role of feedback in practice tests is presented in different “framing” contexts which influence the way in which feedback is interpreted.

Metacognition

Learners possess and can cultivate a sense of awareness of their own cognitive abilities, particularly when learning new information. Awareness of one’s knowledge is often referred to as metacognition, but can also be referred to as metamemory in the context of learning. Traditionally, metamemory has been interpreted using Nelson and Narens’ (1990) theoretical framework for metacognition in which the relationship

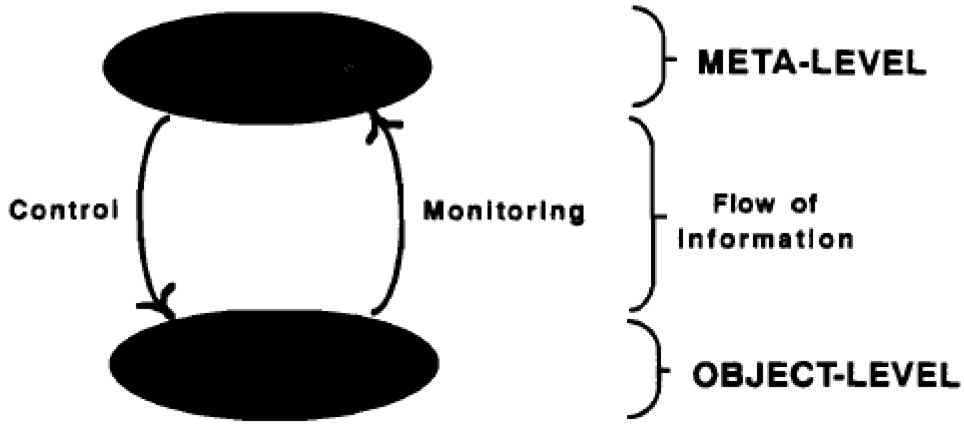

Figure 1. A simple representation of Nelson and Narens' (1990) proposed metacognitive framework. In this model, the object-level is modified by the meta-level using control while the meta-level is informed by the object-level through monitoring.

between awareness and cognitive ability is guided by three abstract principles. The first principle dictates that cognitive processes can be split into two interrelated levels, which the authors define as the "object-level" and the "meta-level". The second principle defines these two levels, where the object-level represents traditional cognitive processes (i.e. encoding processes, retrieval processes, etc.) while the meta-level represents knowledge about cognition (i.e. goals, strategies, and beliefs) as well as a simulation of the object-level which represents ongoing processing at the object level. Nelson and Narens' final principle states that the relationship between the two levels is dictated by two separate mechanisms, "control" and "monitoring", and that these two mechanisms are defined by the flow of information between the object- and meta-levels. Figure 1 illustrates the relationship between the two levels with regard to monitoring and control.

The framework postulates that monitoring and control are independent of one another and represent fundamentally different processes. Consider a situation in which participants are trying to learn word pairs for a final cued-recall test using a procedure in which each presentation of an item involves a practice test (i.e., only the cue word is presented and the participant tries to supply the corresponding target word) followed by feedback (i.e., presentation of the intact pair). The retrieval processes and encoding processes that the participant uses to try to learn the pairs constitute what is happening at the object level, and the meta-level becomes informed of this activity and its consequences via monitoring. Thus, for example, the meta-level keeps track of how well items are perceived to be learned, possibly by assessing the ease with which items are retrieved on practice test trials. Control refers to the meta-level acting on the object level in a way that either changes the state of the process represented by the object-level or changes the process itself, resulting in termination of an action, instantiation of an action, or the continuance of an action that is already in place. An example of control in word-pair learning would be using an assessment of item difficulty in conjunction with strategies or beliefs about memory (both found in the meta-level) to decide if and when the pair should be studied again. Nelson and Narens (1990) postulate that monitoring and control are autonomous, yet function together, which is supported by research showing that focusing on subjective experience (meta-level) makes monitoring more salient while focusing on data-centered information (object-level) makes control more salient (Koriat, Ma'ayan, & Nussinson, 2006).

Control and monitoring are measured using introspective and behavioral methods that are utilized as a function of the stage of learning a person is in. In the acquisition stage, individuals both prepare to learn through evaluating their learning goals and prepare a strategy based on judgments that they make about the information that needs to be remembered. At this stage, these judgements are prospective, or determinations of

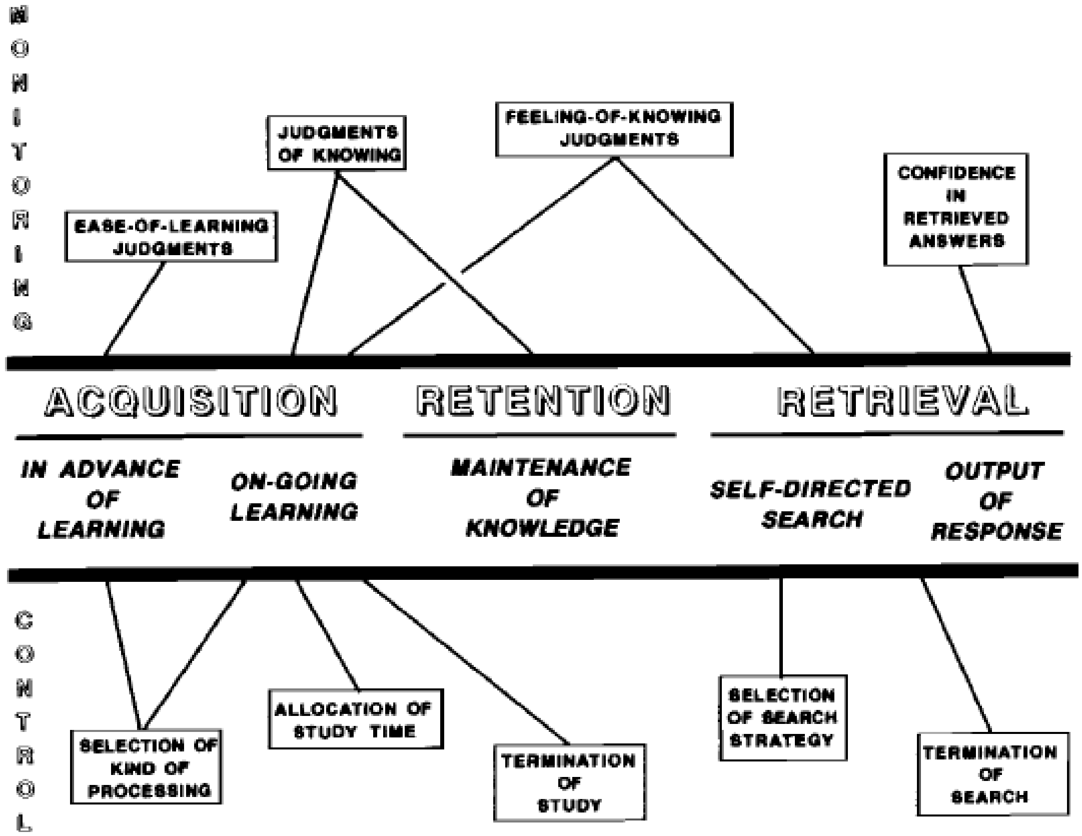

Figure 2. A detailed representation of control and monitoring and the main stages in Nelson and Narens' (1990) metacognitive framework.

future performance. The prospective memory judgment of most interest to this study is a judgment of learning (JOL), which is a subjective rating of the probability that the item will be remembered at a later time. JOLs, along with feelings-of-knowing (FOKs) and ease-of-learning (EOL) judgments, are used to evaluate monitoring during the first presentation of the stimulus. During the acquisition stage, learners also make decisions about how they will study the items. These decisions, such as study time allocation and how an item will be studied again, are used to regulate control. During the retention phase of learning, participants evaluate their memory for previously-seen items and, in some cases, make more decisions about how the items are studied. Finally, the retrieval stage encompasses memory search and item production processes and can be evaluated using judgments of confidence that the retrieved item is correct (monitoring) or the effectiveness of a chosen retrieval strategy (control). Figure 2, taken from Nelson and Narens' (1990) paper, illustrates examples of the tools that can be used to measure control and monitoring during the acquisition, retention, and retrieval stages of learning.

Of the components of Nelson and Narens' (1990) proposed model of metacognition, control during the acquisition phase of learning is of most interest to the proposed research, namely the choices that learners make when presented with items that need to be retrieved at a later time. Two issues will be of particular interest: 1) how participants choose to space the re-presentation of stimuli for future study, and 2) the different ways in which participants choose to process stimuli during second viewing opportunities.

Allocation of Study Time

Metamemory research commonly examines the allocation of study time to measure control during learning. Typically, participants will be given a list of words or word-pairs and asked to remember the items for a memory test later in the experiment. Participants are allowed to view the items for as long as they wish after providing a judgment of how well they believe they will remember the items later without further study (i.e. JOL).

The amount of time that participants devote to the stimuli can be studied with respect to the JOLs that participants give, thereby allowing the allocation of study time to be examined as a function of perceived difficulty. A common finding for this paradigm is that participants spend more time viewing difficult items and less time viewing easier items (Cull & Zechmeister, 1994; Mazzoni, Cornoldi, & Marchitelli, 1990; Nelson, Dunlosky, Graf, & Narens, 1994; Nelson & Leonesio, 1988; Thiede & Dunlosky, 1999). A different method for evaluating the allocation of study time is to instruct participants to remember a list of words for a later test, but to instruct them that only a subset of the items are allowed to be restudied and that they should select the items that they would like to restudy. The items chosen for restudy are presented again after the initial viewing while the others are not shown again until the final test. Results from this paradigm show that, when given the choice, learners choose to restudy significantly more difficult items than easy items, and this pattern of findings is obtained for both young (Dunlosky & Thiede, 1998; Nelson et al., 1994) and older adults (Dunlosky & Hertzog, 1997; Hertzog & Dunlosky, 2011). This interaction between controlled and monitored aspects of learning led Dunlosky & Hertzog (1997) to propose a discrepancy-reduction model. In this model, individuals monitor their learning of either a single item or group of items and determine the level of mastery that they would like to achieve. If an item is thought to be below the level of mastery set by the individual, then more resources will be allocated to that item in order to reduce the discrepancy between the perceived and desired levels of mastery.

Some studies have shown, however, that difficult items are not always the first to be chosen to be restudied (e.g., Mazzoni & Cornoldi, 1993; Mazzoni et al., 1990; Mazzoni, Cornoldi, Tomat, & Vecchi, 1997; Thiede & Dunlosky, 1999). Metcalfe (2002) found that when the amount of time that individuals are given to restudy items is constrained, learners choose to restudy the easiest of the unlearned items. Metcalfe used these results to describe the region of proximal learning model, which states that learners will optimize their study habits and allocate more resources to items that are nearly learned over ones that are the most difficult when study time is limited (see also Metcalfe & Kornell, 2003, 2005; Kornell & Metcalfe, 2006). In short, the model predicts that learners will allocate the most time to the as-yet-unlearned items that are thought to be the most likely to be learned with additional effort.

Both the discrepancy reduction and region of proximal learning accounts explain how learners allocate study time with regard to item difficulty, but there are other factors related to the stimuli and to the learning environment that affect the amount of time that is given to review study items. Ariel, Dunlosky, & Bailey (2009) manipulated the reward structure of the experimental tasks in addition to the difficulty of the items by varying the probability that an item would appear on the final test as well as the amount of reward that the participant would receive for correctly completing the word pair at test. This method allows individuals the freedom to make study time choices based on either the subjective difficulty of an item or the potential for reward. The authors found that participants based their study decisions on the reward structure of the task more than on item difficulty. These results provided support for the agenda-based regulation (ABR) framework, which states that choices related to the allocation of study time can be driven by a learner's agenda (i.e. to collect as much reward as possible) as well as by item difficulty.

Distribution of Practice

Another important decision that learners make when studying new information is how practice will be distributed before a final test. Typically, practice can be distributed by the learner in one of two ways: massing and spacing. When a learner chooses to mass a set of items, she is choosing to see the items again immediately after they are initially presented. When a learner chooses to space a set of items, however, she is choosing to separate the first and second viewing opportunities by a longer interval of time, typically correlated in the literature with the number of items shown between the two viewings. The effect of distributing practice has been studied since the earliest days of psychological research (c.f. Ebbinghaus, 1964) with particular interest centering on the so-called spacing effect, in which items for which study opportunities are separated by longer intervals of time are more likely to be remembered than those whose study opportunities are massed (for reviews, see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Delaney, Verkoeijen, & Spirgel, 2010; Toppino & Gerbier, 2014). Despite the robust support for the spacing effect, learners seem insensitive to spacing when engaging in self-regulated learning, a trend that is found in both metacognitive monitoring and control. A classic example with respect to monitoring is a study done by Baddeley and Longman (1978) in which typing lessons for postal workers were either massed or spaced. Despite having better overall performance, the postal workers whose practice was spaced felt that they were not learning as much as their coworkers were and even threatened to quit the typing lessons if the spaced practice continued. Similarly, learners often rate items that were massed as being more likely or equally likely to be remembered at test than items that were spaced, despite actually recalling fewer massed items (Zechmeister & Shaughnessy, 1980; Dunlosky & Nelson, 1994; Benjamin, Bjork, & Schwartz, 1998). Participants' lack of knowledge of the benefits of spaced practice may influence the decisions that they make when regulating their own practice.

An empirical analysis of metacognitive control over the distribution of practice was first conducted by Son (2004). She presented pairs of GRE vocabulary words with their common synonyms for 1 s each and prompted participants to choose whether they would like to study the items again immediately after the first presentation (massing), to study the items again after all items had been presented at least once (spacing), or to not re-study the items again (done). The “done” option was included because previous research by the author indicated that learners adopt irrelevant response biases when they are forced to choose massing or spacing under circumstances in which they judge it unnecessary to see the item again at all.

Son’s (2004) results indicated that participants are more likely to choose to mass difficult items and space the easier ones. She interpreted these results as being consistent with her “metacognitive hypothesis”, which states that participants space their learning based on their prospective memory judgments for the to-be-remembered items. The author’s hypothesis was related to research on the expanding practice method (see Landauer & Bjork, 1978) in which learning is facilitated when items are initially presented after short lags (or degrees of spacing) and then after successively longer lags on subsequent restudy opportunities. This finding has been interpreted to indicate that short lags are more beneficial when to-be-learned items of information are brand new and difficult to remember but that longer lags become more beneficial as the items become better learned with further practice. Son connected this to her own study because, after the initial study opportunity, participants chose to mass the items that would be the least learned (i.e. high difficulty word pairs) and to space the items that would be better learned (i.e. low difficulty word pairs). The author concluded that learners’ tendency to mass difficult items and space easier ones reflected a metacognitive strategy that is item-specific and strongly guided by perceived difficulty.

Since the publication of Son's (2004) study, the interpretation of the results has been subject to question. Showing that learners choose to space easier items and mass harder ones suggests that they employ more effective study strategies for easier items, which conflicts with previous research examining study-time allocation. Is there an underlying absence of appreciation for the benefits of spacing or are there situations in which learners can utilize effective item-specific spacing strategies?

Benjamin and Bird (2006) related Son's (2004) work to a discrepancy-reduction hypothesis. The discrepancy reduction model states that learners allocate study resources to closing a gap between perceived knowledge and desired mastery (Dunlosky & Hertzog, 1997), and previous research has consistently shown that individuals tend to study difficult items for longer (Son & Metcalfe, 2000). Benjamin and Bird argued that Son's findings challenge the viability of the model as well as the reliability of the results found in previous study-time allocation studies. Therefore, they studied the choice of massing and spacing further in a related, but somewhat different paradigm. The authors dropped the "Done" option and manipulated the difficulty of the items experimentally rather than through the participants' JOLs. They allowed the participants to make item-by-item spacing judgments (i.e. "Sooner" or "Later") with the constraint that half must be spaced and half must be massed. Finally, rather than using Son's presentation durations (1 s and 3 s for the first and second presentations, respectively) they presented pairs for 5 s on both presentations. This was done to create conditions in which the spacing effect could be expected because previous studies had shown that the spacing effect could be reversed by shortening the time of item presentation (Metcalfe & Kornell, 2003). In Experiment 3, they presented stimuli for a much shorter duration, 0.5 s, in an attempt to eliminate spacing effects.

When spacing choices were analyzed, participants in Benjamin and Bird's (2006) study preferred to space difficult items more often than easy ones, consistent with their discrepancy-reduction hypothesis, but contrary to the findings and conclusions of Son (2004). This effect was not found, however, when presentation time was greatly reduced as participants were equally likely to choose to mass or space difficult items. Overall, the authors concluded that constraints in spacing choices elicit performance consistent with the discrepancy-reduction model and that reducing presentation time can eliminate, but not reverse, this trend.

Toppino, Cohen, Davis, and Moors (2009) argued that numerous methodological differences between Son (2004) and Benjamin and Bird (2006) precluded any definite conclusions about learners' spacing choices. In an attempt to resolve the conflicting findings, Toppino et al. used the same stimuli and procedures as Son (2004) and manipulated the amount of time that participants studied each word pair on its initial presentation (1 s vs. 5 s), while holding all other procedural variables constant. The second presentation of a pair was always for 3 s. When the duration of an item's initial presentation was 1 s, Son's finding that the proportion of spaced items declined with increasing difficulty was replicated, but, when the duration of item presentation was extended to 5 s, the participants' choices were reversed. Toppino et al. questioned whether participants could reliably perceive the entire word pair when it was presented for only 1 s. In a follow-up experiment, Toppino et al. replicated the 1 s condition, but asked participants if they had seen the entire pair after each item was initially presented. The participants' choices replicated those made in the 1 s condition of the first experiment, but when items that were judged to be non-perceived were factored out of the analysis, participants did not show a preference to space easy or difficult items. These results suggest that items presented for a short amount of time (e.g., 1 s) are less likely to be fully perceived and encoded by the learner and that the spacing choices made for difficult items were not based solely on an implicit studying strategy, but were influenced by a desire to see a partially perceived word pair again right away. Thus, participants in Son's study chose to mass difficult items not so much because they thought that they would forget the item, but because they did not fully perceive the high-difficulty item after its 1 s presentation. However, when perceptual difficulties are avoided, such as when the duration of presentation is extended to 5 s, participants prefer to distribute rather than mass practice, and this preference becomes stronger with increasing item difficulty (Toppino & Cohen, 2010).

Type of Practice

Choosing how to practice items is another important decision that learners make when studying. Typically, there are two types of practice that are investigated in the metacognitive literature: restudying and taking practice tests. While restudying refers to presenting items in the same manner as they were first shown, practice tests do not show the full item, prompting the learner to retrieve the missing information during practice (e.g. covertly or overtly providing the target word when shown the corresponding cue). A well-established phenomenon in this field is an increase in memory performance for items that are practiced as tests rather than restudied, which is referred to as the testing effect (Carrier & Pashler, 1992). However, some studies have found that the beneficial effect of practice tests is apparent primarily after a relatively long retention interval

(Wheeler, Ewers, & Buonanno, 2003; Roediger & Karpicke, 2006; Toppino & Cohen, 2009).

Monitoring the effects of testing vs. restudying. Using practice tests is a very effective tool in learning, but participants are not always aware of its benefits. In their first experiment, Roediger and Karpicke (2006) prompted participants to read a number of passages which were followed by either a re-presentation of the passage (restudy) or by an initial free-recall test. A final test on their memory for these passages occurred after a 5-minute, 2-day, or one-week interval using a free-recall task. Better performance for practice test items was observed for 2-day and 1-week retention intervals, but for the 5-minute interval, restudying the items was more effective. In their second experiment, the authors varied the type of practice they would experience (i.e. repeated study, repeated study plus single test, or repeated test) and when the final recall test would occur (i.e. after a 5-minute distractor task or after one week). After all of the restudy and practice test trials, participants were asked to indicate how likely they thought it would be that they would remember the items after one week. Similar to their first experiment, participants' final free recall was more likely to benefit from restudying when the retention interval was short (i.e. recall test after 5 minutes) and more likely to benefit from testing when the retention interval was longer (i.e. recall test after one week). The strength of this effect was moderated by the frequency of the additional study opportunities: memory performance for items in the repeated study condition benefitted most from short retention intervals, while memory performance for items in the repeated test condition benefitted most from longer retention intervals. When asked to rate their confidence for remembering the passages later, however, participants indicated higher confidence for the items that were restudied, showing that they were not fully aware of the benefits of testing.

Karpicke and Roediger (2008) also investigated the importance of testing, but incorporated drop-out techniques. In their procedure, participants were presented with foreign language word pairs during a study period, then were given a practice test on that information, then presented the information again, then given another practice test, and so on. In total, the participants alternated between 4 study periods and 4 test periods. There were 4 conditions of drop-out: 1) none of the items were dropped; 2) mastered items were dropped from study periods, but still tested; 3) mastered items were dropped from test but not study periods; and 4) mastered items were dropped from both study and test periods. Mastered items were classified as vocabulary word pairs that had been correctly recalled in the previous test period. After the study phase, the participants were asked to predict how many of the 40 pairs they thought that they would remember after one week. The participants returned for a cued-recall test a week later.

The authors found that final recall was best for the condition in which no items were dropped and the condition in which items were dropped from study, but not test. They concluded that testing, not restudy, is the critical factor in successful long-term retention. Of most importance to this section, however, are the judgments that were provided by the participants. On average, participants believed that they would remember only 50% of the words on a future test, a percentage that did not vary significantly across the four study conditions. These data indicate that participants were not aware of the benefits of repeated retrieval since there was no difference in their judgments between conditions that favored testing and conditions that favored restudy. Karpicke and

Roediger (2008) argue that these results uncover a mismatch between real memory performance and conventional wisdom in education: learners believe that items are learned after they are retrieved a first time and drop them from study, yet repeated testing without dropout yields the best long-term recall performance.

Control of testing and restudying. Much of the research regarding control of type of practice also shows that learners do not appreciate practice tests. This conclusion is based on questionnaires asking how learners would choose to study a list of items as well as on prospective memory judgments. Kornell and Bjork (2007) investigated learners' awareness of the testing effect through a survey assessing the use of practice tests during study. Ninety-one percent of the participants indicated that they use quizzes to help them study, but only 18% believed that quizzes were a learning event. After closer examination, the authors found that the majority of these respondents mostly use these quizzes to gauge how well they know the material. Thus, learners do have some appreciation for the benefits of practice tests, but they seem to feel that tests help them assess how well they have learned something and do not seem to realize that tests are very effective at directly improving their learning.

Karpicke, Butler, and Roediger (2009) also surveyed participants about which strategies they use when studying for exams. In the first section of the survey, participants were prompted to list their preferred studying strategies in a ranked order. Like Kornell and Bjork (2007), the results indicated a lack of appreciation for practice tests. Only 11% of the respondents indicated that they used practice tests and only 1% placed the use of practice tests as their first option. In the next section of the survey, Karpicke et al. asked the respondents to indicate how they would study for an upcoming hypothetical test and were given a list of three ways in which to study. Unsurprisingly, participants were most likely to choose to go back and restudy the entire text (57%) and much less likely to choose to try and recall the material without restudying it (18%). When the test option was amended to include restudying along with practice tests, however, participants were about equally likely to choose to test themselves (42%) and to restudy the material (41%). Similar to Kornell & Bjork, the respondents indicated that they used practice tests to gauge how well they knew the material.

In addition to these questionnaire-based studies, behavioral experiments have also been conducted with regard to metacognitive control of testing vs. restudy. Kornell and Son's (2009) second experiment assessed participants' choices of the type of practice using word-pair stimuli. If they chose restudy, each item was either shown in full (cue + target). If they chose practice tests, each item was presented as a practice test in which only the cue is shown with space for the participants to provide what they believed the target to be. On practice-test trials, participants in different groups either received a 1 s presentation of both the cue and target (feedback), or they were advanced to the next trial (no feedback) after a response was provided. Participants determined the nature of the second presentation of the items by choosing whether they would like to view all of the word pairs again as either practice tests or restudy trials. This procedure was repeated for four different successive lists of word pairs, and participants were tested on their memory for all of the studied items by a recall test at the end of the experiment.

Initially, the authors found no preference for testing over restudying, but the tendency to choose testing increased over successive lists, and this trend was more pronounced when feedback was provided after the tests. This pattern of results, which was confirmed in a subsequent study by LaVan, Pagano, and Toppino (2011), suggests that learners are unaware of the advantages of testing initially but that their appreciation of testing increases with experience in a particular learning context, especially when the tests are followed by feedback.

Karpicke (2009) also examined learners' decisions to restudy or test as well as the conditions that encourage individuals to terminate learning. In this study, Karpicke ran 4 experiments in which all participants were asked to remember a list of Swahili-English word pairs, but each experiment varied by the nature of the study procedure and whether the learning conditions were controlled by the experimenter or the participant. In Experiments 1 and 2, participants studied the word pairs, took an initial practice test, and then were presented with the stimuli again in repeating subsequent viewing trials. The way in which additional trials were presented was dependent on the study condition. When a pair was correctly produced for the first time during the cued-recall practice trials, the remaining trials for that item either changed to all restudy trials, remained as practice tests, or were dropped from future study. These changes were either controlled by the researchers (Experiment 1) or by the participants themselves (Experiment 2). After each test trial, participants were asked to provide a JOL of how well they had learned the pair. A final cued-recall test was given to the participants after one week.

In Experiment 1, the researchers found a large effect of retrieval practice such that recall performance was greater for items that were presented in repeating practice test trials rather than those that were presented in repeated study trials. The participants' confidence judgments, however, indicated that they were not aware of the benefits of testing. In Experiment 2 in which participants were given control over how items were viewed after the initial practice test, the recall and JOL data replicated those in Experiment 1. The JOL results also indicated that the influence of retrieval fluency on their judgments caused participants to become overconfident, which led participants to choose to drop items that were not adequately learned from further practice tests.

The final two experiments explored learning choices during study periods. In Experiment 3, condition assignment was controlled by the experimenter. Similar to Experiments 1 and 2, participants were asked to remember 60 Swahili-English word pairs, but instead of providing JOLs after practice-test trials, participants were asked to provide JOLs after study periods. They were instructed to remember these items for a final test and that, during study, they would view each item over 6 blocks. For the STSTST group, participants were instructed that they would have alternating study and practice-test periods. In the SSSTST group, participants were told that they would receive three study periods before a practice test as well as an additional study period before the second practice test. In the final group, SSSSST, participants were told that they would have 5 study periods before a practice test. Like Experiments 1 and 2, the condition with the most practice-test trials (i.e. the STSTST group) had the highest final recall performance.

The last experiment's procedure was similar to Experiment 3, with the exception that subjects were asked to indicate whether they would like to study the pair in the next period, test the pair at the end of the study period, or drop the pair altogether. In Experiment 4, participants were most likely to choose to study and not test items in the first block. Preference for study decreased over blocks while the preference to drop increased and the preference to test only slightly increased. The results of Experiments 3 and 4 show that retrieval during study is beneficial for memory retention, but that participants do not make learning choices that reflect an appreciation for this benefit. Instead, they choose repeated study in earlier blocks and, in later blocks, drop items that are thought to be remembered rather than testing them.

The experiments by Kornell and Son (2009) and by Karpicke (2009) all involved relatively long spacing intervals between the successive presentations of to-be-learned pairs. This methodological choice could have influenced the degree to which learners chose to take practice tests because they would be less likely to retrieve an item correctly after a long spacing interval. This possibility was tested by LaVan et al. (2011, Experiment 1). In their experiment, participants were asked to remember 3 lists of 12 word pairs each. After initially viewing the pairs and providing a JOL, participants were asked whether they would like to see the item again as a restudy trial (cue + target), as a practice test (cue only), or to skip the second presentation entirely. Participants were assigned to one of two groups depending on whether the second occurrence of a pair occurred after a short spacing interval (2 intervening items) or after a long one (after all items had been presented once). Participants chose to test difficult items (i.e., those with low JOLs) more often when spacing was short than when spacing was long and tended to prefer testing for easy items regardless of the spacing condition. LaVan et al. concluded that the learners' decisions implied a belief that successful retrieval is important when using practice tests, which is why they chose to test difficult items more often when the spacing was short rather than long.

Effect of Feedback on Practice Tests

When pairs are being learned, practice tests involve presenting one member of the pair (the cue word) and asking learners to retrieve the other member of the pair (the target word). In some cases, participants are not informed of the correct alternative after they have responded. In other cases, they are provided feedback following their responses by presenting the entire pair (cue + target). This section examines the effects of providing feedback versus no feedback following practice tests, especially in the context of metacognitive research.

Effect on choosing to test. In general, much, if not all, of the information that is given to learners has been shown to have a significant impact on memory performance. In multiple-choice paradigms, for example, learners are shown the correct answer, but each showing of the correct item is also accompanied by the presentation of incorrect answers. So, while participants can discriminate correctly at the beginning of the task, they are still exposed to misinformation when all of the possible choices are considered by the participants, and participants may judge the wrong answers to be true by the end of the learning phase (Toppino & Luipersbeck, 1993; Brown, Schilling, & Hockensmith, 1999; Roediger & Marsh, 2005). This effect shows that simply repeating information can increase the probability that the information will be judged as true (Hasher, Goldstein, & Toppino, 1977; Bacon, 1979; Begg, Armour, & Kerr, 1985), regardless of whether the information is actually true or not, perhaps by increasing the ease with which it can be retrieved.

With regard to studying, the feedback information that is presented after practice tests has been shown to have a significant influence on later recall. In a meta-analysis of the instructional benefits of feedback after tests, Bangert-Drowns, Kulik, Kulik, & Morgan (1991) proposed that effective feedback simulates the process of updating memory information when correcting erroneous retrievals. Providing feedback after tests has also been shown to be influential in tasks where learning from feedback is incidental. McDaniel and Fisher (1991) examined the effect of testing with feedback using two experiments in which participants were asked to learn a set of trivia questions for a test later on. In both experiments, participants initially studied 54 facts for 6s each and provided a judgment on how comprehensible each fact was. After seeing all of the facts once, 36 items out of the original set were shown again for 5 s each in different conditions of study and feedback. There were two conditions of restudy: the rote rehearsal group, in which participants were instructed to repeat the fact verbally until the next item appeared, and the elaborative rehearsal group, where participants were asked to produce a reason for why the fact is true. The 2 test-with-feedback conditions were slightly different: After presentation of a question (5 s) and the correct answer (5 s), participants in the rote rehearsal group were asked to repeat the answer while those in the elaborative rehearsal group were asked to give a plausible reason as to why the answer is true. At test, participants were asked to answer 54 questions related to the facts presented at the beginning of the study. The results of the study indicated that, whereas individuals had better recall for items that were presented with feedback than items that were not, test performance for the two test-with-feedback groups did not differ. These results show that learning occurs incidentally from feedback, regardless of whether the feedback undergoes low-level processing (i.e. rote rehearsal) or elaborative processing (i.e. "give a reason why this fact is true").

Since learners benefit from practice test feedback, one might expect that learners would appreciate feedback and make adaptive decisions based on the information it provides, but the empirical data is varied. Agarwal, Karpicke, Kang, Roediger, and McDermott (2008) used two experiments to examine the testing effect with open-book tests, in which students are allowed to view the textbook and their own notes during a test, and closed-book tests, in which students were not allowed to view any extra material during a test. In the first experiment, participants were asked to study six prose passages that they might or might not be tested on at the end of the experiment. After each passage, participants either viewed the items again or moved on to the next passage depending on the study condition to which it was assigned. When passages were shown again, they were presented either as an open-book test or as a closed-book test with or without feedback. The authors found the testing effect in both open- and closed-book test conditions, but importantly, they found better memory performance for passages that were presented again as closed-book tests with feedback and as open-book tests as opposed to closed-book tests without feedback. In their second experiment, Agarwal et al. used the same study conditions as in the first experiment, but varied the number of times that study opportunities were given so that learners could read a passage once, twice, or three times. The participants were also asked to give aggregate JOLs for how well they thought they would remember the passages after one week at the end of the study phase. Open-book tests and closed-book tests with feedback were shown to be the most effective for future recall in comparison to other study conditions. However, despite the clear benefits learners received from the information given in the closed-book test with feedback and open-book test conditions, participants judged items that were studied more than once to be more likely to be remembered than items in any other condition.

A mismatch between learner's perception of the role of feedback and the actual benefits of feedback is also found in spacing decision paradigms. In a pair of experiments discussed earlier, Kornell and Son (2009) examined the effectiveness of, and learners' beliefs about, choosing to test, but, in the present context, a key manipulation in their study was the presence or absence of feedback after practice tests. In the feedback condition, participants were shown the correct word pair for 1 s after a practice test. When the experimenters controlled whether items were studied or tested during practice, recall performance for items that were tested during practice was higher than when items were studied, but that JOLs were higher for the studied items. The effect of feedback was significant for neither recall nor JOLs.

When participants were allowed to decide whether each of four successive lists would be restudied or tested on its second presentation, they preferred to study in the first block, but chose to test more often in later blocks. By list 4, participants who received feedback during practice tests chose to test the items significantly more than participants who did not receive feedback at all. However, JOLs again were lower for items that were tested, regardless of the presence of feedback. The authors concluded that the pattern of higher preference for testing with feedback without increased JOLs is due to learners using feedback following practice tests as confirmatory information, which was reflected in the participants' survey responses.

Another previously discussed study by LaVan et al. (2011, Experiment 1) presented three successive lists and varied the presence or absence of feedback following practice tests. When the authors controlled the length of time between initial study trials and second presentations (i.e. short or long), participants' choices differed between the two feedback groups such that those who did not receive feedback had a slight, nonsignificant preference for a restudy strategy while those who did receive feedback preferred to test significantly more than to restudy in the second and third lists. These results are consistent with the results reported by Kornell and Son (2009).

Effect of feedback on choosing to space. The effects of practice test feedback on spacing choices have also been studied. In their second experiment, LaVan et al. (2011) explored how spacing decisions are influenced by expectations of type of study (restudy vs. practice tests) and the presence or absence of feedback following practice tests. Participants only studied one list of 36 words and provided a JOL for each item. They were asked to choose the spacing between the first and second viewing opportunities (i.e. sooner, later, or skip) while the type of presentation (restudy or practice test) was manipulated by the experimenters. When the second presentation was a restudy opportunity, participants preferred to space items, regardless of perceived item difficulty. This pattern of results is generally consistent with the findings of earlier studies in which the procedures ensured full perception of the initial presentation of to-be-learned pairs (e.g., Toppino & Cohen, 2010). In contrast, when the second presentation was a practice test, LaVan et al. found that learners preferred to mass relatively difficult items (low- and medium JOL items). However, the presence or absence of feedback following practice tests did not significantly affect learner's spacing choices.

The authors hypothesized that feedback is unappreciated as a study tool by learners. Choosing to mass difficult practice test items and to space easier ones makes sense in the absence of feedback because successful retrieval is necessary if anything is to be learned from the second presentation of an item. However, these choices would not be advantageous when feedback is present and the correct information is always given.

Under the latter circumstances, greater spacing for all items would be the optimal strategy, and learners adopt just such an appropriate strategy when the second presentation of pairs involves a restudy opportunity. The fact that they do not adopt the same strategy when the second presentation involves a practice test followed by feedback suggests that they do not interpret feedback to be a learning or restudy opportunity.

Pagano and Toppino (2013) directly assessed the efficacy of learners' spacing choices when the second presentation of an item involved a practice test, with or without subsequent feedback. In their first experiment, participants studied 36 word pairs for 6s each, gave a JOL for each item, and were presented with three spacing options: sooner (after 2 intervening items), later (at the end of the list), or skip. All of the items would be presented as practice tests unless they chose to "skip", in which it would be dropped from study. For practice test trials without feedback, the cue was presented alone for 8s. For practice test trials with feedback, however, the cue was presented alone for 6s followed by presentation the cue + target for 2s. The participants were randomly placed into either a feedback or no-feedback condition. Unlike LaVan et al. (2011), however, one-third of the spacing choices were dishonored, allowing final recall performance to be unconfounded by learners' choices. The results show that dishonoring "sooner" choices in the no-feedback condition (i.e. presenting the word pair as "later") impaired final recall compared to honoring the choices, indicating that choosing to mass practice tests is an efficacious strategy when feedback is not given. When "sooner" choices were dishonored in the feedback condition, though, recall was better than if the choices had been honored, indicating that choosing to mass test items when feedback is available is not optimal for memory performance.

In the second experiment, the methods were the same as Experiment 1, but with the exception of the feedback conditions. In addition to a no-feedback group, there were two feedback groups, differing in the duration of feedback (1 s vs. 4s). The results from this configuration show that spacing choices were not affected by the presence or duration of feedback but that recall was better when practice tests without feedback occurred sooner and when practice tests with feedback occurred later. These results suggest that, although the effectiveness of spacing varies with the presence or absence of feedback following practice tests, learners do not view practice tests with feedback to be different from practice tests without feedback. Thus, learners receive the benefits of spacing practice tests with feedback, but still choose to mass practice tests to ensure successful retrieval. They apparently view feedback as information that can confirm their practice-test response rather than as information from which they can learn.

Effect of framing feedback. LaVan et al. (2011) and Pagano and Toppino (2013) showed that providing feedback after practice tests does not affect learners' choices. When given the choice to mass or space practice-test trials, learners make decisions that indicate a desire to ensure a successful retrieval. That is, they space items that are thought to have a greater chance to be retrieved at longer intervals (i.e. easy word pairs) and mass items that are unlikely to be retrieved at longer intervals (i.e. difficult word pairs). When an opportunity to study the full word pair is added after retrieval practice (i.e. practice test feedback), the potential benefits of this restudy trial are ignored, causing learners to make the same study choices as when feedback is not given. Ignoring feedback could reflect an implicit belief that self-testing is an evaluative tool and that feedback provides confirmatory information about whether the item is remembered or not (Bjork, Dunlosky,

& Kornell, 2013). If this is the case, then framing the feedback that follows practice tests as opportunities to learn within the instructions of the experiment could alter how learners perceive self-testing and, in turn, affect their study decisions.

Several studies in other areas of metacognition have shown that framing task instructions can affect how memory is monitored. Koriat, Bjork, Sheffer, and Bar (2004) provided an early account of framing when examining theory- and experience-based memory predictions over varying retention intervals. While participants are generally good at predicting performance on tests that occur immediately after study (see Nelson et al., 1994), the accuracy of their judgments significantly decreases when asked to make judgments about performance 24 hours and 1 week in the future. The relative inaccuracy of these JOLs is due in part to the influence of heuristics that are based on experience, such as processing fluency and target retrieval. Recall predictions that are based on participants' personal theories or beliefs however, are more accurate at predicting "forgetting" that occurs over longer retention intervals. In order to emphasize the importance of theory-based judgments, the researchers presented seven experiments that varied with regard to framing task instructions.

When participants were asked to remember a list of word-pairs and judge how many they would remember after 3 different retention intervals (i.e. immediately, after 24 hours, and after 1 week), JOL accuracy for tests at longer retention intervals was significantly worse (due to overconfidence) than JOL accuracy for a test occurring immediately after study. When participants were asked to judge how others would perform on a recall test after the three retention intervals and then perform the experiment themselves, JOL accuracy for both themselves and others was high for all three retention intervals. Finally, when the instructions were framed to emphasize forgetting (e.g. "How likely do you feel that you will forget this item?") rather than remembering (e.g. "How likely do you feel that you will remember this item?"), participants were more accurate at predicting their own memory performance, regardless of the retention interval. Overall, these results suggest that, in making JOLs for different retention intervals, learners are inherently biased towards basing their judgments on experience (i.e. fluency and ease of retrieval), causing them to make inaccurate recall judgments. When the instructions are framed to lessen the emphasis of experience-based judgments or to emphasize the role of forgetting over longer retention intervals, however, participants provide more accurate recall predictions.

Ariel, Hines and Hertzog (2014) studied framing in a related context. In a previous series of experiments, Kornell and Bjork (2009) had presented a list of pairs for up to four trials with each trial involving a study opportunity and then a test. After each pair was presented on the first study trial, participants were asked to make a prediction of learning (POL) with respect to how well they would remember the pair after varying numbers of additional study and test trials (up to four total trials). They found evidence of a "stability bias" in which participants' POLs were relatively unchanged (i.e. stable) as the number of trials increased from one to four. Ariel et al. proposed that this stability bias was caused by how the situation was framed for the participants. Kornell and Bjork's instructions emphasized testing over studying by asking participants to predict their performance on a particular test on the first through the fourth trials.

Ariel et al. (2014) modified Kornell and Bjork's (2009) procedure so that the instructions emphasized testing or the number of study opportunities before the critical test. Overall, participants made more accurate POLs when the additional viewing opportunities were framed as study, but exhibited more of a stability bias when the viewing trials were framed as test. The authors concluded that learners believe that multiple study opportunities will increase recall and that this belief influences memory predictions when it is accessed while the predictions are being made. However, framing the prediction in terms of tests may reduce the accessibility of this belief so that participants exhibit a stability bias by underestimating how much they will learn on successive trials.

Ariel et al.'s (2014) results provided an interesting link to research on practice test feedback. In agreement with LaVan et al. (2011), Ariel et al.'s results imply that learners do not believe that testing will benefit their learning, which is shown by their tendency to provide similar recall predictions for items whose instructions emphasize testing, regardless of the benefits that repeated practice will have on recall performance. Apparently, learners do not consider the learning potential of practice tests.

Returning to the issue of feedback, two experiments reported by Heslin et al. (2014) addressed several reasons why the presence or absence of feedback on practice tests does not affect whether learners choose to mass or space these tests. Their first experiment addressed possible procedural explanations for this phenomenon, and their second experiment explored the possibility of a framing effect.

Heslin et al.'s (2014) Experiment 1 addressed two possible procedural explanations for this phenomenon. First, in previous studies, a cue-target pair was presented more briefly as feedback than as a restudy opportunity. This has been true even in experiments that varied the duration of the feedback information (Pagano &

Toppino, 2013). Thus, the relative brevity of feedback may contribute to learners underestimating its effectiveness. Heslin et al. presented 48 pairs for study, and learners made JOLs and then decided whether to take the practice test sooner, later, or not at all (skip). They compared learners' choice behavior under three different conditions which varied in the nature of the second presentation of to-be-learned items: (1) a restudy condition in which the entire pair was re-presented for 5 s; (2) a practice test condition without feedback in which the cue alone was presented for 5 s; and (3) a practice-test condition with feedback in which the cue alone was presented for 5 s, followed by the cue plus target (feedback) for another 5 s. Thus, the full cue-target pair was presented for the same duration in both the restudy and in the test-with-feedback conditions.

The second concern addressed by Heslin et al.'s (2014) Experiment 1 was that, in previous studies, there may have been implicit demands that influenced learners' choices. Both the LaVan et al. (2011) and Pagano and Toppino's (2013) studies required overt responses for both practice-test conditions (i.e. feedback and no-feedback), so it is possible that implicit demands to look good by performing well on the practice tests may have led participants to focus on practice-test performance rather than on the long-term benefits of their spacing choices. Therefore, Heslin et al.'s participants were not allowed to provide an overt response during practice-test trials and were instead asked to quietly think of their answer, thus, minimizing or removing the impetus to be concerned about practice-test performance. The emphasis was on final test performance.

Despite equalizing the time during which the cue and target were presented and despite requiring covert responses during practice tests, the results were similar to those obtained by LaVan et al. (2011). Spacing was chosen most for difficult items that were shown again as restudy trials while massing was chosen most for difficult items that were shown again as practice tests trials, regardless of whether feedback was given or not. Like before, these choices reflect an over-reliance on practice-test retrieval as well as a fundamental lack of appreciation for the benefits of feedback following practice tests.

In view of the fact that Experiment 1 failed to alter the usual pattern of results, Heslin et al. (2014) considered a framing explanation of the effect in Experiment 2. Because learners indicate that they use practice tests to evaluate their knowledge rather than to learn from them (e.g. Kornell & Bjork, 2007; Bjork et al., 2013), perhaps referring to these trials as "practice tests" in the instructions influenced participants to view them as providing evaluative information rather than as opportunities to restudy. In Experiment 2, Heslin et al. used similar techniques to the first experiment, but modified the instructions to emphasize either restudying or testing. As in Experiment 1, participants studied 48 word pairs, provided JOLs for those items, and were asked to give a decision on how to space each item (i.e. "Sooner", "Later", or "Done"). There were three conditions: a restudy condition and two conditions involving practice tests with feedback. The restudy condition was identical to that used in Experiment 1. The "test-framing" condition was identical to the practice-test-with-feedback condition in Experiment 1. The instructions described the task simply as involving practice tests with feedback in which their answer to the practice test would be followed by the full cue-target pair for 5 s. The "restudy-framing" condition was identical to the test framing condition except that the instructions described the task as involving restudy opportunities preceded by practice tests.

The results for Experiment 2 showed an effect of framing. Learners' choices for the restudy- and test-framing conditions replicated the choices made in the restudy and practice test conditions in Experiment 1, respectively, while learners' choices in the restudy-framing condition mirrored those in the restudy condition. Specifically, participants in the restudy-framing condition chose to space rather than mass difficult practice test items when the instructions emphasized the role of feedback as a study tool. These results show that learners can show an appreciation for the benefits of spacing tests, but do so only when the restudy potential of feedback is explicitly emphasized.

Heslin et al's. (2014) results show a direct influence of framing on how learners view practice test feedback. When framed to emphasize the restudy benefits of feedback, learners are more likely to treat practice test feedback as information that will be useful for studying rather than as a confirmatory tool as previous research has indicated. At the level of study decisions, these results show that learners can change how feedback is processed, but a final question regarding how learners process the feedback itself still remains. If learning choices made when feedback is framed as a study opportunity can reflect an appreciation for the benefits of practice tests, then it may be more likely that learners utilize and learn from practice test feedback when its benefits are explicitly stated. Although Pagano & Toppino (2013) found that participants do learn from feedback even while their choice behavior indicates that they discount feedback's effect, it remains possible that participants learn better when feedback is considered to be a learning opportunity. If this is the case, then recall should be better for participants who view practice tests with feedback as viable learning opportunities than for participants who do not.

The Present Research

The purpose of the present experiment is to assess the effect on final recall of framing practice tests as being followed by feedback versus being followed by a restudy opportunity. Although the participants in Heslin et al.'s (2014) Experiment 2 took a final recall test, the data are difficult to interpret because the participants decided which items would be massed and which would be spaced. This problem will be avoided in the current study because item presentation will be fully controlled by the experimenter.

All participants studied a list of 48 word pairs of varying associative strengths and were asked to remember these items for a test later in the experiment. After viewing each word pair, the participants made a prospective memory judgment on how well they thought the item would be remembered later (JOL). Each item was shown again after 4 – 12 other item presentations as a practice test with feedback. Participants were randomly assigned to three conditions: Test-framing, restudy-framing, or a no-feedback control. In the test-framing condition, the directions indicated that participants would experience a practice test followed by feedback, and in the restudy-framing condition, the directions will explain that they would receive a restudy opportunity preceded by a practice test. In the control condition, no feedback followed the practice tests. The primary question is whether final cued recall will be facilitated in the condition in which feedback is framed as a restudy opportunity relative to the condition in which it is framed simply as feedback. Inclusion of the no-feedback control allows us to determine the extent to which learners benefit from feedback at all.

LaVan et al. (2011) showed that participants' spacing decisions for items that were shown again as practice tests did not differ when feedback was present or absent, but did differ from the spacing decisions made when the second presentation was a restudy trial. Heslin et al. (2014) demonstrated the same effect when feedback was framed as feedback. However, when feedback was framed as a restudy opportunity preceded by a practice test, spacing decisions were the same as when the second presentation involved a restudy opportunity instead of a practice test. These results suggest that learners usually use feedback to gauge the correctness of their practice-test responses and do not think of the presentation of feedback as a learning opportunity in itself. This opens the possibility that feedback often is processed in a way that minimizes learning. However, when the potential for feedback to assist learning is explicitly stated in the instructions, then it is possible that the implicit benefits of feedback will be complemented by more effective forms of processing, leading to better final test performance. At the very least, framing feedback as a learning tool could influence learners to allocate more resources to processing feedback during studying, which may enhance learning from feedback under these circumstances.

Method

Participants

Ninety-three undergraduates recruited from Villanova University and the Georgia Institute of Technology participated in this study. During the summer semester, 29 undergraduate participants were recruited from Villanova University and participated in exchange for $10. During the fall semester, 18 participants from Georgia Tech and 46 from Villanova University were recruited using each institution's respective undergraduate subject pool in exchange for class credit. Each participant completed the experiment individually and was randomly assigned to either the no-feedback, test-framing, or restudy-framing condition. All participants were required to provide written consent before taking part in the study using a form that has been approved by the Institutional Review Board at both Villanova University and Georgia Tech.

The data from 3 participants in this study were excluded from the final analyses which were based on the data of 90 participants, 30 in each condition. In order to analyze the influence of the items' perceived difficulty or memorability, participants made JOLs to each item during learning, and their JOLs were later divided into categories of low, medium, and high JOLs for the purpose of data analysis. The 3 participants whose data were excluded did not produce enough variance in their JOLs to make their Vincentized data meaningful because the participants apparently used strategies such as relying on only one JOL for the majority of their responses.

Materials

The stimuli in the experiment consisted of 56 pairs of common English words that vary in difficulty. As in the LaVan et al. (2011) study, all of the items were selected from the University of South Florida Free Association Norms database (Nelson, McEvoy, & Schreiber, 1998) and relative pair difficulty was determined by the associative strength of the cues and targets. Half of the items were randomly selected from a list of word pairs of moderate associative strength (.050 - .054). These were the relatively easy items (e.g. biscuit – cookie). The rest were relatively hard items and were generated by randomly pairing unassociated cue and target words (e.g. kite – perfume). Any items that were judged to be incorrectly classified were removed from the list and replaced by a randomly-selected word pair of the correct associative strength.

The pairs were presented individually on the computer screen with one word (the cue word) above a dashed line and another word (the target word) below the line for initial study trials and feedback trials. All of the practice test trials had an initial presentation of the cue word above the dashed line and a space below it for the participant to complete the word pair. Test trials were followed immediately by presentation of the correct cue and target as feedback. The items were presented using the E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA).

Procedure

After providing informed consent, participants were seated at a computer with the experiment module and were asked to begin reading the directions on the screen detailing the procedure. They were informed that the main goal of the task was to learn the word pairs for a recall test at the end of the experiment. The instructions also provided details about the procedure and the scaling involved in making JOLs. In this, participants were instructed to use the full range of JOLs. If the participant was assigned to one of the two feedback conditions, then the instructions also specified how feedback would be given to the learner. Participants in the test-framing condition were instructed that the second viewing trial for each pair would be a practice test followed by feedback, whereas participants in the restudy-framing condition were instructed that the second viewing trial for each pair would be a restudy opportunity preceded by a practice test. Participants in the no feedback condition were simply instructed that the second viewing trial would be a practice test and that they would not be informed whether they were correct. At the end of the instructions, any questions that the participant had about the procedure were answered by the researcher.

After reading through the instructions, the participants performed two practice trials in order to become familiar with the procedure for making JOLs. Neither of the practice items was on the final test. For each pair, the participants viewed the item for 5 seconds and made a JOL based on how well they thought they would remember the item on a future test. The participants did not see these items a second time during the practice trials; instead, a screen appeared that reminded them that the second presentation would involve a practice test. Participants were encouraged to ask questions during this time.

During the initial phase of the experiment, the 56 selected word pairs were each presented twice: once during an initial study trial and again during a subsequent practice test. During the first study trial, a cue and target were presented for 5 s and immediately were followed by a prompt to make a JOL. Participants were instructed to make JOLs by depressing one of 11 keys, labeled 0 to 10, with lower numbers indicating less confidence that a particular item will be recalled at test and higher numbers indicating more confidence that the item will be recalled at test. The order in which words were presented initially was determined randomly. Afterwards, each word pair was presented again as a practice test after 4 – 12 intervening presentations, with the exact number of intervening presentations being randomly selected on a trial-by-trial basis. For all of the participants, the second viewing opportunity displayed only the cue with a space below it for the participant to complete the word pair. For participants in the two feedback conditions (the test-framing and restudy-framing conditions), the practice test trials were followed by 5 s of feedback, i.e. presentation of both the cue and target. Participants had 8 s to type an answer on the keyboard before either receiving feedback or advancing to the next trial in the no-feedback condition. Practice test trials did not require a JOL.

After all of the word pairs were presented twice, participants engaged in a 5-minute distractor task consisting of a mixed sequence of addition and subtraction problems. Participants either added or subtracted 3-digit numbers and were prompted to type the answer using a numeric keypad. After the distractor task was presented, the cues for each word pair were presented a final time in a cued-recall test. During this task, participants were presented with a cue and were asked to provide the target that completed the word pair using the keyboard provided. No feedback was provided. Participants had the opportunity to continue to the next test trial by pressing the “Enter” key, but were given 10 seconds to answer before being automatically advanced. If the participant provided no answer after continuing to the next trial, then the response was recorded as a blank, and if the participant only provided a partial or incomplete answer before advancing, then the incomplete response was recorded as their answer. Responses that were close to the correct target word but were incorrectly inflected or misspelled were recorded as being correct. After the final test, the individuals were compensated for their participation and thanked for their time.

Results

JOL Data

After the first viewing of each word pair, participants were asked to make a judgment of learning (JOL) based on how well they believed they would remember the item later. Judgements ranged from 0 – 10 with higher numbers indicating greater confidence that the item will be recalled and lower numbers indicating less confidence that the item will be recalled. The mean JOLs for each participant were analyzed using a 3 (Condition: no feedback, test-framing, restudy-framing) x 2 (normative word pair difficulty: moderate association, low association) mixed analysis of variance (ANOVA) with repeated measures on the second factor. Only a main effect of word pair difficulty was significant, $F(1,87) = 316.695$, $MSE = 19{,}702.756$, $p < .001$, $\eta_p^2 = 1$, such that participants gave higher mean JOLs (i.e., higher perceived likelihood of future retrieval) to word pairs of moderate associative strength ($M = 63.952$, $SEM = 1.593$) and lower mean JOLs (i.e., lower perceived likelihood of future retrieval) to word pairs of low associative strength ($M = 43.028$, $SEM = 1.467$). As expected, the mean JOL ratings did not differ amongst the three feedback condition, $F(2,87) = .061$, $MSE = 21.806$, $p = .941$, $\eta_p^2 = .059$, and the word pair difficulty x feedback condition interaction did not approach significance, $F(2,87) = .558$, $MSE = 34.705$, $p = .574$, $\eta_p^2 = .140$.

The JOL data were Vincentized (Vincent, 1912) in order to better analyze the relative difficulty of the word pairs. In this procedure, each participants' JOLs were roughly divided into three groups: low, medium, and high JOL levels. Items given JOLs in the low category are considered to be difficult while items given JOLs in the medium and high categories are considered to be medium in difficulty and easy, respectively.

These Vincentized JOL levels were used in the analysis of the participants' practice test and final recall accuracy.

Practice Test Recall

When items were shown a second time, participants were asked to type what they believed was the target when prompted with the cue and then were shown the correct cue and target as feedback if they were in one of the feedback conditions. The proportion of correct items recalled at test was submitted to a 3 (Condition: no-feedback vs. test-framing vs. restudy-framing) x 3 (JOL level: low vs. medium vs. high) mixed ANOVA with repeated measures on the second factor (Figure 3). Mauchley's Test indicated that the assumption of sphericity had been violated ($X^2 = 6.665$, $p = .0365$), so all within-subjects degrees of freedom were adjusted using the Greenhouse-Geisser correction ($\varepsilon = .931$). The main effect of JOL level was significant, $F(1.861, 161.925) = 129.797$, $MSE = 2.094$, $p < .001$, $\eta_p^2 = 1$, such that mean practice test recall differed for each level of Vincentized JOLs (i.e., perceived difficulty). As expected, participants recalled a lower proportion of practice test items when those items had been given a low JOL rating ($M = .569$, $SEM = .022$) than when they had been given a medium JOL rating ($M = .685$, $SEM = .021$) or a high JOL rating ($M = .863$, $SEM = .014$). This result is consistent with previous studies in which an effect of JOL ratings on subsequent recall performance has been demonstrated persuasively (though metacognitive accuracy is variable; see Nelson & Narens, 1990). In contrast, the condition to which participants had been assigned (no-feedback vs. test framing vs. restudy framing) had no significant effect on practice-test recall. Neither the main effect of Condition, $F(2, 87) = 1.673$, $MSE = .119$, $p = .194$, $\eta_p^2 = .344$, nor the JOL level x Condition interaction, $F(3.722, 161.925) = .377$, $MSE = .006$,

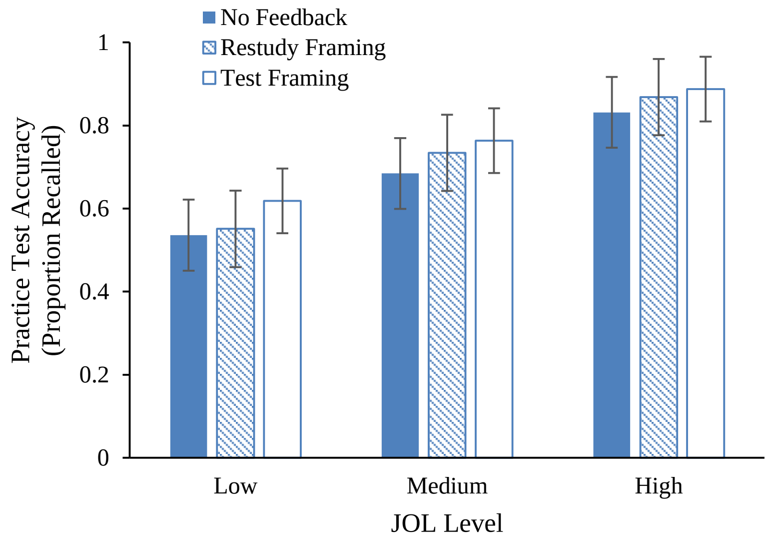

Figure 3. Mean proportion correct on the practice recall test as a function of feedback condition and JOL level. Error bars depict one standard error of the mean.

p = .819, η_p^2 = .121, was reliable.

Final Recall Test Data

After viewing each word pair twice, participants completed a 5-minute distractor task consisting of simple arithmetic and were then given a cued-recall test for each of the items shown during study. The proportion of items correctly recalled on the final test was submitted to a 3 (Condition: no feedback vs. test-framing vs. restudy-framing) x 3 (JOL level: low vs. medium vs. high) mixed ANOVA with repeated measures on the second factor (Figure 4). For this analysis, Mauchley's Test was not significant, X^2 = 5.002, p = .082, indicating that the assumption of sphericity had not been violated. [1] As in the previous analysis, the main effect of JOL level was significant, $F(2, 174) = 109.925$, *MSE* = 1.638, $p < .001$, η_p^2 = 1. Participants were more likely to recall items given high JOLs

[1] Since Mauchley's test came close to reaching significance, the within-subjects effects were also analyzed using the Greenhouse-Geisser correction. The correction did not affect the interpretation of either the main effect of JOL or the JOL x Condition interaction.

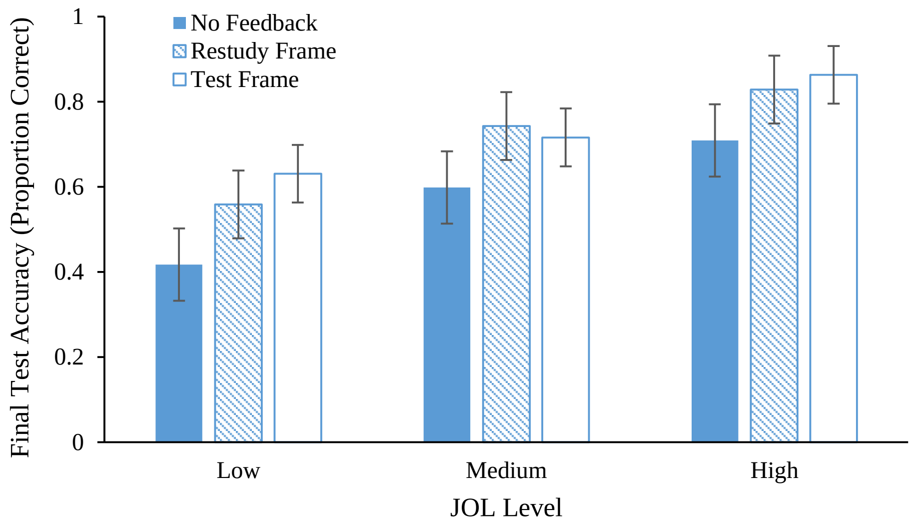

Figure 4. Mean proportion correct on the final recall test as a function of feedback condition and JOL level. Error bars depict one standard error of the mean.

(*M* = .804, *SEM* = .016) than items given medium JOLs (*M* = .685, *SEM* = .023) or low JOLs (*M* = .535, *SEM* = .024).

The main effect of condition was also significant, $F(2, 87) = 7.464$, $MSE = .684$, $p = .001$, $\eta_p^2 = .935$, such that mean proportion of items recalled on the final test was not equal for the three between-participant conditions. These data were further analyzed using Tukey's Honestly Significant Difference (HSD) paired comparisons, which showed that participants in the no feedback condition (*M* = .575, *SEM* = .032) recalled a significantly lower proportion of items than participants in the restudy frame condition (*M* = .712, *SEM* = .032) and the test frame condition (*M* = .737, *SEM* = .032). Final recall for participants in the test frame and restudy frame conditions, however, did not differ significantly from each other ($p = .847$). The framing condition x JOL level interaction was not significant, $F(4, 174) = 1.617$, $MSE = .024$, $p = .172$, $\eta_p^2 = .476$.

Supplemental Analysis of Final Recall

An additional analysis of final recall examined the influence of correct practice test recall on final cued recall. The potential to learn from feedback may differ depending on whether an item was successfully retrieved or not on the practice test. Therefore, the proportion of items recalled correctly on the final test were submitted to a 3 (Condition: no feedback vs. feedback framing vs. restudy framing) x 2 (Prior Retrieval History: correctly recalled vs. not recalled on the practice test) mixed ANOVA with repeated measures on the second factor. [2] As in the previous analysis, the main effect of condition was significant, $F(2, 87) = 21.844$, $MSE = .816$, $p < .001$, $\eta_p^2 = 1$. Again, these data were further analyzed using Tukey's HSD, revealing that, as in the original analysis, final recall for participants in the no feedback condition was significantly lower than final recall for participants in both the restudy frame condition and test frame condition. Final recall for test frame and restudy frame conditions were not significantly different.

The main effect of Prior Retrieval History was also significant, $F(1, 87) = 879.044$, $MSE = 12.632$, $p < .001$, $\eta_p^2 = 1$, such that word pairs that were correctly recalled during practice tests were more likely to be recalled at the final test ($M = .833$, $SEM = .011$) than word pairs that were not correctly recalled during practice tests ($M = .303$, $SEM = .021$).

Finally, the interaction between Prior Retrieval History and Condition was significant, $F(2, 87) = 38.406$, $MSE = .552$, $p < .001$, $\eta_p^2 = 1$, such that final recall for

[2] Items were collapsed across JOL conditions for this analysis because including JOL level in the analysis resulted in too many cells with no data (e.g., high JOL items were more likely to have no failures to recall on the practice test while low JOL items were more likely to no successful recalls on the practice test). The effect of JOL does not interact with Condition in any previous analysis, so it is unlikely that differences in JOL will contaminate the current analysis.

Table 1

Mean Proportion of Correct Final Recall and Standard Errors of the Mean, as a Function of Feedback Condition and Prior Retrieval History

	Feedback Condition	***M***	***SEM***
Items Correctly Recalled During Practice	No Feedback	0.808	0.016
	Test Framing	0.840	0.024
	Restudy Framing	0.840	0.017
Items Not Correctly Recalled During Practice	No Feedback	0.067	0.019
	Test Framing	0.480	0.049
	Restudy Framing	0.373	0.034

each of the different between-participant conditions was dependent on whether the items were recalled during practice tests or not. This interaction was further analyzed using a simple main effects analysis. In order to perform this analysis, the final recall data for each of the three conditions was separated into two conditions: recall for items that were recalled during practice tests and recall for items that were not recalled during practice tests. (For a comparison of conditionalized recall performance for the 3 feedback conditions, see Table 1.) Recall for items that were successfully remembered during practice tests was analyzed first and submitted to a one-way ANOVA, which indicated that final recall in the three feedback conditions did not significantly differ from each other, $F(2, 145) = .541$, $MSE = .026$, $p = .583$.[3] We then analyzed the simple main effect of feedback condition for items that were not successfully recalled on the practice test. Results of the analysis involving items that were not successfully retrieved in the practice test indicated that the recall proportions for the three feedback conditions were significantly different from each other, $F(2,145) = 52.36$, $MSE = .026$, $p < .001$. Tukey's HSD paired comparisons indicated that final recall was better in both feedback conditions than in the no feedback condition. Moreover, final recall was significantly better in the test-framing condition than in the restudy-framing condition.

---

[3] The simple effect analyses were carried out with a pooled error term derived from the omnibus ANOVA. Because the error terms involved in the pooling procedure were heterogeneous, the degrees of freedom were estimated using Satterthwaite's approximation (Winer, 1971).

Discussion

The current experiment set out to clarify the relationship between how practice test feedback is framed and final recall performance. Our research follows concerns raised by the results of LaVan et al. (2011) and Heslin et al. (2014), where learners seem to discount the memorial benefits of feedback and do not think of the presentation of feedback as a viable learning opportunity unless its benefits are explicitly stated. The results of the current experiment are consistent with the notion that practice test feedback aids in long-term memory retention, but the results of the *overall* analysis do not seem to support the hypothesis that learners process feedback differently depending on how it is framed. Overall, the benefits of practice test feedback for final recall were significant, but there was no significant difference in these benefits for individuals whose instructions framed feedback as a restudy opportunity and for individuals whose instructions framed feedback as simply following a practice test. However, subsequent analyses indicated that the restudy framing and test framing conditions differ with respect to final recall when one considers only items that were not previously recalled during practice tests. This last finding will be discussed later in this section.

The main focus of this experiment was the examination of final recall as a function of feedback following a practice test and as a function of how the feedback was framed in the instructions. Recall that in the research of both LaVan et al. (2011) and Heslin et al. (2014), learners were required to make their own decisions regarding the spacing between the initial and second viewing opportunities. While this provided valuable insight into how restudy opportunities, practice tests with feedback, and practice tests with feedback framed as a restudy opportunity affect study decisions about spacing, the final recall results were confounded with these decisions and are therefore uninterpretable. In the current experiment, participants did not self-assign items to conditions. All items were given spaced practice with the degree of spacing for each item being determined randomly. As a consequence, differences in final recall among the no feedback, test framing, and restudy framing conditions could be meaningfully interpreted. The final recall results show that participants in the no feedback control were significantly worse at recalling cue-target pairs than participants in either the restudy framing or test framing conditions. This result is consistent with previous findings regarding the benefits of practice test feedback (e.g., Pashler, Cepeda, Wixted, & Rohrer, 2005). The critical comparison in this experiment, however, involves the final recall for the two feedback conditions (i.e. test framing and restudy framing). The overall analysis indicated that final recall in the two framing conditions did not significantly differ from one another. We had thought it was possible that the restudy framing condition would perform better, indicating that learners devote more resources or more effective study strategies (e.g., deeper encoding) to processing practice test feedback when it is described as a restudy opportunity than when it is described as simply feedback. The fact that the final recall of the two framing groups did not differ significantly suggests that learners benefit from the presence of feedback, no matter how it is framed in the instructions.

Previous research by Heslin et al. (2014) indicated that participants chose to space practice tests more when feedback was framed as a restudy opportunity than when it was not. However, if framing causes feedback to be processed differently when it occurs, the difference in processing does not lead to an overall recall advantage for the restudy framing condition. Perhaps this should not be surprising in light of findings reported by

McDaniel and Fisher (1991) who explicitly controlled how feedback was processed. Participants were asked to remember trivia facts for a later test and given practice tests followed by presentation of the full memory item as feedback. Processing of this feedback was manipulated using different task instructions in two conditions. In a rote rehearsal condition, they asked participants to verbally repeat feedback information until the next item appeared, while in an elaborative rehearsal condition, they asked participants to provide a reason as to why the to-be-remembered fact was true. McDaniel and Fisher found that recall performance was not affected significantly by how feedback was processed.

In addition to overall final recall, we also were interested in final recall as a function of whether or not an item had been recalled during a practice test. The resulting analyses were intended as a more sensitive test of the hypothesis that framing affects the degree to which participants learn from feedback. Although there were no significant differences in overall final recall between the test framing and restudy framing conditions, it remained possible that there could be an effect with one kind of item (e.g., items that were not retrieved in the practice test) but not the other type (e.g., item that were retrieved on the practice test). Such an effect could be obscured in the overall recall data, depending on the size of the effect and the proportion of items that were and were not retrieved.

For items that were remembered during practice tests, there was no difference in final recall performance amongst the conditions, even between participants who received feedback (i.e. the test framing and restudy framing conditions) and those who did not (i.e. the no feedback condition). When items were not remembered during practice tests and no feedback was given, final recall was extremely low. When feedback was given, final recall was significantly higher, though it did not approach the level for items that had been correctly recalled on the practice test. A similar pattern of results was obtained by Pashler et al. (2005).

The most interesting finding was that final recall for items that had not been recalled during practice tests differed significantly between the two feedback groups. Participants in the test framing condition unexpectedly had higher final recall than those in the restudy framing condition. This result is different from the expectation under our original hypothesis, as participants in the restudy framing condition were expected to utilize feedback more effectively when it is framed in terms of its restudy benefits. Instead, participants performed significantly better on these items when feedback was simply described as feedback following a practice test.

The influence of framing on the effect of feedback on items that were not successfully recalled on the practice test is potentially a very interesting finding. However, it needs to be replicated in view of the fact that the effect was not predicted and was not apparent in the overall analysis of final recall. In the meantime, caution in interpreting the effect is appropriate although its possible importance demands some consideration of the potential underlying mechanisms.

The instructions in the restudy-framing condition were intended to focus participants' attention on the learning potential of feedback with the expectation that this would lead to more effective processing and better final memory performance. To the extent that participants in the test-framing condition did not focus on feedback, the framing instructions may have encouraged them to focus primarily on the act of taking the practice test. Thus, it is possible that participants in the test framing condition put more emphasis on attempting to retrieve the memory items during practice tests whereas participants in the restudy-framing condition did not. This analysis suggests that the effectiveness of feedback may be greater to the extent that it is preceded by effortful attempts to retrieve. In short, superior recall in the test-framing condition may be a manifestation of an effect known in the literature as "test-potentiated learning" (e.g., Hays, Kornell, & Bjork, 2013; Izawa, 1970; Kornell, Hays, & Bjork, 2009).

The phenomenon of test-potentiated learning was first described by Izawa (1966), who defined it as the improvement of later encoding of an item after a retrieval attempt has been made. Importantly, she suggests that the learning during tests does not only occur when items are retrieved, but also when they are not retrieved during practice (Izawa, 1970). This was supported by her finding that multiple unsuccessful retrieval attempts of a target led to better final recall than did a single unsuccessful retrieval attempt (Izawa, 1966; 1967). More recent studies on test-potentiated learning have suggested potential mechanisms to explain the phenomenon. Kornell and colleagues (Kornell et al., 2009; Hays et al., 2013), for example, argue for a semantic priming hypothesis. They suggest that taking a test primes knowledge related to the cue, specifically associations between the cue and target. In the case of a failed retrieval attempt, a broad network of semantic information related to the target is primed, and when corrective feedback is given immediately afterwards, this activation allows the correct cue-target pair to be effectively mapped for later retrieval.

A related but alternative account of test-potentiated learning involves the role of retrieval inhibition. A common finding in the memory literature is that the act of attempting to retrieve memory items suppresses the retrieval of other items (i.e. retrieval inhibition; Nickerson, 1984; Bjork, 1989; Anderson & Bjork; 1994). In the case of a failed cued-recall test, searching a semantic network activates incorrect semantic associates of the cue and temporarily suppresses the association between the cue and the correct target (Anderson, Bjork, & Bjork, 1994). If corrective feedback is given while the correct semantic association is suppressed, however, there is evidence that relearning of the cue-target pair may be more potent and more beneficial for future recall (Storm, Bjork, & Bjork, 2008). It is possible that in the current experiment, the increased performance for items that were not recalled during practice tests was facilitated by feedback following retrieval inhibition.

In summary, our results show that how practice-test feedback is framed affects final recall only for those items that were not recalled during practice tests. In this case, *not* emphasizing the restudy opportunity provided by feedback actually seems to enhance cued-recall for items misremembered during retrieval practice, possibly by encouraging more effortful retrieval and enabling greater test-potentiated learning. Framing feedback as a restudy opportunity may actually induce participants to use a suboptimal learning strategy by de-emphasizing the act of retrieval. Further research will be required to establish firmly how framing feedback on practice tests affects final recall as well as to establish the underlying mechanisms.

References

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

Anderson, M. C., & Bjork, R. A. (1994). Mechanisms of inhibition in long-term memory: A new taxonomy. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 265-325). San Diego: Academic Press.

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.

Ariel, R., Dunlosky, J., & Bailey, H. (2009). Agenda-based regulation of study-time allocation: When agendas override item-based monitoring. *Journal of Experimental Psychology: General*, *138,* 432-447.

Ariel, R., Hines, J. C., & Hertzog, C. (2014). Test framing generates a stability bias for predictions of learning by causing people to discount their learning beliefs. *Journal of Memory and Language*, *75*, 181-198.

Bacon, F. T. (1979). Credibility of repeated statements: Memory for trivia. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 241-252.

Baddeley, A. D., & Longman, D. J. A. (1978). The influence of length and frequency of training session on the rate of learning to type. *Ergonomics*, *21*, 627-635.

Bangert-Drowns, R. L., Kulik, C. L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, *61*, 213-238.

Begg, I., Armour, V., & Kerr, T. (1985). On believing what we remember. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *17*, 199-214.

Benjamin, A. S., & Bird, R. D. (2006). Metacognitive control of the spacing of study repetitions. *Journal of Memory and Language*, *55*, 126-137.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, *127*, 55-68.

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 & consciousness: Essays in honor of Endel Tulving* (pp. 195-210). Hillsdale, NJ: Erlbaum.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417-444.

Brown, A. S., Schilling, H. E., & Hockensmith, M. L. (1999). The negative suggestion effect: Pondering incorrect alternatives may be hazardous to your knowledge. *Journal of Educational Psychology*, *91*, 756-764.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, *20*, 633-642.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*, 354-380.

Cull, W. L., & Zechmeister, E. B. (1994). The learning ability paradox in adult metamemory research: Where are the metamemory differences between good and poor learners? *Memory & Cognition*, *22*, 249-257.

Delaney, P. F., Verkoeijen, P. P., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 53, pp. 63-147). New York: Academic Press.

Dunlosky, J., & Hertzog, C. (1997). Older and younger adults use a functionally identical algorithm to select items for restudy during multitrial learning. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *52*, 178-186.

Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language*, *33*, 545-565.

Dunlosky, J., & Thiede, K. W. (1998). What makes people study more? An evaluation of factors that affect self-paced study. *Acta Psychologica*, *98*, 37-56.

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology* (H.A. Ruger & C.E. Bussenius, Trans.). New York: Dover. (Original work published 1885).

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, *16*, 107-112.

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*, 290-296.

Hertzog, C., & Dunlosky, J. (2011). Metacognition in later adulthood: Spared monitoring can benefit older adults' self-regulation. *Current Directions in Psychological Science, 20*, 167-173.

Heslin, K.A., Curley, T.M., Jackiewicz, M.K., Flowers, C.S., Phelan, H.A. & Toppino, T.C. (2013, November). *Influence of feedback on metacognitive decisions about spacing practice tests: a framing effect?* Poster presented at the 55th annual meeting of the Psychonomic Society, Long Beach, California.

Izawa, C. (1966). Reinforcement-test sequences in paired-associate learning. *Psychological Reports, 18,* 879 –919.

Izawa, C. (1967). Function of test trials in paired-associate learning. *Journal of Experimental Psychology, 75*, 194-209.

Izawa, C. (1970). Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. *Journal of Experimental Psychology, 83,* 340-344.

Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General, 138,* 469-486.

Karpicke, J. D., Butler, A. C., & Roediger III, H. L. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory*, *17*, 471-479.

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, *319*, 966-968.

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, *133*, 643-656.

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General*, *135*, 36-69.

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review*, *14*, 219-224.

Kornell, N., & Bjork, R. A. (2009). A stability bias in human memory: Overestimating remembering and underestimating learning. *Journal of Experimental Psychology: General*, *138*, 449-468.

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

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 609-622.

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

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. *Practical Aspects of Memory*, *1*, 625-632.

LaVan, M.H., Pagano, M.J., & Toppino, T.C. (2011, November). *Metacognitive control of spacing and testing during learning: Are they related?* Poster presented at the 52nd annual meeting of the Psychonomic Society, Seattle, Washington.

Mazzoni, G., & Cornoldi, C. (1993). Strategies in study time allocation: Why is study time sometimes not effective? *Journal of Experimental Psychology: General*, *122*, 47-60.

Mazzoni, G., Cornoldi, C., & Marchitelli, G. (1990). Do memorability ratings affect study-time allocation? *Memory & Cognition*, *18*, 196-204.

Mazzoni, G., Cornoldi, C., Tomat, L., & Vecchi, T. (1997). Remembering the grocery shopping list: A study on metacognitive biases. *Applied Cognitive Psychology, 11*, 253-267.

McDaniel, M. A., & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology*, *16*, 192-201.

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

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*, 530-542.

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language*, *52*, 463-477.

Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science, 5*, 207-213.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 676-686.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South

Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*, 402-407.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G.H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–141). New York: Academic Press.

Nickerson, R. S. (1984). Retrieval inhibition from part-set cuing: A persisting enigma in memory research. *Memory & Cognition*, *12*, 531-552.

Pagano, M.J. & Toppino, T.C. (2013, November). *Efficacy of metacognitive decisions about practice tests: Spacing and feedback*. Poster presented at the 54th annual meeting of the Psychonomic Society, Toronto, Canada.

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31,* 3-9.

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

Roediger, H. L., & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1155-1159.

Son, L. K. (2004). Spacing one's study: Evidence for a metacognitive control strategy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 601-604.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and*

Cognition, *26*, 204-221.

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.

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*, 1024-1037.

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

Toppino, T. C., & Cohen, M. S. (2010). Metacognitive control and spaced practice: Clarifying what people do and why. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1480-1491.

Toppino, T. C., Cohen, M. S., Davis, M. L., & Moors, A. C. (2009). Metacognitive control over the distribution of practice: When is spacing preferred? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1352-1358.

Toppino, T. C., & Gerbier, E. (2014). About practice: repetition, spacing, and abstraction. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 60, pp. 113-189). New York: Academic Press.

Toppino, T. C., & Luipersbeck, S. M. (1993). Generality of the negative suggestion effect in objective tests. *The Journal of Educational Research*, *86*, 357-362.

Vincent, S. B. (1912). The function of the vibrissae in the behavior of the white rat. *Behavior Monographs*, 1 (No. 5).

Wheeler, M., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory*, *11*, 571-580.

Winer, B. J. (1971). *Statistical principles in experimental design.* New York: McGraw-Hill.

Zechmeister, E. B., & Shaughnessy, J. J. (1980). When you know that you know and when you think that you know but you don't. *Bulletin of the Psychonomic Society*, *15*, 41-44.