An Examination of Note Review and the Testing Effect on Test Performance

Vivian Song

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2018

ABSTRACT

An Examination of Note Review and the Testing Effect on Test Performance

Vivian Song

Traditionally, classroom testing is utilized and viewed as a way to measure students' knowledge of material. However, research has shown that test taking also enhances long-term learning and retention of material, a phenomenon known as the *testing effect*. Across settings, research has found that compared to rereading or repeated reviewing, repeated testing leads to poorer performance on immediate tests, but stronger long-term learning of material on delayed tests. These results have been produced with various materials, such as prose passages, word-pair associates, and educational materials such as textbook chapters. However, the testing effect has not been examined in relation to student-generated materials, such as lecture notes. Lecture notetaking is widely embraced in postsecondary education. Both taking and reviewing notes have significant benefits on students' academic and test performance. However, it is a complex cognitive task, which often results in students taking poor or incomplete notes and thus, limiting the benefits of notetaking and note review. There are many interventions to support students in taking better notes, but there is limited research on the effectiveness of the types of strategies used to review notes. This dissertation examined the effects of different note reviewing strategies on test performance: repeated review, self-testing, and rewriting.

In two experiments, 69 and 117 undergraduate students watched a recorded lecture while taking notes. Students then studied the notes through the use of repeated review (reread), self-testing (repeated recall), or rewriting before taking either an immediate or delayed final multiple-choice test on the materials. The independent variables included study method (repeated review/reread vs. self-testing/repeated recall vs. rewriting) and time of test (immediate vs.

delayed). The delayed variables included total test score, memory item performance, and inference item performance. Due to attrition in participants in Study 1, only study method was analyzed.

Results of these studies did not find a testing effect. There was only a significant main effect of study method on the total test and inference items in Study 1, in which the repeated review group performed significantly better on the immediate test than the self-testing and the rewriting groups. There was no significant main effect of study method for Study 2. Instead, there was a significant main effect of time across the three dependent variables. Students performed significantly better on the immediate test than the delayed test. There was no significant study method x time of test interaction.

These studies also examined whether quality and quantity of students' notes had an effect on test performance. Three covariates were examined: note themes, number of propositions, and number of main ideas. In Study 1, number of propositions and number of main ideas were significantly related to all dependent variables. In Study 2, the results were mixed. Number of propositions and main ideas were significantly related to total test performance and memory items, but not inference items. However, for number of main ideas, there was a trend that approached conventional significance for inference items.

Results also examined the effects of the notes taken during the study trials on test performance. In Study 1, the number of propositions recalled by students in the self-testing group was predictive of performance only on the total test score. The number of main ideas and propositions generated by students in the rewriting group were not significantly related to test performance. Results were similarly mixed in Study 2. Number of propositions and main ideas recalled by students in the self-testing group were not significantly related to test performance.

In contrast, number of main ideas included in students' notes in the rewriting group was related to performance on memory items and the total test items. Future research should continue to explore the testing effect in conjunction with note taking.

TABLE OF CONTENTS

LIST OF TABLES

Acknowledgments

Looking back on my five years at Teacher College, it is still amazing to me that I have gotten this far. This would not have been possible without many key figures. The first is Dr. Stephen Peverly, who has been a wealth of knowledge, wisdom, and support. I am so grateful for the time and effort spent not only reading and giving feedback on my manuscript drafts, but also extending words of support and guidance whenever I hit a bump on the road in the program. I also like to extend my deepest appreciation to my program faculty member, Dr. Marla Brassard. I am so thankful for all of the honest and constructive feedback over the years. Thank you both for the unwavering support and belief in my abilities. I am also grateful to Drs. Sumowski, Corter, Son, and Farber for their time and participation on my dissertation committee.

I would also like to thank the individuals who were already there from the start of my journey, to those who have been there at the end. Thank you to Hans for being there every step of the way since college. I am also appreciative towards Missy, who has been a constant source of support and fun times in NYC. I would also like to thank my cohort of interns and externs at JCCA, who have been the most supportive group in regards to everything from dissertations to difficult clients. I am fortunate to have made new friends at the end of the tunnel, Molly and Samantha, who have listened to and read endless streams of conversations and texts about my dissertation.

Lastly, I am thankful for the support of the individuals who have really been there every step of the way during my life: my family. I am grateful to my brother, Michael for always lending an ear to listen. Thank you mom and dad for providing endless amounts of encouragement and support throughout my life, and always considering my wellbeing.

CHAPTER I: INTRODUCTION

Since lecturing is the predominant form of instruction in U.S. schools from middle school to college, it is not surprising that notetaking has been widely embraced by students as a useful strategy to facilitate understanding, as well as comprehension, of lecture content (Armbruster, 2009). Notetaking serves two functions to students: *encoding* (taking notes), and *external storage* (reviewing notes). Encoding encourages learning by requiring students to pay attention to the lecture, in addition to concurrently generating connections within the presented information, and connecting it with prior knowledge. External storage facilitates further cognitive processing of the information for students, which helps long-term retention and leads to relearning of forgotten material (DiVesta & Gray, 1972). Both encoding and external storage have been shown to contribute to increased academic outcomes, (e.g. Armbruster, 2009; Kiewra, 1985; Kobayashi, 2006; Peverly, Brobst, Graham, & Shaw, 2003; Peverly et al., 2007); indicating that it is an effective study strategy.

Research has sought to determine which of the two functions has the greatest impact on student academic outcomes. The literature indicates that the external storage function has proven to be more beneficial. Kobayashi (2005) found a positive, but moderate effect size for encoding (.22) in a meta-analysis that compared students who took notes with those who did not. In the same meta-analysis, when he compared students who took and reviewed notes with those who did not, and/or those who were able to only mentally review notes, he found a much larger effect size (.75) for external storage. Students who take and review their notes are more likely to perform better on exams than those who do not, indicating the extra benefits of reviewing in addition to taking notes.

However, notetaking is a cognitively demanding task, as students need to attend to the lecture, select important ideas, manipulate and interpret information, and then write down the information (Piolat, Olive & Kellogg, 2005). The literature indicates that individual differences in cognitive variables such as handwriting speed, sustained attention, and language comprehension can play an important role in note-taking abilities (Armbruster, 2009; Peverly et al., 2007; Peverly, Garner & Vekaria, 2014; Peverly, et al., 2013).

Thus, given the cognitive complexity of notetaking, it is not a surprise that the literature shows students often take poor notes (e.g. Armbruster, 2009; Kiewra et al., 1989, Kiewra, Benton, & Lewis, 1987). To compensate for the difficulty students may experience when taking notes, instructors can provide supports for notetaking to help them fully integrate and process lecture information during encoding. Two types of support, skeletal outlines and guided notes, provide cues to help students extract key information from the lecture. Skeletal or linear outlines consist of headings for the main ideas and subtopics, as well as space for students to record pertinent information. Studies have shown that use of skeletal outlines can improve note quality (Kiewra, Benton, Christensen, Kim & Risch, 1995), as well as performance on a multiple-choice test (Cohn, Cohn, & Bradley, 1995; Peverly et al., 2013).

Guided notes, which are similar to skeletal outlines, provide additional cues such as lines, to indicate when, where, and how many key ideas students should record. Austin, Lee, and Carr (2004) compared the effects of traditional lectures, lectures with slides, and lectures with slides and guided notes on the quality of students' notetaking. They found that guided notes improved the quality of students' notes. In terms of effects on students' test performance, Cornelius (2008) compared test outcomes between those who received partial notes and those who received complete notes. Results showed that students who received partial notes performed significantly better on tests later in the semester.

Instructors can also provide supports that help students extract and reorganize key ideas from lectures. One such example is matrix notes. Matrix notes present topics horizontally and subtopics vertically, creating a grid for students to take notes that connect the intersecting topic and subtopic at each cell. The evidence on matrix notes is mixed. Kiewra et al. (1989) found that students who used a matrix framework recorded significantly more lecture ideas than those who took conventional notes (47% of lecture ideas recorded vs. 32%), and Kiewra et al. (1991) found that matrix notetaking was more effective than an outline or conventional notetaking when measured by performance on a cued-recall test of lecture content. However, Kiewra et al. (1995) found the opposite results; skeletal outlines lead to better performance on tests of recall and relational learning than matrix and conventional notetaking.

Lastly, instructors can also provide students with complete notes to use during the lecture. While it would make sense that having access to the instructor's complete notes would lead to higher performance outcomes, the literature suggests that this is not the case, with results indicating there is no significant difference in performance between those who receive partial notes and those who receive full notes (Cornelius, 2008; Neef, McCord, & Ferreri, 2006). It may be that receiving instructors' complete notes reduces and interferes with student notetaking during the lecture, which restricts encoding and understanding.

While research has examined the effects of reviewing notes on students' academic performance in regards to their completeness and the source of the notes, there is less literature on the effects of strategies used to review notes and other text material. Graphic organizers are one such way to improve review; they present relationships among concepts by using the relative locations of topics and subtopics. Katayama and Robinson (2000) found that students who studied partially filled graphic organizers performed better than students who studied partially filled outlines on an application test.

To investigate students' real-world study behaviors, Karpicke, Butler, and Roediger (2009) surveyed 177 college students on the types of strategies they use for studying. They found that the most common study strategy was rereading text material, with 86% of students listing it as a study strategy, and 55% of students indicating it as their number one strategy. However, it is also ineffective, as research has shown that rereading prose passages produces limited benefits (Callender & McDaniel, 2009). Passive studying, such as rereading or highlighting the text, results in more shallow understanding of the material, with the goal to only memorize and produce information needed on a quiz or test. Active studying, such as rewriting notes or self-testing, allows for deeper processing, which results in better understanding and long-term retention of the material (Marton and Saljo, 1976; Tomes, 2011).

Self-testing in particular has been shown to be a powerful study strategy. There have been numerous studies on how retrieval of information from long-term memory results in better retention than merely rereading the text, a phenomenon called the *testing effect* (Roediger & Butler, 2010). Roediger and Karpicke (2006b) demonstrated the testing effect in one of their classic studies using prose passages and free recall tests with college students. There were two conditions: a study only group (SS) and a single-test group (ST). The study group (SS) was given two chances to study or repeatedly review the passage before taking a final recall test, while the testing group (ST) studied the passage once and then recalled the passage once before taking a final recall test. During the repeated review sessions, subjects had seven minutes to read and study the passage, whereas during the intermittent testing sessions, subjects were given a test sheet with the passage title on top and had seven minutes to recall as much information from the passage as they could. All two groups either took a free recall test five minutes, two days, or one week later. Results show that while repeated review produced more benefits than repeated recall after a five-minute delay, the testing group had better retention of information both after two days (68% vs. 54%) and one week (56% vs. 42%).

Similar results were found when Roediger and Karpicke (2006b) conducted a follow up study on the effects of repeated testing, with three conditions: a study only group (SSSS), a single-test group (SSST), and a testing group (STTT). The study group (SSSS) was given four chances to study or review the passage before taking a final recall test, while the testing group received one chance to study the passage and three opportunities to repeatedly recall the passage before taking a final recall test (STTT). The third group, the single-test group, was allowed to study the passage three times and recall the passage once before taking a final recall test (SSST). During the study/review trials, subjects had five minutes to review the passages, whereas during the testing/recall trials, subjects received a blank piece of paper and had ten minutes to recall as much information from the passage as they could. All three groups took a free recall test five minutes or one week later. Results showed that after a five-minute delay, the study group (SSSS), the single-test group (SSST), and the testing group (STTT) remembered 83%, 78%, and 71% of the passage, respectively. Following a one-week delay, results indicated that the testing group (STTT) and the single-test group (SSST) remembered the most (61% and 56% of the passage), while the study group (SSSS) only recalled 40% of the passage. These results suggest that testing helps promote long-term retention of material better than merely restudying.

The literature on the testing effect has predominantly focused on the use of paired associates or text passages. However, studying lecture notes is one of the most common of all study strategies for students. There has been less research comparing the efficacy of the testing effect to repeated review of notes on test performance. This study aims to compare the efficacy of different approaches to studying notes on test performance: repeated review, repeated recall (self-testing), or rewriting.

CHAPTER II: REVIEW OF THE LITERATURE

Lecture Notetaking

In higher education, the predominant form of instruction is lecturing. According to a 2001 National Center of Education report, 83% of postsecondary instructional faculty reported that lecturing was one of their primary instructional methods for at least one of their classes (Wirt et al., 2001). Armbruster (2009) estimated that undergraduate students spend 80% of their time in class listening to lectures. Given that the amount and complexity of information students are expected to process during lectures increase exponentially in college, research has historically shown that lecture notetaking is a widely accepted practice among college students (Williams & Eggert, 2002; Palmatier & Bennett 1974; Dunkel & Davy, 1989; Hartley & Davies, 1978). Research by Palmatier and Bennett (1974) reported that 99% of undergraduates take notes, while Williams and Eggert (2002) found that students take notes during class without explicitly being told to do so. Additionally, in an ongoing survey of undergraduates' notetaking behaviors in our lab, 93% of students indicated that they often or always take notes in class.

Both professors and students share a similar positive perception of the value of notetaking. In a survey evaluating professors' opinions on notetaking, Landrum (2010) reported that 83% of the professors expected their students to take notes. Palmatier and Bennett (1974) found that 96% of undergraduates felt that notetaking was essential to performing well in college. Dunkel and Davy (1989) indicated that 94% of U.S. students and 92% international students viewed notetaking as an important activity. There are many benefits of notetaking: students reported they took notes because it helped them remember, understand, and review lecture content, in addition to pay attention in class and prepare for exams.

Notetaking serves two functions: *encoding* (taking notes), and *external storage* (reviewing notes). Encoding is a cognitively demanding task, as students need to pay attention, select important ideas from the lecture, manipulate and interpret information, and then write down the information, all within the limits of working memory (Piolat, Olive, & Kellogg, 2005). However, encoding is beneficial in that it assists in learning of lecture material (Armbruster, 2009). External storage facilitates further cognitive processing of the information for students after the act of notetaking, which helps long-term retention and leads to relearning of forgotten material (DiVesta & Gray, 1972). There is much more research on the encoding function than the external storage function, although as we shall see, the latter is much more important than the former.

The literature on the relationship between taking lecture notes and test and academic performance is extensive. The encoding function is commonly measured by comparing the performance of students who listen to the lecture and take notes, to those who listen and do not take notes. Neither group is permitted to review their notes after the lecture (DiVesta & Gray, 1972).

Results have been mixed; however, they typically indicate that there is a positive, albeit modest effect of encoding on test performance (DiVesta & Gray, 1972; Kiewra, 1989; Fisher & Harris, 1974; Kobayashi, 2005). Hartley and Davies (1978) conducted a meta-analysis of 35 studies, with results showing that 17 studies found beneficial encoding effects, 16 studies indicated there was no difference in performance between notetakers and listeners, and two studies showed listening without notetaking was more facilitative of performance than taking notes. Kiewra's (1985a) meta-analysis of 56 studies found similar results, with 38 studies favoring notetaking, 21 indicating no significant differences, and two finding notetaking to be harmful. Additionally, Kobayashi (2005) found a positive, but moderate effect size for encoding (.22) in a meta-analysis across 131 independent samples that compared students who took notes with those who did not.

The value of encoding on recall of information specifically has been demonstrated through various studies. Kiewra and Fletcher (1984) reported a correlation of .72 between the number of points recorded in notes and immediate cued recall performance. Similar results have also been found with immediate and delayed free recall. Howe (1970a) found that information had a 34% chance of being recalled on a delayed free recall test if they were included in notes, but only 5% if they were not. Similarly, in a study using passages, Bretzing and Kulhavy (1981) found that idea units were 58% more likely to be recalled immediately if they were included in participants' notes, and only 15% if they were not.

The relationship between the quantity, as well as quality, of notes and academic performance in natural classroom conditions has also been investigated. A correlation of .97 was found between the total number of words in notes and exam performance for the top performing students (Nye et al., 1984), in addition to a correlation of .72 between the number of points recorded in notes and immediate cued recall performance (Kiewra & Fletcher, 1984). Kiewra (1985b) found correlations of .61 between the amount of information in lecture notes and exam performance, and .78 with solely lecture-related test items. Similarly, other studies have found correlations of +.20 to +.60 between the amount of lecture material included in students' notes and their exam performance (Hartley & Davies, 1978), and shown that notes with fewer words are negatively correlated with academic performance (Fisher & Harris, 1973; Kiewra & Frank, 1988). These results have been replicated across various exam formats, such as immediate and delayed multiple-choice tests (Baker & Lombardi, 1985; E. Cohn et al., 1995); cued recall tests

(Kiewra et al., 1991), free recall tests (O'Donnell & Danserau, 1993), and compare-contrast essays (Benton et al., 1993). Taken together, these results suggest that extensive notes are more conducive to recall and academic performance.

Research also has indicated that the type of information included in notes, especially main ideas, is predictive of academic achievement (Kiewra, 1987; Kiewra & Benton, 1988; Kiewra & Frank, 1988). Kiewra and Fletcher (1984) demonstrated that lecture notes emphasizing main points rather than details lead to increased performance on immediate and delayed test items regarding main ideas, details, and integration of ideas. Similarly, Baker and Lombardi (1985) found that the content of students' lecture notes was related to subsequent test performance, with the more information they included of a certain type (e.g. main points, details) in their notes, the better they performed on corresponding test questions. Students rarely answered questions incorrectly as long as the information were present in notes. A study conducted by Peverly et al. (2007) scored students' notes on lecture content for quality and quantity, which were highly correlated with each other (.93). Note quality was a significant predictor of performance on a written recall task, with a correlation of .37 between note quality and recall quality.

Overall, the encoding literature indicates that the act of taking notes facilitates the recall of information, with both quality and quantity of notes impacting students' academic performance. However, students are generally poor note takers, which can make it more difficult for them to acquire the benefits from encoding (e.g. Armbruster, 2009; Kiewra et al., 1989, Kiewra, Benton & Lewis, 1987). Locke (1977) found that students included only 60% of relevant lecture material in their notes, whereas other studies demonstrated that college students only record about 20% to 40% of lecture material (Kiewra, 1985c; Kiewra et al., 1987; Kiewra, DuBois, Christian & McShane, 1988; O'Donnell & Danserau, 1993). While this may partially be due to students choosing to record only material they feel is relevant, this may also be partially accounted for by the fact that notetaking is cognitively complex, and requires the use of higher-level cognitive resources such as handwriting speed, sustained attention, and language comprehension. Weaknesses in any of these processes may be related to difficulty in taking good quality notes (Armbruster, 2009; Peverly et al., 2007; Peverly, Garner & Vekaria, 2014; Peverly, et al., 2013). Additionally, a major component of notetaking is generative processing, as to take good notes; students must actively generate connections within the presented information, as well as with prior knowledge in order to encourage learning (Benton, Kiewra, Whitfill & Dennison, 1991; Suritsky & Hughes, 1991; Wittrock, 1990). This is best done by paraphrasing and interpreting the lecture in one's own words, in order to facilitate deeper processing and understanding (Kiewra, 1985a). However, since some students mistakenly assume that the act of simply taking notes results in deeper processing, they transcribe the lecture information verbatim, possibly leading to rote memorization of the lecturer's words and terminology, and undermining comprehension (Williams & Eggert, 2002).

As referenced earlier, the literature on the external storage function of note taking is much more limited, as compared to encoding. However, review of notes is a common study strategy, as indicated by the results of a study conducted by Hartley and Davies (1978), which found that 98% of U.S. students and 86% of international students take notes so that they can review them for exams. Research has shown that this study strategy is effective, as students who review their notes, typically perform better than those who do not (Kiewra, 1985a).

The external storage function is commonly investigated by having subjects listen to a lecture and either take notes or simply listen, and then compare the performance of those who reviewed their own notes, reviewed no notes, or reviewed provided notes. Results indicate there is a strong effect for review, one that is much more significant than those for encoding. Kiewra (1985a) reviewed 22 studies, and found that 77% showed a positive effect for review, 23% indicated no significant differences, and none demonstrated a negative effect of reviewing notes. Similarly, Kobayashi (2006) found a large effect size of .75 in favor of review when he compared students who took and reviewed notes with those who did not, and/or those who were able to only mentally review notes. Students who review their own notes are more likely to perform better on exams than those who do not, indicating the extra benefits of reviewing in addition to taking notes.

Taken together, research supports the value of both encoding and external review in benefitting students' recall and academic performance. However, there is no guarantee that these facilitative processes will occur. Research indicates that students typically take notes that are transcribed verbatim, as well as incomplete (e.g. Williams & Eggert, 2000; Kiewra, 1989), which affects the quality of notes and their review, as well as students' subsequent academic performance. Additionally, there is little research on how to effectively review notes (Kiewra, 1987). Thus, it is important to consider how educators can help improve how students take notes and how they review them. The following sections describe interventions to help improve note quality and review.

Notetaking Interventions

Given the research that indicates students are typically poor notetakers, it is important to consider how educators can help improve students' note quality. Poor notes can have implications for both encoding and external review; insufficient encoding affects students' ability to generate connections with the lecture material, which subsequently affects the quality of notes used for review and test performance.

To compensate for the difficulty students may experience when taking notes, instructors can provide supports for notetaking to help them fully integrate and process lecture information during encoding. One method is to provide students with handouts that provide cues to help them extract key information from lecture. Handouts typically fall into two categories: skeletal outlines and guided notes. Skeletal or linear outlines consist of headings for the main ideas and subtopics, as well as space for students to record pertinent information. Studies have shown that outlines can improve note quality as well as recall (Kiewra et al.,, 1995; Peverly et al., 2013). Cohn et al., (1995) also found that students who took notes in an outline form recorded significantly more complete notes than those who took conventional notes, and subsequently performed better on a multiple-choice exam.

Guided notes, which are similar to skeletal outlines, provide additional cues such as lines, to indicate when, where, and how many key ideas students should record. Austin, Lee, and Carr (2004) investigated students' note quality under three presentation conditions: 1) traditional lecture, 2) lecture with slides, and 3) lecture with slides and guided notes. The authors reported that students in the guided notes condition recorded significantly more critical points and examples than those in the traditional lecture and lecture with slides conditions. For example, students in the guided notes condition recorded 100% of the critical points, compared to only 62% in the traditional lecture condition. Guided notes have also been found to have a significant effect on test and academic performance. Cornelius (2008) compared test and final course outcomes between students who received partial notes and those who received complete notes.

Results demonstrated that those in the partial notes condition achieved better test performance later in the semester (including on the final exam) and better overall course grades.

Instructors can also provide supports that help students extract and reorganize key ideas from lectures. Two such types are matrix notes and graphic organizers. Matrix notes present topics horizontally and subtopics vertically, creating a grid for students to take notes that connect the intersecting topic and subtopic at each cell. Graphic organizers present relationships among concepts by using the relative locations of topics and subtopics. Since most of the literature with graphic organizers examines its effects on recalling or studying text, rather than lecture content, it will be referenced later when reviewing external storage interventions.

The evidence on matrix notes is mixed in regards to note quality and test performance. Kiewra et al. (1989) found that students who used a matrix framework recorded significantly more lecture ideas than those who took conventional notes (47% of lecture ideas recorded vs. 32%), and Kiewra et al. (1991) found that matrix notetaking was more effective than an outline or conventional notetaking when measured by a cued-recall test of lecture content. However, Kiewra et al. (1989) demonstrated that students using skeletal outlines to take notes recorded more idea units than those using matrix notes and significantly more than those taking conventional notes. Kiewra et al. (1995) found similar results when comparing the effects of conventional notes, outlines, and matrix notes. Those who took outline notes recorded more complete notes than those who recorded matrix notes and conventional notes. As for test performance, Kiewra et al. (1995) demonstrated that skeletal outlines lead to better performance on tests of recall and relational learning than matrix and conventional notetaking.

Lastly, instructors can also provide students with complete notes. Full notes include all of the pertinent main ideas as well as supporting details. While it would make sense that having access to the instructor's complete notes would lead to higher performance outcomes, the literature suggests that this is not the case. Morgan, Lilley, and Boreham (1988) examined the effects of the amount of detail provided in notes on student notetaking and test performance, with four conditions: 1) headings and full text; 2) headings and key points; 3) headings only; and 4) no supplementary materials. Students in the headings only condition recalled more lecture material after two days than those in the headings and full text condition, as well as heading and key points condition. While a delayed recall test two weeks later demonstrated those in the full text condition performed better, they only performed marginally better than those in the partial text conditions. Results also showed that the amount of notes student recorded was inversely related to the amount of details given on the handout, suggesting that those who received partial notes transcribed more of their own notes. Similarly, other studies indicate there is no significant difference in performance between those who receive partial notes and those who receive full notes (Cornelius, 2008; Neef et al., 2006). It may be that receiving complete notes during the lecture reduces student notetaking of their own personal notes, which restricts encoding and understanding.

Note Review Interventions

While the literature supports the value of providing certain supplemental notes to facilitate students' encoding abilities during lecture, there are more mixed results in regards to the efficiency of reviewing instructor notes. Under immediate review and testing conditions, reviewing full instructor notes appeared to interfere with recall (Fisher & Harris, 1973). Kiewra (1984) also demonstrated that students who took and reviewed personal notes performed better on an immediate recognition test (93% correct) than those who took notes but reviewed instructor notes (71% correct), and those who only listened to the lecture and reviewed instructor notes (79% correct). These results suggest that reviewing instructor's notes after listening and recording personal notes is more of a hindrance to immediate recall than simply listening to the lecture and then reviewing instructor's notes. Kiewra suggested that reviewing instructor notes after taking their own notes interfered with students' initial processing of the lecture.

In contrast, under delayed conditions, reviewing full instructor notes appears to have more benefits for recall of factual knowledge than studying personal notes. The literature suggests that by having an adequate review period prior to a delayed exam, this leads to higher achievement for students who reviewed provided notes compared to those who reviewed personal notes (Annis & Davis, 1975; Fisher & Harris, 1973; Kiewra, 1985d; Masqud, 1980; Thomas, 1978). When Kiewra (1985b) compared the performance of students who reviewed provided notes to those who reviewed personal notes on a delayed test, the former achieved significantly better. In this experiment, students were given 25 minutes to review the notes, as opposed to 10 minutes, the amount of time used in studies that found no significant differences between reviewing provided notes and personal notes. Kiewra suggested that the effectiveness of the provided notes was due to the longer review period, which allowed for a lengthy delay between acquisition and review. Allowing for a longer delay between encoding and review reduced the saliency of the original acquisition cues during the lecture, so that what was learned during the recording of notes did not interfere with what was learned later during review. Additionally, instructor notes are typically more complete in breadth and organization, which may have helped improve the efficacy of the notes and subsequent review.

However, when Kiewra examined the results on tests of higher order learning (e.g. application, analysis, problem solving, synthesis), there were no significant differences in performance between personal notes and provided notes conditions. The author posited that this was due to lack of generative processing in both groups: the instructors did not include internal connections within the provided notes, and the students did not generate internal or external connections in their own notes. In his review of the literature, Kiewra (1985c) also concluded that studying instructor notes and personal notes are more efficacious than only studying one or the other; reviewing both sets of notes not only provides positive generative learning, but also improves accuracy and completeness of information. This may also reduce the possible interfering effects that appeared when students took personal notes but reviewed instructor provided notes.

There is also literature comparing the benefits of reviewing different types of notes. Kiewra, DuBois, Christian, and McShane (1988) had college students view a lecture without taking notes, and then a week later, they were placed in one of four conditions: 1) mental review (no notes), 2) complete notes (complete transcript of the lecture), 3) notes in outline form, and 4) notes in matrix form. Results of cued recall, recognition, and transfer (synthesis and application) tests demonstrated the importance of the external storage function, as reviewing any form of notes was better than reviewing the material mentally, without any form of notes. Additionally, there were also differences in performance amongst the note review conditions within test type. Outline and matrix notes produced higher recall than complete notes, while matrix notes alone facilitated higher transfer performance. The authors suggested that outline and matrix notes lead to greater processing of internal connections and improved recall, while matrix notes lead to a more integrative understanding of the information, leading to better performance of the synthesis and application test. Lastly, all three note review groups performed similarly on the recognition test, most likely due to the items being less affected by forming internal connections.

Other studies have also demonstrated that review of different types of notes impacts learning. Results have demonstrated that when students are provided with graphic organizers to study along with when reading text, they perform better on tests measuring concept relation knowledge (e.g. Kiewra et al., 1988; Robinson & Kiewra, 1995). Katayama and Robinson (2000) also found that students who studied partially filled graphic organizers performed better than students who studied partially filled outlines on an application test. Benton et al. (1993) demonstrated similar effects where participants viewed a lecture and were then required to write compare and contrast essays a week later. Participants were placed in one of three note review conditions: 1) conventional, 2) outline, and 3) matrix. Half of the participants were allowed to review their notes while writing the essay, while the other half did not use any study material. Results showed that those who had use of their notes wrote more cohesive and more coherent essays. Within the note review groups, students in the outline notes and matrix notes conditions, compared to the conventional note condition, included more text units in their essays. Students in the matrix notes condition also wrote more coherent essays. These results suggest that reviewing notes that help generate internal and external connections are more conducive for performance across application tests and essays.

Study Strategies

While research has examined the effects of the completeness of notes, source of notes, and types of notes on test performance, there is less research on the effects of strategies on *how* to review notes and other text material. Broadly, students' study strategies can be organized into two categories: passive and active. Passive studying, such as rereading or highlighting the text, results in more shallow understanding of the material and limited benefits, with the goal to memorize and produce information needed on a quiz or test (Marton and Saljo, 1976; Tomes, 2011). Active studying, such as self-testing, allows for deeper processing, which results in better understanding and long-term retention of the material (Marton and Saljo, 1976; Tomes, 2011). Strategies can range from creating flashcards, mnemonic devices, or rewriting notes.

In an investigation on students' real-world study behaviors, Karpicke, Butler, and Roediger (2009) surveyed 177 college students on the types of strategies they used for studying. Some of the most common strategies included doing practice problems (43%), using flashcards (42%) and rewriting notes (29%). However, the most common study strategy was rereading text material, with 84% of students listing it as a study strategy, and 55% of students indicating it as their number one strategy. This is consistent with the literature, which demonstrates that one of the most commonly reported study methods by students is rereading their textbook (Carrier, 2003; Goetz & Palmer, 1991).

Despite the prevalence of this strategy, it has proven to be ineffective, as research has shown that rereading prose passages produces limited benefits (Callender & McDaniel, 2009; McDaniel & Callender, 2008). In fact, Tomes (2011) found that passive reading was negatively associated with test grades, final course grades, and overall GPA. Callender and McDaniel (2009) conducted an extensive series of experiments with lengthy educational texts (approximately 2000 words) to determine the effects of rereading on test outcomes. In Experiment 1, participants were randomly assigned to two conditions: one where they read the text once, and the other where they read the text twice. Afterwards, they were administered a summative assessment of 22 multiple-choice questions and four short-answer questions. Results indicated that there were no significant benefits to rereading the text twice. In Experiment 2, participants at a more selective institution were given a more unfamiliar text to reread, with the thought that more familiar texts could have a positive effect on outcomes. Results showed a small effect for rereading on multiple-choice items, with the authors positing that rereading might be more valuable with an unfamiliar text. However, in Experiment 3, in which the authors tested their theory by mixing familiar and unfamiliar texts, there were no significant effects for rereading across all texts. These results suggest that rereading has limited effects on test outcomes regardless of the tests' format, and with educational texts commonly found in classrooms.

Tomes (2011) found that students who engaged in active strategies achieved higher test grades, final course grades, and overall GPA. The two active strategies most associated with the higher outcomes were "creating/writing/processing study materials" and "quizzing and testing." The first strategy encompassed answering questions in the textbook, drawing diagrams and charts, writing out summaries, writing out notes, making a study guide, and creating study notes. The second ranged from recalling material, predicting test items, using flash cards, and quizzing by self or with a partner. Similarly, Wittrock (1990) demonstrated that college students were able to improve reading retention and comprehension by constructing verbal analogies or summary sentences in their own words of text material. Students may learn the best via active studying, because it allows them to reconstruct the presented material in their own words, and make connections within the information and with their prior knowledge (Wittrock, 1990).

Given the research in favor of the effects of reviewing notes on academic performance, it is important to consider what are the most efficacious means for students to review their notes. One such method of review is self-testing. The next section of this dissertation will thus focus on the testing effect, which is a form of self-testing.

Testing Effect

Traditionally, testing refers to the standardized measures used to assess students, whether it is what they learn in the classroom, or their knowledge and aptitude. From kindergarten to high school, testing is frequently used to evaluate students' progress. However, as students advance to higher education, tests are given less frequently. At universities, it is common for even the most basic courses to only include a midterm and a final exam. Across educational levels, it appears that both teachers and students have similarly negative viewpoints regarding testing, as students do not enjoy taking them, and teachers do not like grading them.

However, testing can do more than simply measure knowledge. It also has the ability to improve long-term retention of knowledge. For example, if students are tested on material and they successfully recognize or recall it, this improves their ability to retain it in the future. This phenomenon in which testing can lead to improved retention is called the testing effect.

The idea that testing improves retention is not new. Francis Bacon, as well as William James, both argued in favor of learning through self-testing (as cited in Roediger & Karpicke, 2006b, p.181). The effects of testing on learning were first investigated by two classic studies conducted by Arthur Gates and Herbert Spitzer. Gates (1917) tested groups of children across different grade levels (Grades 1, 3, 4, 5, 6, 8) on two different stimulus materials: nonsense syllables and brief biographies. The experiment was divided into two phases. In the first phase, the children studied the material themselves. In the second phase, the children were instructed to look away from the materials and attempt to recall the information, a form of self-testing (recitation phase). The children were allowed to look back at the materials if they needed to refresh their memories during the recitation phase. Gates also manipulated the amount of time the children spent reciting by instructing them when to stop reading (end of the first phase) and begin reciting (start of the second phase). Children at all grade levels spent 0, 20, 40, 60, 80, or 90% of the study period involved in recitation. At the end of the study period, Gates asked the children to write down as many items as they could remember in order of appearance. The children were then retested again three to four hours later. All of the children except the first graders showed strong effects of recitation with nonsense syllables, while all grade levels showed these same recitation effects with the biographies, albeit less strongly on the initial tests than on the delayed tests. It appeared that for the biographies, the optimal amount of time for recitation was 60% of the study period. Gates concluded that recalling information during learning is an efficacious way to study.

Spitzer (1939) conducted the second landmark study demonstrating the effects of testing. He tested the entire sixth grade population in nine Iowa cities, using 600 word articles that were similar to material they might study in school. Across 63 days, the students took tests according to varying schedules. Some students took a single test 63 days later, while others took earlier intermittent tests to see if they had an effect on later test performance. Each test consisted of 25 multiple-choice items with five answer choices. There were three main conclusions from the results: the longer the first test was delayed, the worst the test performance; second, giving an initial test before a final test appeared to delay forgetting, since their performance did not drop as much (and sometimes increased); and third, the earlier the initial test was given to students, the better they did on later tests. Spitzer suggested that this showed that an initial test must be given immediately after study, in order for the student to be better able to recall/recognize the material at some subsequent point in time.

Since these two classic studies, other researchers have added many variations to research on the testing effect. Beyond examining the testing effect with free recall, researchers also commonly test participants on paired associates or word lists through the use of cued recall tests. These results also demonstrate that repeated testing leads to greater benefits than repeated studying (Allen, Mahler & Estes, 1969; Izawa, 1966, 1967, 1970; Jacoby, 1978).

Given the research illustrating the benefits of repeated testing, what is the best sequence to promote long-term retention? Paired-associates have been used to investigate if there is an optimal testing schedule for learning material. The literature indicates that spaced retrieval practice is more effective than massed retrieval practice for long-term retention (Melton, 1970; Cepeda, Pashler, Vul, Wixted & Rohrer, 2006). Additionally, beyond the laboratory, frequent testing has also been shown to encourage students to study consistently and space out their studying throughout the semester, rather than simply mass studying before an examination (Bangert-Drowns, Kulik, & Kulik, 1991; Roediger & Karpicke, 2006a). Kornell and Bjork (2007) and Hartwig and Dunlosky (2012) surveyed college students on their beliefs and strategies in managing their studying, and demonstrated that students do not study continuously in the absence of an evaluation. Both found that students' study schedules were driven by specific events (e.g. an upcoming exam), rather than systematic studying, with 59% (Kornell & Bjork, 2007) and 56% (Hartwig & Dunlosky, 2012) of students reporting they decide what to study next based on "whatever's due soonest/overdue." This was found to have an adverse effect on GPA. Hartwig and Dunlosky (2012) demonstrated that lower achievers not only never planned their studying times, but also studied later at night and focused on impending deadlines, compared to the higher achievers. Students who scheduled their studying time in a spaced manner also used significantly more study strategies, which can also improve academic success. Research has demonstrated that massed studying is only effective for immediate tests, and thus is poor for delayed tests (Roediger, Finn, & Weinstein, 2012).

However, the literature is more mixed when comparing the effects of spaced retrieval practice and expanding retrieval practice. In spaced retrieval practice, the intervals between studying and testing are equally spaced; while in expanding retrieval practice, the intervals gradually expand with each subsequent testing trial (e.g. the first retrieval occurs one minute post studying, the second retrieval occurs five minutes post studying, the third retrieval occurs 10 minutes post studying etc.). In an experiment conducted by Landauer and Bjork (1978) that compared expanding retrieval practice and spaced retrieval practice in paired-associate learning, they found that the former produced better recall on the final test. However, it is of note that the difference was not significant. Most subsequent studies have found that there is typically no difference in performance on the final test between the two schedules (Balota, Duchek & Logan, 2007; Balota, Duchek, Sergent-Marshall, & Roediger, 2006; Carpenter & DeLosh, 2005). There have been some exceptions; for example, Karpicke and Roediger (2006a) similarly compared the results of massed practice, expanding retrieval practice, and spaced retrieval practice on recall for paired associates, and found that on the final test given 10 minutes after the study session, results were moderately in favor of expanding retrieval practice. However, on a delayed test 48 hours later, spaced retrieval practice instead produced better recall. The authors propose that the mixed results are dependent on the placement of the first retrieval attempt. The literature suggests that longer delays between studying and the initial retrieval attempt lead to greater difficulty and better later long-term retention. Thus, the expanding retrieval schedule may not be as beneficial since the initial recall is typically immediately after the learning session.

The benefits of the testing effect have also been generalized to research involving educationally relevant materials, such as U.S History course material and other factual information (Carpenter, Pashler, & Cepeda, 2009; Butler & Roedieger, 2007; McDaniel &

Fisher, 1991), long chapters (Butler et al., 2006; Kang et al., 2007; Weinstein, McDermott & Roediger, 2010) and short prose (Agarwal et al., 2008; Roediger & Karpicke, 2006a). Roediger and Karpicke (2006b), for example, used short prose passages and free recall tests in their classic study with college students. There were two conditions: a study only group (SS) and a single-test group (ST). The study group (SS) was given two chances to study the passage before taking a final recall test, while the testing group (ST) received only one chance to study the passage and one opportunity to recall the passage before taking a final recall test. During the study sessions, subjects had seven minutes to read the passage, whereas during the testing sessions, subjects were given a test sheet with the passage title on top and also had seven minutes to recall as much information from the passage as they could. All two groups either took a free recall test five minutes, two days, or one week later. Results show that while restudying produced more benefits than testing after a five-minute delay, the testing group had better recall of information both after two days (68% vs. 54%) and one week (56% vs. 42%).

Roediger and Karpicke (2006b) found similar results when they conducted a second experiment to focus on the effects of repeated testing. There were three conditions: a study only group (SSSS), a single-test group (SSST), and a testing group (STTT). The study group (SSSS) was given four chances to study the passage before taking a final recall test, while the testing group received one chance to study the passage and three opportunities to recall the passage before taking a final recall test (STTT). The third group, the single-test group, was allowed to study the passage three times and one opportunity to recall the passage before taking a final recall test (SSST). During the study sessions, subjects had five minutes to review the passages, whereas during the testing sessions, subjects received a blank piece of paper and had ten minutes to recall as much information from the passage as they could. All three groups took a free recall test five minutes or one week later. Results showed that after a five-minute delay, the study only group (SSSS), the single-test group (SSST), and the testing group (STTT) remembered 83%, 78%, and 71% of the passage, respectively. Following a one-week delay, results indicated that the testing group (STTT) and the single-test group (SSST) remembered the most (61% and 56% of the passage), while the study group (SSSS) only recalled 40% of the passage. These results suggest the powerful effects of testing, in that even just one testing period before taking a final test helped promote long-term retention better than merely restudying the passage.

Beyond the use of educationally relevant materials, research has also focused on studying the testing effect via test formats found in classrooms, such as short answer and multiple-choice. Results show that both test formats are related to positive testing effects on long-term retention, even when the format of the initial and final tests differ (Nungester & Duchastel, 1982; Laporte & Voss, 1975). Nungester and Duchastel (1981) also conducted a follow up to their experiment that compared performance on a final delayed test amongst three groups: a group that studied a prose passage and then reviewed it, another that studied and then took an initial test, and lastly, a group studied it only once. In the follow up study, the participants took a multiple-choice retention test five months after the initial learning session; they found similar pattern of results, where the tested group performed significantly better compared to groups who only reviewed it once or twice. This demonstrates that the testing effect not only occurs across multiple choice and short answer formats, but also persists over long intervals of time.

Within the common multiple-choice format of classroom tests, two types of multiple-choice questions are commonly asked: memory and inference items. Memory items assess students' recall for material explicitly stated in a text or lecture. Inference items measure students' abilities to analyze and understand what is beyond what is explicitly stated in text or lecture. While memory items only require students to have memory of the information but not necessarily any deeper understanding, inference items require students to integrate and apply information beyond what is explicitly stated in the text (McNamara, E. Kintsch, Songer & W. Kintsch, 1996; Peverly, Brobst, Graham & Shaw, 2003; Peverly & Sumowski, 2012).

There has been extensive research conducted by Peverly and colleagues on the relationship between notes quality and performance on these two types of multiple-choice items. Peverly et al. (2003) conducted an experiment in which college students were given a history text, and either asked to take notes on it or to study it without taking notes.   Students then took a free recall essay test and multiple-choice test that consisted of memory and inference items. Results indicated that notes quality and background knowledge were predictive of performance on inference items. However, for the notes group, there were no significant predictors for performance on memory items. Conversely, when Peverly and Sumowski (2012) evaluated the relationship between notes and test performance, they found that notes were only predictive of performance on memory items, but not the inference items.

There has been less research on the testing effect and types of multiple-choice items. Karpicke and Blunt (2011) compared the effectiveness of repeated retrieval and elaborative studying through the use of concept maps on delayed recall. Students studied a text in one of four conditions: 1) study once; 2) repeated study; 3) elaborative mapping and 4) retrieval practice. The study-once group studied the text in one study period and the repeated study group studied the text in four consecutive study periods. The elaborative mapping group learned about concept maps, viewed an example, and then created concept maps on paper while reviewing the text. Lastly, the retrieval practice group studied the text once and then practiced retrieval by recalling as much information as they could on a free recall test. After a one-week delay, students took a short answer test that included both verbatim and inference questions. Results indicated that retrieval practice produced the best learning across both types of questions, with an increase of 50% over elaborative concept mapping in long-term retention scores. Within the retrieval practice group, there were no significant differences in performance between verbatim and inference items.

Researchers also have examined the effects of different test formats and feedback, as well as whether one type of test is more effective for long-term retention, or if the format of the initial test facilitates the testing effect only for final tests of the same format. The literature indicates that feedback provides benefits for long-term retention, but its effectiveness is dependent on when it occurs. Agrawal, Karpicke, Kang, Roediger, and McDermott (2006) provided feedback to students in two conditions: in one condition, students received continual feedback by having access to the passage while taking the test, very similar to open-book testing conditions or taking notes while studying a textbook. In the other condition, students took the test and then restudied the passage (delayed-feedback condition). Results indicated that the immediate feedback condition performed better than the delayed feedback group on the initial test, but the latter group demonstrated better long-term retention. Beyond the direct effects on learning, feedback also gives students a chance to further inform their learning; not only can they learn from the feedback itself, but they can also use the test outcomes to guide their future studying, to focus on material they have not yet mastered (Roediger & Karpicke, 2006a). This may be why Hartwig and Dunlosky (2012) found that almost all high achievers used self-testing to study, largely as a metacognitive strategy to monitor their learning progress.

As for test format, laboratory research has examined the testing effect in studies with an initial recall test and a later recognition test (e.g. Darley & Murdock, 1971; Lockhart, 1975;

Wenger, Thompson, & Bartling 1980), as well as an initial recognition test and a later recall test (e.g. Carpenter & DeLosh 2006). Studies have also moved towards using educationally relevant materials to investigate the effects of different test formats. Dunchastel (1981) tested students on a prose passage via an initial short answer, multiple-choice, or free recall test, and then a final short answer or cued recall test two weeks later. Students who took the initial short answer test demonstrated better long-term retention than those who took the initial multiple-choice or free recall tests on the final cued recall exam. The author suggested that the final short answer test was not sensitive enough to produce differences between the groups. However, Dunchastel and Nungester (1982) found different results when they had high school students read a passage and then either take an initial short answer or multiple-choice test. A final test took place two weeks later, which was composed of 24 items; 12 from the initial short answer test and 12 from the initial multiple-choice test. While results showed that the initial short answer test group performed better than a non-tested control group, there were no significant differences between the two initial test groups on long-term retention.

More recently, Kang, McDermott, and Roediger (2007) and Butler and Roediger (2007) examined not only the effects of differing test formats, but also feedback in enhancing the testing effect. In the study conducted by Kang et al. (2007), students studied articles from a psychology journal, and then either took a short answer or multiple-choice test for each article. A final test that included both short answer and multiple-choice items on all articles occurred three days later. The authors conducted two versions of this study – one where students were given immediate corrective feedback after the initial test, and one where they were not. Results from the experiment where students were provided immediate feedback were consistent with most of the literature, where students who took the initial short answer test produced stronger recall compared to those who took the initial multiple-choice test, regardless of the final test format. Interestingly, the experiment where students were not given immediate feedback provided opposite results; the initial multiple-choice test promoted superior performance compared to the initial short answer test on the final test, also regardless of test format.

Butler and Roediger (2007) conducted a replication of the aforementioned study through a simulated classroom environment, by having students view a video recorded lecture prior to taking a short answer test, a multiple-choice test, or only restudying the facts. Students received feedback on only half of the test items. They then took a final short answer test 30 days later. Results showed students in the initial short answer test condition produced the greatest long-term retention of information on the delayed test, while those in the restudying and the initial multiple-choice test conditions performed equally. All three experimental conditions outperformed the control condition (different post lecture activity). Unlike Kang et al. (2007), this study did not demonstrate any significant effect for feedback. These results suggest the efficacy of an initial short answer test over multiple-choice exams in producing long-term retention of information.

Beyond the laboratory, studies on the testing effect have also been generalized to actual classrooms. Positive results of classroom testing on long-term retention have been found across middle school, high school, and college populations (e.g. McDaniel, Anderson, Derbish & Morrisette, 2007; McDaniel, Roediger, & McDermott, 2007). McDaniel, Agarwal, Huelser, McDermott and Roediger (2011) demonstrated these findings through the use of low stake quizzes with middle school students. The testing effect persisted until the end of the semester, with quizzing increasing students' performance on unit exams from a baseline of 79% correct (performance when items were non-quizzed) to levels of more than 90%. In regards to grades, this performance gain represented a change from a C+ grade to an A- grade. Additionally, another indirect benefit was that students indicated on an end of the semester survey that taking the quizzes reduced anxiety (64% of respondents) and increased learning (89% of respondents).

Summary and Research Questions

Notetaking is widely used by students from middle school to postsecondary education as one of the primary strategies to record and later review lecture information. There are two primary functions of notetaking: taking notes (encoding) and reviewing notes (external storage). The literature suggests that students are generally poor notetakers, as transcribing notes require students to actively engage with the material while selecting and writing down specific information from the lecture. Thus, later review allows for students to fully understand and conceptualize the material that may have not occurred during encoding. Research has demonstrated such benefits of review, as students who review their notes are more likely to perform better. However, given the limited research on which review strategies are effective, what is the best way for students to review?

One method is self-testing; with research demonstrating that testing promotes long-term learning and retention of material. The testing effect has been shown to be efficacious across grade levels and with educationally relevant materials, as well as beyond the laboratory and in the classroom. However, there has been limited research on the testing effect in relation to students' notes, which is one of the most commonly studied materials.

The purpose of this investigation is to extend the notetaking and testing effect literature by examining the effects of different note reviewing strategies on test performance. In this study, all participants will watch a recorded lecture and take their own notes. They will then be assigned to one of three conditions. In the traditional repeated review condition, participants will review their notes over two sessions. In the traditional testing effect condition (self-testing), students will review their notes in the first session and then recall as much of the lecture as possible in the second session. In the rewriting condition, participants will rewrite their notes in the first session and then review them in the second session. All students will either take a multiple-choice test five minutes or two days after the last study session.

The investigation will seek to answer the following questions: 1) Will there be a significant difference between repeated reviewing, self-testing (the testing effect), and rewriting and reviewing of notes, on test performance? 2) Will there be a significant difference in test performance based on time of test (immediate vs. delayed)? 3) Will there be a significant difference in performance based on type of multiple-choice question (memory vs. inference)? 4) Will form of studying and time of test interact to affect test performance?

CHAPTER THREE: METHODS AND RESULTS

STUDY 1

Method

Participants

Participants were 69 undergraduate students from a large public university located in the northeastern United States. The students' ages ranged from 18-26, with an approximate mean age of 19.10 (SD = 1.3). The sample consisted of 87% female students and 13% male students. The sample was White (84.3%), Asian (5.7%), Hispanic (4.3%), African American (2.9%), and Other (1.4%). Ninety four percent of the sample spoke English as their first language. Students identified themselves as undergraduates in their first year (72.9%), second year (13%), third year (10.1%), or fourth year (1.4%) of college. One student reported being in graduate school. Students reported a mean GPA of 3.4 (SD = .69). Students received extra course credit for participation, as approved by the university's IRB. If participants chose to leave the study early, they had the option of receiving extra course credit by finishing an alternative task. The task was to read two assigned articles from the *Journal of Educational Psychology* and outline and summarize them in detail. All participants chose to participate in the study to receive credit.

Experimental Design and Analysis

The experiment consisted of two between-subjects variables: study method (repeated review/reread vs. self-testing/repeated recall vs. rewriting) and time of test (immediate or delayed). Participants indicated their scheduling availability, and then were randomly assigned to different time blocks for each of the different study groups and time conditions.

All participants watched a video-recorded lecture and took their own notes. Participants were randomly assigned to a study condition: repeated review/reread, self-testing/repeated recall, or rewriting) as well as time of test within each group: immediate or delayed. The multiple-choice test consisted of 37 multiple-choice items, with 20 memory items and 17 inference items.

There was a significant issue with attrition in the delayed condition, where many participants attended the first session but did not return for the second session. A total of only 19 participants participated in both sessions for the delayed condition: 7 participants in the repeated review group, 7 in the self-testing group, and 5 in the rewriting group. Due to the insufficient number of participants in the delayed condition, only the data for the immediate condition were analyzed. Thus, the data for Study 1 consisted of one between-subject variable: study method (repeated review/reread vs. self-testing/repeated recall vs. rewriting). The dependent variables were total test score, memory items score, and inference items score.

The data were first analyzed with a one-way ANOVA. Subsequently, the data were analyzed with an ANCOVA, with notes as the covariate (note quality and quantity).

Materials

Materials consisted of a video-recorded lecture on emotion, a 37-item multiple-choice exam, which was taken either immediately or following a two-day delay, and a word search distractor task. Since all participants took their own notes, everyone was provided with sheets of blank paper, as part of their packets of material prior to the start of the study (See Appendix C). Participants in the rewriting and the self-testing conditions were given additional sheets of blank paper with either a line of instructions to organize their notes and to write down any additional or missing information from the lecture (See Appendix D), or to write down everything they could recall from the lecture, respectively (See Appendix E). A full copy of the verbal instructions for each time period can be found in Appendix G.

Lecture. A 16-minute long video-recorded lecture on the history of the psychology of emotion was presented using a projector. The lecture consisted of 2,662 words and was recorded at 150 words per minute. The lecture was chosen because it was short and structured, but still contained enough information for students to take notes on. There were no additional visuals used besides the video of the lecture. A full transcript of the lecture can be found in Appendix B.

Immediate and Delayed Tests. A 37-item multiple-choice test assessed participants' understanding of the lecture material (See Appendix F). Each multiple-choice item included four answer choices. Similar to other measures used in experiments on notetaking (Brown, 2005; Peverly & Sumowski, 2012), the questions were based on Kintsch's (1998) model of text comprehension, with 20 *memory* items assessing participants' ability to recall information explicitly stated in the lecture, and 17 *inference* items measuring their ability to infer information implied in the lecture. An inter-rater agreement of 89% was established regarding question classification (memory vs. inference), by having two raters go through the test items and decide whether they were memory or inference items. Disagreements were settled by consensus. The Cronbach's Alpha for the total, memory, and inference items were .69, .49, and .52, respectively. Item analysis indicated that item nine had a more untoward effect on reliability than the other items, so it was removed from the analysis. Subsequently, Cronbach's Alphas for the total and inference items were .70 and .55, respectively. The Cronbach's Alpha for memory items remained .49.

Scoring Tests. Trained raters using answer keys scored the tests, which were created and validated by our research team. Multiple-choice items answered correctly were scored one point, or if answered incorrectly, zero points. The total points were then summed, with the total possible points ranging from 0 to 37.

Scoring Notes. Notes were scored for quality and quantity. The number of propositions determined the note quantity scores. Note quality was determined by number of main ideas. Scorers created a rubric that detailed the main ideas, propositions, supplementary details, and themes from the lecture. Themes were scored one point if present, or zero points if not. The presence of each proposition was scored one point. Propositions that had multiple components were also scored for a total of one point (e.g. if there were four parts to the proposition, each part was scored .25 points). If all of the necessary propositions needed to establish a main idea were present, main ideas were scored for one point. Main ideas were scored for zero points if there were any missing propositions. Supplementary details were each scored one point. Similar to propositions, if they had multiple components, each component was scored accordingly so together, the proposition was still scored one point. Inter-rater agreement in scoring was established by having three independent raters randomly select and score 25% of the protocols. Overall, there were four themes, 30 main ideas, 95 propositions, and 21 supplementary details, which summed to a total of 150 decisions for each protocol. The total agreements for the protocols were added, and then divided by the total number of decisions for the protocols. Inter-rater agreements were 97%, 98%, and 96% for each pair of raters. Average inter-rater agreement was 96%. Disagreements were settled by consensus.

Word Search Task. A word search task was used as a distractor task for two minutes between the initial encoding period (T1) and the initial study period (T2), and then between T2 and the final study period (T3) (See Appendix H). In order to provide participants with a sufficient time delay and control for short-term memory effects on the exam, it was also used for five minutes between the final studying period (T3) and the immediate test (T4). To make sure participants continued to engage in each of the distractor tasks, a different word search puzzle was used at each break. The word search puzzles were from www.puzzles.ca, and consisted of a 19x19 matrix of letters and a list of 46 words hidden in the matrix; a 19x19 matrix of letters and a list of 60 words hidden in the matrix; and a 20x2x matrix of letters and a list of 51 words hidden in the matrix.

Procedure

Participants of the study were tested in small groups, over the course of one or two sessions, depending on their testing condition (immediate or delayed). Those in the immediate testing condition completed the experiment in one session, whereas those in the delayed testing condition completed it in two sessions. The total time for the whole experiment was approximately 65-75 minutes (depending on the study group), with 20 minutes for the final multiple-choice test. The videotaped lecture was administered in a classroom through the use of an electronic overhead projector and associated speaker system.

Prior to the start of the experiment, participants were randomly assigned to a study condition with participants in the same group placed in a different classroom (e.g. participants in the reread/review condition were all together). Those assigned to the immediate testing condition took the multiple-choice test five minutes after the final study period, while those in the delayed condition came back and took the test two days later. Participants were provided with a packet of materials for the study, dependent on their condition. The first pages of the material packets for all participants included a description of the study, information detailing their rights as participants, and an informed consent form for them to sign. Participants had the opportunity to ask questions at the beginning of each task. All participants first watched and listened to the lecture. Since all participants took their own notes during this task, they were given four blank pages for this purpose in their packets. The experimenter informed them prior to the start of the lecture that they would be able to use their notes to review for a test occurring later in the experiment; thus, they should take the best possible notes.

After all participants finished watching the lecture and taking notes, they turned to the next page, which consisted of a page with a word search task that they completed for two minutes. Subsequently, during the initial study period (T2), participants in the repeated review and self-testing groups reread/reviewed their notes for seven minutes. Those in the rewriting group had ten minutes to rewrite their notes (the differences in time were based on feedback from a pilot study). Participants in this condition turned to the next page in the packet, which was a blank page of paper with instructions to rewrite and organize their notes (they were instructed to add any missing information to make their notes more coherent and integrated). They were provided with four blank sheets of paper in total for this task.

Following this first study period, there was another short two-minute break where participants again turned to the next page and completed a different word search task. For T3, the final study period, participants either reread/reviewed their original notes, their rewritten notes, or took a written recall test for seven minutes. Similarly to the first study period at T2, participants in the reread/review or rewriting conditions reviewed their original or rewritten notes, without adding any additional notes on the paper. In the self-testing condition, participants turned to the next page in the packet, which was one sheet of paper with instructions to spend seven minutes writing down everything they could recall from the lecture. They were provided with four blank sheets of paper in total for this task.

At the end of the study periods, participants in the immediate test condition completed another different word search distractor task for five minutes, to provide a sufficient delay between reviewing and taking the test. This controlled for any short-term memory effects on test performance. They then took the 37-item multiple-choice test on information from the lecture. Lastly, they filled out a demographics questionnaire. These participants were then thanked and dismissed. As for the participants in the delayed test condition, they returned after two days to take the same 37-item multiple-choice test. Then they also filled out a demographics questionnaire, and were thanked and dismissed.

Results

Dependent Variable: Total Test Score

A one-way, between-subjects ANOVA was conducted to determine whether study method (repeated review (reread/review) vs. self-testing (test/retrieve) vs. rewriting (rewrite/review)) affected performance on Total Item Performance. See Table 1 for the means, standard deviations, and number of participants in each cell.

Table 1
Total Item Performance: Means and Standard Deviations by Study Method

Group	*M*	*SD*	*N*
Review	26.09	3.90	23
Self-Test	23.08	3.97	25
Rewrite	22.33	4.37	21
Total	23.86	4.33	69

Results indicated a statistically significant difference among the three different study methods on the immediate test ($F(2, 66) = 5.374$, $p = .007$). Post-hoc analysis indicated that performance was significantly higher for the repeated review group ($M = 26.09$, $SD = 3.90$), compared to the self-testing group ($M = 23.08$, $SD = 3.97$) and the rewriting group ($M = 22.33$, $SD = 4.37$). The self-testing and the rewriting groups were not significantly different ($p = .810$). See Table 2.

Table 2
Total Item Performance: Analysis of Variance Results for the Main Effect of Study Method

Variable	*df*	*MS*	*F*	*p*
Study Method	2	89.11	5.37	.007
Error	66	16.58		
Total	69			

Dependent Variable: Memory

A one-way between-subjects ANOVA was conducted to determine whether study method affected performance on Memory items. See Table 3 for the means, standard deviations, and number of participants in each cell.

Table 3
Memory Item Performance: Means and Standard Deviations by Study Method

Group	*M*	*SD*	*N*
Review	15.52	2.13	23
Self-Test	14.88	2.22	25
Rewrite	14.33	2.71	21
Total	14.93	2.37	69

Results indicated there was no statistically significant difference among the three study methods on memory items (p = .252). See Table 4.

Table 4

Memory Item Performance: Analysis of Variance Results for Main Effects of Study Method

Variable	*df*	*MS*	*F*	*p*
Study Method	2	7.80	1.41	.252
Error	66	5.53		
Total	69			

Dependent Variable: Inference

A one-way, between-subjects ANOVA was conducted to determine whether study method affected performance on Inference items. See Table 5 for the means, standard deviations, and number of participants in each cell.

Table 5
Inference Item Performance: Means and Standard Deviations by Study Method

Group	*M*	*SD*	*N*
Review	10.57	2.19	23
Self-Test	8.20	2.33	25
Rewrite	8.00	2.32	21
Total	8.93	2.53	69

Results indicated there was a statistically significant difference among the different study methods on performance on inference items ($F(2, 66) = 8.93$, $p = .000$). A comparison of the means and standard deviations for each study group indicated that the mean inference score was significantly higher in the repeated review group ($M = 10.57$, $SD = 2.19$), compared to the self-testing group ($M = 8.20$, $SD = 2.33$) and the rewriting group ($M = 8.00$, $SD = 2.32$). Similar to the previous results, post hoc analysis indicated there was no significant difference in performance between the self-testing group and the rewriting group ($p = .953$). See Table 6.

Table 6
Inference Item Performance: Analysis of Variance Results for Main Effects of Study Method

Variable	*df*	*MS*	*F*	*p*
Study Method	2	46.49	8.93	.000
Error	66	5.21		
Total	69			

Additional Analyses

Given the literature suggesting that there is a relationship between the quality and quantity of notes and test performance, the presence of notes was used as a covariate. Even though notes were scored for themes, propositions, main ideas, and supplementary details, only the first three were chosen as covariates since they are the most theoretically and empirically relevant (Peverly et al., 2003). Results indicated that themes were not significantly related to test outcomes, but propositions and main ideas were significantly related. The ANCOVAs for main ideas and propositions were virtually identical. Additionally, the correlations indicated that main ideas were the most predictive of all test outcomes. Thus, only the analyses for main ideas are discussed below. See Table 7 for the means, standard deviations, and number of participants in each cell for note variables. See Table 8 for correlations among note variables and test outcomes.

Table 7
Mean and Standard Deviations of Note Variables

Variable	*M*	*SD*	*N*
Themes	3.99	.13	69
Propositions	52.94	8.55	69
Main Ideas	14.48	4.41	69
Supplementary Details	3.70	2.17	69

Note: Themes = note themes; propositions = number of propositions; main ideas = number of main ideas; Supp. details = note supplementary details

Table 8
Pearson Correlations Among Note Variables and Test Outcomes

Variable	*1*	*2*	*3*	*4*	*5*	*6*	*7*
1. Themes	--						
2. Propositions	.442**	--					
3. Main Ideas	.346**	.919**	--				
4. Supp. Details	.096	.557**	.480**	--			
5. DV: Total	-.004	.398**	.400**	.289*	--		
6. DV: Memory	.-.004	.422**	.425**	.283*	.874**	--	
7. DV: Inference	-.003	.286*	.286*	.229	.891**	.558**	--

N = 69

Main Ideas. The number of main ideas present in participants' notes was used as a measure of note quality. Even after controlling for main ideas, participants in the repeated review group still performed significantly higher on total test and inference items than those in the self-testing and rewriting group. There was still no significant difference between subjects in the self-testing and rewriting group. As for the memory items, similar to previous results, there was no significant difference in performance between the three groups after controlling for number of main ideas. Main ideas had a significant effect on all three dependent variables. See Table 9, 10, and 11.

Table 9
Total Item Performance: Analysis of Covariance Results for Main Effects and Interaction Effects of Study Method and Number of Main Ideas

Variable	*df*	*MS*	*F*	*p*
Study Method	2	112.96	8.71	.000
Number of Main Ideas	1	251.72	19.42	.000
Error	65	12.96		
Total	69			

Table 10
Memory Item Performance: Analysis of Covariance Results for Main Effects and Interaction Effects of Study Method and Number of Main Ideas

Variable	*df*	*MS*	*F*	*p*
Study Method	2	12.40	2.81	.068
Number of Main Ideas	1	78.06	17.68	.000
Error	65	4.42		
Total	69			

Table 11
Inference Item Performance: Analysis of Covariance Results for Main Effects and Interaction Effects of Study Method and Number of Main Ideas

Variable	*df*	*MS*	*F*	*p*
Study Method	2	53.30	11.78	.000
Number of Main Ideas	1	49.43	10.92	.002
Error	65	4.53		
Total	69			

Quality and Quantity of Recalled Information. The quality and quantity of information recalled during study test trials was also examined in relation to test outcomes. An analysis of correlations indicated that only the number of propositions students recalled (quantity of notes) was significantly correlated with total test performance. See Table 12.

Table 12
Pearson Correlations Among Note Quality and Quantity of Recalled Information and Test Outcomes

Variable	*1*	*2*	*3*	*4*	*5*
1. Propositions	--				
2. Main Ideas	.789**	--			
3. DV: Total	.416*	.179	--		
4. DV: Memory	.396	.206	.866**	--	
5. DV: Inference	.331	.108	.878**	.520**	--

Note: N = 69*;* Propositions = number of propositions; main ideas = number of main ideas; total = total test score; memory = memory score; inference = memory score *$p < .05$ **$p < .01$

Quality and Quantity of Rewritten Notes. The number of propositions and main ideas included in students' rewritten notes were also examined in relation to test outcomes. Analyses indicated that quality and quantity of students' rewritten notes had no effect on test outcomes. See Table 13.

Table 13
Pearson Correlations Among Note Quality and Quantity of Rewritten Notes and Test Outcomes

Variable	*1*	*2*	*3*	*4*	*5*
1. Propositions	--				
2. Main Ideas	.898**	--			
3. DV: Total	-.016	.022	--		
4. DV: Memory	.168	.133	.889**	--	
5. DV: Inference	-.226	-.114	.846**	.509*	--

Note: N = 69*;* Propositions = number of propositions; main ideas = number of main ideas; total = total test score; memory = memory score; inference = memory score
*$p < .05$ **$p < .01$

Study 2

Method

Participants

Participants were 117 undergraduate students from a large private university located in the northeastern United States. This sample was more diverse across the various demographics, as compared to the sample in study one. The students' ages ranged from 18-45, with an approximate mean age of 20.94 (SD= 3.9) (three participants chose not to respond). The sample consisted of 55.6% female students, 37.1% male students, and .8% transgender students (one participant chose not to respond). The sample was White (54.8%), Asian (20%), Hispanic (9.6%), African American (7%), Native American (1.7%), and Other (7%) (two participants chose not to respond). Seventy five percent of the sample spoke English as their first language (one participant chose not to respond). Students identified themselves as undergraduates in their first year (29.3%), second year (36.2%), third year (20.7%), or fourth year (9.5%) of college (one participant chose not to respond). A small portion of the students reported being in a post baccalaureate program (4.22%). Students reported a mean GPA of 3.7 (SD = .48) (thirteen participants chose not to respond). Students received research hours to fulfill course requirements for their participation, as approved by the university's IRB. If participants chose to leave the study early, they had the option of receiving extra course credit by finishing an alternative task. The task was to read two assigned articles from the *Journal of Psychology,* and outline and summarize them in detail. All participants chose to participate in the study to receive credit. None of the participants in study one participated in study two.

Experimental Design and Analysis

The experimental design was the same as the one for the first study. The data were analyzed using 3 (repeated review/reread vs. self-testing/repeated recall vs. rewriting) x 2 (immediate vs. delayed) ANOVA. The dependent variables were total test score, memory items score, and inference items score. The data were then additionally analyzed using a 3 x 2 ANCOVA, with notes as the covariate (note quality and quantity).

Materials

The materials used were the same as those used in study one.

Scoring Tests. Trained raters using answer keys scored the tests. Multiple-choice items answered correctly were scored one point, or if answered incorrectly, zero points. The total points were then summed, with the total possible points ranging from 0 to 37. The Cronbach's Alphas for the total, memory, and inference items were .68, .52, and .53, respectively. Item analysis indicated that items five, twenty-three, twenty-six, and thirty-two had more untoward effects on reliability than the other items, so they were removed from the analysis. Afterwards, Cronbach's Alphas for total, memory, and inference items were .70, .54, and .56, respectively.

Scoring Notes. Participants' notes were scored for quality and quantity. The procedure used to score the notes was the same one used in study one. The number of propositions students mentioned in their notes determined note quantity scores. Note quality was evaluated by accuracy and presence of information, which was determined by number of main ideas. Inter-rater agreement in scoring was established by having three independent raters randomly select and score 25% of the protocols. The procedure used to establish inter-rater agreement in Study 1 was used in Study 2. The inter-rater agreement for each pair was 91%, 89%, and 94%. Disagreements were settled by consensus. Average inter-rater agreement was high (92%).

Procedure

The procedure was the same as that of study one. Participants were tested in small groups over the course of one or two sessions, depending on their testing condition (immediate or delayed). Those in the immediate testing condition completed the experiment in one session, whereas those in the delayed testing condition completed it in two sessions. The total time for the whole experiment was approximately 65-75 minutes (dependent on study group), with 20 minutes for the final multiple-choice test. The videotaped lecture was administered in a library room through the use of an electronic projector and associated speaker system.

Results

Dependent Variable: Total Test Score

A two-way, between subjects ANOVA was conducted to determine whether study method (repeated review/reread vs. self-testing/repeated recall vs. rewriting) and time of test (immediate vs. delayed) affected Total Item Performance. See Table 14 for the means, standard deviations, and number of participants in each cell. The Levene's test for the equality of variance was significant ($F(5,111) = 2.67$, $p = .026$).

Table 14
Total Item Performance: Means and Standard Deviations by Study Method and Time of Test

Group	*Time*	*M*	*SD*	*N*
Review	Immediate	30.81	2.98	21
	Delayed	26.24	5.38	21
Total		28.52	4.88	42
Self-Test	Immediate	31.00	3.79	18
	Delayed	26.65	2.50	17
Total		28.89	3.87	35
Rewrite	Immediate	29.25	4.42	20

	Delayed	26.40	2.94	20
Total		27.83	3.98	40
Total	Immediate	30.34	3.78	59
	Delayed	26.41	3.85	58
Total		28.39	4.28	117

Results indicated that the Time of Test main effect was significant, ($F(1, 111) = 30.35$, $p = .000$). Participants in the immediate condition performed better on total test items than those in the delayed condition. There was no significant difference in performance among the three study method groups ($p = .509$). Additionally, the interaction of study method x time of test was not statistically significant ($p = .554$). See Table 15.

Table 15
Total Item Performance: Analysis of Variance Results for Main Effects and Interaction Effects of Study Method and Time of Test

Variable	*df*	*MS*	*F*	*p*
Study Method	2	10.03	.680	.509
Time of Test (ToT)	1	447.77	30.35	.000
Study Method x ToT	2	8.77	.595	.554
Error	111	14.75		
Total	117			

Dependent Variable: Memory

A two-way, between-subjects ANOVA was conducted to determine whether study method and time of test affected performance on Memory items. See Table 16 for the means, standard deviations, and number of participants in each cell. The Levene's test for the equality of variance was significant ($F(5,111) = 3.60$, $p = .005$).

Table 16
Memory Item Performance: Means and Standard Deviations by Study Method and Time of Test

Group	*Time*	*M*	*SD*	*N*
Review	Immediate	18.24	1.26	21

	Delayed	15.57	2.16	21
Total		16.90	2.21	42
Self-Test	Immediate	17.89	1.68	18
	Delayed	15.88	1.73	17
Total		16.91	1.97	35
Rewrite	Immediate	17.00	2.66	20
	Delayed	15.80	1.77	20
Total		16.40	2.31	40
Total	Immediate	17.71	1.99	59
	Delayed	15.74	1.88	58
Total		16.74	2.17	117

Similar to previous results, the main effect of Time of Test was significant, $F(1, 111) = 29.89$, $p = .000$). Participants in the immediate condition performed significantly higher than those in the delayed condition. The main effect of study method was not statistically significant ($p = .422$). The interaction between study method x time of test was not significant ($p = .232$). See Table 17.

Table 17
Memory Item Performance: Analysis of Variance Results for Main Effects and Interaction Effects of Study Method and Time of Test

Variable	*df*	*MS*	*F*	*p*
Study Method	2	3.24	.87	.422
Time of Test (ToT)	1	111.41	29.89	.000
Study Method x ToT	2	5.52	1.48	.232
Error	111	3.73		
Total	117			

Dependent Variable: Inference

A two-way, between-subjects ANOVA was conducted to determine whether study method x time of test affected performance on Inference items. See Table 18 for the means, standard deviations, and number of participants in each cell. The Levene's test for the equality of variance was not significant ($F(5,111) = 1.58$, $p = .173$).

Table 18

Inference Item Performance: Means and Standard Deviations by Study Method and Time of Test

Group	*Time*	*M*	*SD*	*N*
Review	Immediate	12.57	2.23	21
	Delayed	11.62	3.04	21
Total		12.10	2.68	42
Self-Test	Immediate	13.11	2.56	18
	Delayed	10.76	1.35	17
Total		11.97	2.36	35
Rewrite	Immediate	12.25	2.59	20
	Delayed	10.60	2.04	20
Total		11.43	2.45	40
Total	Immediate	12.63	2.44	59
	Delayed	11.02	2.31	58
Total		11.83	2.50	117

Similar to previous findings, only the time of test main effect was found to have a significant effect on inference item performance ($F(1, 111) = 13.90$, $p = .000$). Again, participants in the immediate condition scored significantly higher than those in the delayed condition. There was no significant difference in performance on the inference items among the three groups ($p = .421$). The interaction of study method x time of test was not significant ($p = .444$). See Table 19.

Table 19
Inference Item Performance: Analysis of Variance Results for Main Effects and Interaction Effects of Study Method and Time of Test

Variable	*df*	*MS*	*F*	*p*
Study Method	2	4.95	.87	.421
Time of Test (ToT)	1	79.10	13.90	.000
Study Method x ToT	2	4.65	.82	.444
Error	111	5.69		
Total	117			

Additional Analyses

Similar to study one, notes were scored for themes, propositions, main ideas and supplementary details, but only the first three were used as covariates. Results indicated that themes were not significantly related to test outcomes, and that the analyses for note propositions and main ideas were virtually identical; thus, only the analyses for main ideas are presented here. Seel Table 20 for the means, standard deviations, and number of participants in each cell for note variables. See Table 21 for correlations among note variables and test outcomes.

Table 20
Mean and Standard Deviations of Note Variables

Variable	*M*	*SD*	*N*
Themes	3.99	.09	117
Propositions	56.50	8.19	117
Main Ideas	16.20	3.90	117
Supplementary Details	4.16	2.18	117

Note: Themes = note themes; propositions = number of propositions; main ideas = number of main ideas; Supp. details = note supplementary details

Table 21
Pearson Correlations Among Note Variables and Test Outcomes

Variable	*1*	*2*	*3*	*4*	*5*	*6*	*7*
1. Themes	--						
2. Propositions	.324**	--					
3. Main Ideas	.244**	.908**	--				
4. Supp. Details	.135	.481**	.414**	--			
5. DV: Total	-.100	.219*	.257**	.115	--		
6. DV: Memory	.-.054	.249**	.308**	.164	.778**	--	
7. DV: Inference	-.118	.064	.187*	.081	.773**	.473**	--

Note: N = 117*;* Themes = note themes; propositions = number of propositions; main ideas = number of main ideas; Supp. details = note supplementary details; total = total test score; memory = memory score; inference = inference score
*$p < .05$ **$p < .01$

Main Ideas. The number of main ideas present in notes was used as a measure of note quality. Time of test still had a statistically significant effect on total test, memory, and inference item performance, even after controlling for main ideas. Additionally, there were still no significant differences in test outcomes among the three groups after controlling for main ideas. There was a significant effect of main ideas only on total test and memory item performance, even though the results for inference items approached conventional significance (p = .052). Again, study method x time of test interaction was not significant. See Table 22, 23, and 24.

Table 22
Total Item Performance: Analysis of Covariance Results for Main Effects and Interaction Effects of Note Main Ideas, Study Method, and Time of Test

Variable	*df*	*MS*	*F*	*p*
Study Method	2	9.76	.70	.498
Time of Test (ToT)	1	421.26	30.29	.000
Study Method x ToT	2	5.64	.41	.668
Number of Main Ideas	1	107.80	7.75	.006
Error	110	13.91		
Total	117			

Table 23
Memory Item Performance: Analysis of Covariance Results for Main Effects and Interaction Effects of Note Main Ideas, Study Method, and Time of Test

Variable	*df*	*MS*	*F*	*p*
Study Method	2	3.63	1.07	.345
Time of Test (ToT)	1	103.26	30.58	.000
Study Method x ToT	2	4.31	1.28	.283
Number of Main Ideas	1	42.19	12.49	.001
Error	110	3.38		
Total	117			

Table 24
Inference Item Performance: Analysis of Covariance Results for Main Effects and Interaction Effects of Note Main Ideas, Study Method, and Time of Test

Variable	*df*	*MS*	*F*	*p*
Study Method	2	5.46	.98	.377
Time of Test (ToT)	1	74.16	13.37	.000
Study Method x ToT	2	4.36	.79	.458
Number of Main Ideas	1	21.34	3.85	.052
Error	110	5.55		
Total	117			

Quality and Quantity of Recalled Information. Study 2 also examined the relations between the number of propositions and the number of main ideas recalled during study test trials and test outcomes. An analysis of correlations indicated that the quality and quantity of information recalled had no significant effect on total test performance. See Table 25.

Table 25
Pearson Correlations Among Note Quality and Quantity of Recalled Information and Test Outcomes

	1	*2*	*3*	*4*	*5*
1. Propositions	--				
2. Main Ideas	.768**	--			
3. DV: Total	.159	.034	--		
4. DV: Memory	.157	.082	.874**	--	
5. DV: Inference	.131	-.013	.915**	.604**	--

Note: N = 117; Propositions = number of propositions; main ideas = number of main ideas; total = total test score; memory = memory score; inference = memory score
*$p < .05$ **$p < .01$

Quality and Quantity of Rewritten Notes. The number of propositions and main ideas included in students' rewritten notes were also examined in relation to test outcomes. Analyses indicated that quality of students' rewritten notes had a significant effect only on memory and total item performance. See Table 26.

Table 26
Pearson Correlations Among Note Quality and Quantity of Rewritten Notes and Test Outcomes

	1	*2*	*3*	*4*	*5*
1. Propositions	--				
2. Main Ideas	.783**	--			
3. DV: Total	.286	.349*	--		
4. DV: Memory	.293	.321*	.826**	--	
5. DV: Inference	.188	.264	.847**	.400*	--

Note: N = 117; Propositions = number of propositions; main ideas = number of main ideas; total = total test score; memory = memory score; inference = memory score
**p*< .05 ***p* < .01

CHAPTER IV: DISCUSSION

The studies conducted for this dissertation attempted to extend the notetaking and testing effect literature by investigating the effects of different note review strategies, including the testing effect, on test performance. Given that the predominant form of instruction in secondary education is lecturing, lecture notetaking is a widely embraced practice among college students. Notetaking serves two functions: *encoding* (taking notes) and *external storage* (reviewing notes). The literature indicates that while both encoding and external storage have strong effects on test performance, note review is much more predictive of academic outcomes by a factor of three (DiVesta & Gray, 1972; Fisher & Harris 1974; Kiewra, 1984a; Kiewra, 1989; Kobayashi, 2005; Kobayashi, 2006). However, students are generally poor notetakers because they typically take verbatim and/or incomplete notes (Williams & Eggert, 2000; Kiewra, 1989). Students also often choose passive study strategies, which result in more shallow understanding and less long-term retention (Marton & Saljo, 1976; Tomes, 2011). Rereading is the most common passive strategy, despite its ineffectiveness (Carrier, 2003; Goetz & Palmer, 1991; Callender & McDaniel, 2009; McDaniel & Callender, 2008). Instead, more active study strategies such as writing out notes, making a study guide, or drawing diagrams are more efficacious (Tomes, 2011). While there is a substantial amount of research on notetaking interventions (e.g. guided notes, partial notes, matrixes), there is very little research on effective note review interventions (Kiewra, 1987).

One method to consider is self-testing. A robust body of research suggests that self-testing, commonly known as the testing effect, promotes long-term learning and retention of material better than most other strategies (Butler et al., 2006; Kang et al., 2007; Weinstein, McDermott & Roediger, 2010; Agarwal et al., 2008; Roediger & Karpicke, 2006a; Nungester &

Duchastel, 1982; McDaniel, Roediger, & McDermott, 2007).

In the studies presented earlier in this dissertation, participants watched a recorded lecture and took notes. Students either studied their notes through repeated review, self-testing (the testing effect), or rewriting, before taking an immediate or delayed multiple-choice test. The independent variables included study method (repeated review/reread vs. self-testing/repeated recall vs. rewriting), and time of test (immediate vs. delayed). The dependent variables were total test score, memory item performance, and inference item performance. Covariates included note quality and quantity, as measured by the number of main ideas and propositions present in students' notes, respectively.

These studies sought to answer the following questions: a) Will there be a significant group main effect? b) Will there be a significant main effect for time of test? c) Will there be a significant group x time of test interaction? d) Will there be a significant difference in performance based on type of multiple-choice question (memory vs. inference)?

Both studies failed to support the efficacy of the testing effect, as compared to the other conditions. In Study 1, there was a main effect of study method on the total test and inference items. However, it was the students in the repeated review group who performed significantly better on the immediate test than those in the self-testing and the rewriting groups. These findings are consistent with the testing effect literature, in which students who repeatedly reviewed the material performed significantly better on the immediate test than those who repeatedly recalled. They also performed significantly better than those who rewrote their notes. Unfortunately, due to attrition in the delayed condition, the data were not sufficient to determine if the testing effect group would have performed significantly better than the other groups on the delayed test.

In Study 2, study method was not significant. Instead, Study 2 found a significant main effect of time across the three dependent variables. This indicates that students performed significantly better on the immediate test than the delayed test. The study method x time of test interaction was not significant for any of the three dependent variables.

Based on the literature demonstrating a significant relationship between the quantity and quality of notes and test outcomes, these investigations also examined whether students' notes taken during the lecture affected test performance. Three covariates were examined: note themes, number of propositions, and number of main ideas. Only number of propositions and main ideas had significant effects on test outcomes. In Study 1, number of propositions and number of main ideas were significantly related to all dependent variables. In Study 2, the results were mixed. Number of propositions and main ideas were significantly related to total test performance and memory items, but not inference items. However, for number of main ideas, there was a trend that approached significance for inference items.

These studies also examined the quality and quantity of notes taken during the study/test trials. In Study 1, the number of propositions generated by students in the self-testing group was predictive of performance only on the total test. The number of main ideas and propositions generated by students in the rewriting group were not significantly related to test performance. Results were similarly mixed in Study 2. Number of propositions and main ideas recalled by students in the self-testing group were not significantly related to test performance. In contrast, number of main ideas included by students in the rewriting group in their rewritten notes was related to performance on memory items and the total test.

Limitations and Future Research

There were several differences between this dissertation and studies in the testing effect literature, which could account for the failure to find an effect for repeated recall of notes on test performance. One difference was the number of study trials. Typically, the literature on the testing effect utilizes multiple test trials. For example, Experiment 2 in Roediger and Karpicke (2006b) compared the effects of multiple test trials (STTT), a single test trial (SSST), and no test trial (SSSS) on final recall. Results indicated that after the delay of a week, the multiple-test trials and the single test trial groups recalled the most information (61% and 56%, respectively), compared to the no test trial group (40% of the information), which suggests that there is not much difference between three test trials and one test trial on recall. Relatedly, Roediger and Karpicke (2006b, Exp. 1) found a significant testing effect when comparing participants who studied twice (SS), to those who studied once and self-tested once (ST).

Given that the materials used in the studies reported in this dissertation were lengthier than those typically used by Roediger and colleagues (e.g. short prose passages or short pair associates), which could add to its tediousness, the design of Studies 1 and 2 followed the lead of Roediger and Karpicke (2006b, Exp. 1), by reducing the number of study trials to reduce the length of the study. However, the same procedure and retention intervals used by Roediger and Karpicke (2006b, Exp. 1) were utilized to maintain comparability between the studies reported here and those of Roediger and Karpicke. Thus, it is unlikely that failure to replicate is due to the number of recall trials or the procedures.

Another more likely reason why Studies 1 and 2 did not replicate the testing effect is encoding. Craik and Tulving (1975) found that more elaborate encoding of information leads to better future memory performance. In the experiments conducted by Roediger and colleagues (e.g., Roediger and Karpicke, 2006b), all participants were exposed to the same materials (e.g., short prose passages; paired associates). Thus, individual differences in encoding were minimized and differences in performance were due to the strategies students used to study the information, but not to variations in the information itself. The nature of the initial encoding task in the experiments reported here, which consisted of students taking notes during a lecture, could have hampered encoding and limited the power of the testing effect. In other words, the materials students studied in this dissertation were student-generated and varied noticeably from student to student. It is well known that there are strong differences in the quantity and quality of notes (e.g., some are more verbatim than others) among students (e.g., Armbruster, 2009; Kiewra et al., 1989; Kiewra et al., 1987), which can limit the benefits of review. Research has shown that note quality and quantity is predictive of academic achievement (Kiewra, 1987; Kiewra & Benton, 1988; Kiewra & Frank, 1988), and students are much less likely to respond correctly to test questions if information included in the test was not written in their notes (Baker & Lombardi, 1985). This is consistent with the results found here, in which the quality and quantity of notes taken during lecture had a significant impact on test performance.

Interventions to Improve Recall as a Study Method

The nature of the initial encoding task, notetaking, may have mitigated the effects of repeated recall in the studies in this dissertation. To facilitate the benefits of the testing effect on notetaking, future studies may need to make modifications that change the traditional paradigm of self-testing, when used as a method for studying notes.

In Studies 1 and 2, students in the self-testing group free recalled as much information as they could from the lecture. Results indicated the quantity of the information students recalled correlated with performance only on total test items (.42), and only in Study 1, suggesting that recall had limited effects on test performance. In contrast to these results, Butler and Roediger (2007) produced the testing effect when they examined the benefits of testing on long-term retention of lecture material in a simulated classroom setting. Students watched a different lecture on three consecutive days and then completed a different learning activity that focused on that lecture: a multiple-choice test, a short-answer test, or review of a lecture summary. They were instructed to take notes, solely in order to simulate a classroom experience. Thus, students did not study their notes, and they were not used in the analyses. All students took a one-month delayed cumulative short-answer test that covered material from the three lectures. Results indicated that the intermittent short-answer test produced significantly better retention of the material on the final test than the multiple-choice test or review of the lecture summary.

Research indicates that free recall or other methods of testing that include fewer retrieval cues, typically invoke more elaborate and organized retrieval structures over time than tests with more retrieval cues (e.g. cued recall or recognition tests). Studies have shown that when comparing the effects of different types of intermittent tests, such as free recall, cued recall, and recognition, on final test performance, those that require production of information (e.g. essay or short-answer), rather than simply recognition (e.g. multiple-choice) lead to a more significant testing effect, most likely because of the more elaborate and generative nature of the retrieval. Future iterations of these studies here should examine whether a stronger testing effect may be produced if students in the self-testing group complete a short-answer test that allows for more elaborate retrieval during the test periods.

Another possible modification is the use of self-questioning as the intermittent test during test periods. This method of repeated recall may generate more elaborate retrieval and provide students with valuable corrective feedback. Research has shown that self-questioning is one of the study strategies that produces the most benefits in long-term retention and learning of material when studying (Tomes, 2011). King (1989, 1990, 1991, 1992) investigated the use of self-questioning in comprehending orally presented material during lectures. After the lecture, high school and college students engaged in study sessions in which they used a set of question stems ("What is the main idea of…" or "How does…relate to…)" to facilitate their processing of lecture content. These students performed better on lecture comprehension tests than those who used other independent review strategies.

King (1992) extended these experiments by comparing the efficacy of learning lecture content when students used three different strategies: self-questioning, reviewing, and summarizing of notes. The author found that students in the self-questioning group performed significantly better than those in the note review or summarizing groups, on a one-week delayed multiple-choice test on lecture content. The effectiveness of this study strategy is thought to be due to its cognitive and metacognitive functions. Self-questioning encourages students to organize the material and integrate new information with existing knowledge (Brown et al., 1983; Palincsar & Brown, 1984). It also encourages students to monitor their understanding of what they are studying, and helps them tailor their studying to focus on areas that are lacking.

In future replications of the studies in this dissertation, having students utilize question stems to test themselves during the test periods may facilitate deeper processing and improve accurate retrieval. For example, prior to the lecture, participants in the self-questioning group would receive generic question stems that would facilitate: a) understanding of the lecture content, b) generating connections within the lecture information, and c) making connections between prior knowledge and lecture knowledge. Researchers could also provide participants in the self-questioning condition with question stems that relate directly to the lecture (e.g. "How are the theories of Hippocrates and Aristotle similar?”). Participants would be told that after viewing the lecture and taking notes, they would use the question stems to review their notes and then generate and answer questions. A modification of Experiment 1 in Roediger and Karpicke (2006b), which incorporates these suggestions, would look something like this; students would generate and answer questions during the first test period, and then generate and answer more questions during a second test period. This would be an externally valid strategy since many students repeatedly test themselves in one study time period. Despite its potential benefits, a limitation of this replication could be that students may run out of time to generate and answer questions within a test trial. Future research should also examine whether the testing effect could be produced if students spent one test trial generating the questions, and then the next one answering them.

It may also be beneficial for repeated recall to allow self-testing participants to have the opportunity to receive feedback between testing trials. This can occur via explicit or implicit feedback. Those in the repeated review condition would continue to review their notes without feedback during the study periods (e.g. SSSS). Kang et al. (2007) found that students who received explicit immediate feedback after an initial short-answer test had superior retention on the final test three days later, regardless of the format (short-answer or multiple choice). Another implicit method to provide feedback to students is by having them re-encode information between testing trials, in the form of alternating between study and test trials (STST). For example, Karpicke and Roediger (2007) compared the retention of freely recalled words for three groups: repeated study (SSSS), repeated test (STTT), and standard (STST). The last condition consisted of students studying and recalling the list of words during alternating study and test trials. Results indicated that the STST group had superior recall compared to the other two groups. The authors postulated that the study trials allowed students to reexamine materials after seeing what they did and did not remember, assess their knowledge, and then focus their attention in studying the areas that were lacking. This more targeted focus on forgotten information may have improved retention. Future replications of these studies could include brief study trials between the test trials for the self-testing participants, modeled after the STST group in Karpicke and Roediger (2007). More specifically, students would watch a lecture and take notes as they usually do, and then some of them would repeatedly review their notes (SS), while others would alternate between studying their notes during the study trials, and then recalling the information from the lecture during test trials (ST). This would allow participants to receive immediate feedback to assess what they know and do not know, as well as more frequent chances for them to encode the lecture information. This procedure also has educational implications for students when studying. When testing themselves, students may benefit from reexamining their notes, testing themselves to receive feedback in terms of areas of missing information, and subsequently tailoring their studying towards these missing areas.

Interventions to Improve Encoding during Lectures

As previously mentioned, encoding a lecture (notetaking), compared to reading short passages or paired associates, may have reduced the power of the testing effect. Another possible modification to facilitate the effects of repeated recall may be to improve the act of encoding during the lecture. This can be done either through improving the lecture presentation itself, or the encoding process (notetaking).

Lecture Presentation. The lecture in these studies was presented at a typical pace compared to what is seen in the classroom (150 words per minute); however, students still may have had difficulty concurrently attending to the lecture and taking notes. To help students compensate for the relatively rapid pace of lectures, Bacchel and Thaman (2014) examined the effects of the pause procedure on lecture learning. For the experimental group, two-to-three minute pauses occurred three times during a 50-minute lecture. During the pauses, the students worked in pairs to discuss their notes, reorganize them, or add in missing information. The control group attended to the same lecture in the traditional lecture format. All students took a multiple-choice test 15 days later. Results indicated that students in the experimental group performed significantly better than those in the control group on the multiple-choice test. They also reported that they felt that having the opportunity to immediately review the material during the lecture enhanced their lecture recall and improved their understanding of concepts. Ruhl, Hughes and Schloss (1990) conducted a similar experiment and found that implementing the pause procedure during lectures also facilitated immediate and free recall for students. These results were also replicated with learning disabled students (Ruhl, Hughes, & Gajar, 1990), suggesting that the pause procedure facilitates retention of lecture information among struggling learners as well.

Making the aforementioned modifications to the lecture may facilitate improved notetaking and recall of information. Future replications of these studies should consider incorporating pauses during the lecture, in order to give students the opportunity to review and rework their notes. This can be done individually, or with a partner. Not only would this facilitate deeper encoding, but it would also provide implicit feedback for students that could improve the accuracy of their notes and later recall. These modifications have educational implications for instructors. Students' understanding and retention of lecture content may benefit from receiving short breaks during the lecture to assess their knowledge with a partner, examine and rework their notes, and ask questions to the teacher.

Another possibility to consider is embedding the testing effect during notetaking, rather than during the review of notes. Research has shown that college students report frequent lapses of attention and mind wandering during lectures (Bunce, Flens, Neiles, 2011; Lindquist & McLean, 2011; Wilson & Korn, 2007; Smallwood, McSpadden, Schooler, 2008; Smallwood & Schooler, 2006). To increase students' learning of lecture content and facilitate more task-relevant activities, such as notetaking, researchers should consider interpolating the lecture with memory tests.

In two experiments conducted by Szpunar, Khan, and Schacter (2013), students watched a video-recorded lecture that was divided into four segments, with a break between each segment. They were given copies of lecture slides on which they could take notes. During each break, all students spent the first minute completing arithmetic problems. In Experiment 1, students in the not-tested group continued to complete arithmetic problems for another two minutes at the end of each of the first three segments, and then took a two-minute test about the fourth lecture segment. The other students (tested group) took a two-minute test at the end of each segment on the recently learned material, rather than the continued arithmetic problems. In Experiment 2, a third group was added. In this group, called the restudy group, students were exposed to test questions paired with answers for an additional two-minute period after the first minute of arithmetic problems at the end of each of the first three lecture segments. They also took an additional short-two minute test at the end of the fourth lecture segment. All students took a final cumulative test on all four lecture segments that were comprised of the previous test questions in a new random order.

Results indicated that interpolating the lecture with memory tests reduced the occurrence of mind wandering, increased the frequency of notetaking, and facilitated learning. Students in the tested group retained more information from the fourth lecture segment on the final test, and scored better on the final cumulative test, than those in the non-tested and restudy groups. The tested group also took significantly more notes during the lecture (24% of slides with additional notes) than the restudy (9%) or non-tested groups (7%). Given that the students were essentially tested on the same questions twice, this could have contributed to the authors' positive results. Future replications of this study should use interpolating tests with questions that are different than those used in the final test.

It would also be interesting to see if other recall tests would have similar effects when interpolated during the lecture. For example, while the durations of the breaks between lecture segments would have to be lengthier, the effects of interpolating the lecture with other free or cued recall tests (e.g. short-answer tests) could be examined. Given the positive benefits of short-answer tests when used as intermittent tests, the testing effect may be increased if used during the lecture as well as the test trials. Alternatively, embedding self-questioning techniques during the lecture pauses could be beneficial as well. As mentioned previously, students who utilized question stems during studying demonstrated better retention of material on a delayed test (King, 1992). Students who are provided with question stems, and then generate and answer questions during lecture pauses may experience more elaborative encoding of the information, and subsequently greater recall. Since students would need more time to generate and answer questions, lectures should be shorter and segmented into several longer breaks.

Notetaking. Another possible way to improve the power of the testing effect is to better structure students' notetaking processes during the lecture. As discussed earlier, the materials participants studied in experiments 1 and 2 were student-generated and varied from student to student. The short prose passages used by Roediger and Karpicke (2006b) did not vary among students. This problem is compounded by variations among students in the quality (some are more verbatim than others) and quantity of notes (notes are often incomplete, e.g., Armbruster, 2009; Kiewra et al., 1989; Kiewra, Benton & Lewis, 1987). Also, students have a tendency to lose attention and focus during lectures (more so than they might reading short prose passages), which further decreases the possibility of including relevant information in their notes. This limits the benefits of both encoding and review. Research has shown that note quality and quantity is predictive of academic achievement (Kiewra, 1987; Kiewra & Benton, 1988; Kiewra & Frank, 1988) and students are much less likely to respond correctly to test questions if information included in the test was not written in their notes (Baker & Lombardi, 1985). This is consistent with the results found here, in which the quality and quantity of notes taken during lecture had a significant impact on test performance.

There are many research-supported interventions to help facilitate students' notetaking. They include guided notes, matrix notes, or instructor-provided complete notes (e.g. Cornelius, 2008; Cohn et al., Bradley, 1995; Kiewra et al., 1999), although guided notes is the method with the most empirical support. For example, Austin, Lee, and Carr (2004) investigated students' note quality under three presentation conditions: 1) traditional lecture, 2) lecture with slides, and 3) lecture with slides and guided notes. Students in the guided notes group recorded significantly more critical points and examples than those in the other two groups. Future replications of these studies should consider providing students in the repeated review and self-testing groups with guided notes during the lecture, in order to help them fully integrate and process lecture information during notetaking. This may improve students' encoding of the material, facilitate understanding of the main ideas, and generate connections within the material, resulting in more elaborate notes that can provide more cues in memory for better recall.

Interventions to Improve Rewriting as a Study Method

Experiments 1 and 2 sought to extend the literature by comparing self-testing and repeated review with another frequently used study method, rewriting. Research shows that more active study strategies facilitate deeper understanding and long-term retention of material (Marton & Saljo, 1976). Specifically, two of the active study strategies most associated with higher academic outcomes are self-testing and creating/writing/processing study materials. This included writing out notes. Karpicke et al. (2009) also found that rewriting notes is one of the most common study strategies, with 29% of students reporting they engage in this study practice. Thus, it was postulated that rewriting notes could be another beneficial method of reviewing notes, resulting in significantly better long-term recall of information. However, this was not the case. The students in the rewriting group did not perform significantly better on any of the dependent variables. This may have been due to several factors listed below.

For students in the rewriting group, the instructions stated to rewrite and organize their notes, and gave examples such as adding in missing information from the lecture or re-organizing the notes to make the information more coherent. The instructions also stated that the students should rewrite their notes so that they were good enough for someone who did not take notes to study for the test. Inspection of the notes indicated that many students simply rewrote their notes verbatim, rather than reorganizing the notes by different themes, generating and integrating information across topics, or adding in supplementary information from the lecture. Since many of the students in this condition rewrote their notes from beginning to end, they ran out of time and missed recopying information form the end of the lecture.

Despite this, results from Study 2 found that the number of main ideas included in students' rewritten notes was correlated to performance on memory and total test items (.32 and

.35, respectively). This is consistent with the notetaking literature which shows students were more likely to remember information if it were present in their notes (Kiewra, 1987). Memory items assess students' memory of explicit information expressed in the material, while inference items assess students' understanding and ability to apply information beyond what is stated in the material. Since many of the students rewrote their notes verbatim, it was most likely easier to recall the information stated explicitly in the lecture, resulting in the effect for memory items. Rewriting notes verbatim most likely had no benefit in facilitating deeper understanding of the material, resulting in limited effects for inference items.

Research has indicated that students benefit from paraphrasing and interpreting the information in their own words, which facilitates deeper processing and understanding (Kiewra, 1985a). It also helps students make connections within the text and with their prior knowledge (Wittrock, 1990). In a study conducted by Kiewra (1989), participants either reviewed their notes, or used their notes to write an integrative essay. They then took four tests: recall, application, factual recognition, and synthesis tests. The two groups performed similarly on the recall, application and factual recognition tests, while the review group outperformed the rewrite group on the synthesis test. The authors speculated that this might have been a result of the rewrite group only listing the categories of the lecture, rather than comparing and contrasting the topics to generate more connections and make the essay more integrative. Similarly, Shimmerlick and Nolan (1976) found that students who reorganized previously acquired textual information performed better on immediate and delayed free recall tests, than those who simply listed the information. That is, it is not enough to simply rewrite notes to facilitate learning.

However, students may need more explicit guidance that is similar to the methods used in self-testing to rewrite notes that benefit learning and recall. In a previously mentioned study, King (1992) found that students in the summarizing group performed no better than those in the note review group, on a one-week delayed multiple-choice test on lecture content. This finding is similar to the results here from Experiments 1 and 2, in which there was no significant difference in test performance between the rewriting group and the repeated review group. King (1992) suggested that this was due to the instructions of the rewriting group, which were not facilitative enough to generate accurate and elaborative processing. The summarizers were only instructed to use their own words to identify and generate connections among the main topic, subtopics, and main ideas. There were also no explicit instructions to go beyond the lecture, which may have limited the generation of elaborate external connections. King (1992) postulated that summarizing was not enough for students to construct a representation of the lecture in their long-term memory that was integrated enough to contain sufficient cues for recall. Additionally, since the students were merely summarizing their own notes, this strategy may have lacked the metacognitive benefits of self-testing, and limited the accuracy of their recall and retention of lecture content.

Future replications of these studies should consider operationalizing the rewriting condition by giving students more explicit instructions or guiding prompts in making the notes more integrative and coherent (e.g. using concept mapping). Rewriting notes may have also had limited benefits if the original notes were lacking relevant information and students merely rewrote them as is. Teachers should encourage students to use study strategies that include rewriting their notes and processing information in a more integrative manner. They can do this by providing explicit guidelines or guiding questions to facilitate more accurate and coherent generation of internal and external connections, which may result in a more complete representation of material in students' learning.

There were several limitations to keep in mind in regards to these studies. Specifically, in reference to the sample, participants in Study 1 were 87% women and 14% men. Participants were also 84.3% white, and 72.9% were in their first year of college. Thus, it was a homogenous sample. Another limitation was the operationalization of the rewriting condition in both Study 1 and Study 2. Even though rewriting notes is a study strategy many students utilize, there is typically no time limit when they are studying. This could be addressed by using a much shorter lecture with less content, to ensure all students would be able to finish rewriting their notes within a specific time frame.

Conclusion

The purpose of this dissertation was to investigate the efficacy of different note review strategies on test performance, in order to further explore the testing effect in relation with lecture notetaking. The two studies did not find evidence for the testing effect in the context of lecture notetaking, contradicting the numerous studies in support of the testing effect with text. In Study 1, repeated review produced significantly better performance on the immediate test, compared to self-testing and rewriting. While this is consistent with the testing effect literature, it is unknown whether repeated recall would have produced superior performance on the delayed test. Results for Study 2 instead found that performance was significantly better on all dependent variables (total test score, memory items, and inference items) for the immediate test. There was no interaction between study method x time of test. In terms of the relationship between the quality and quantity of notes and test outcomes, results generally indicated that number of propositions and main ideas were predictive of test performance. Quantity and quality of recalled information and rewritten notes appeared to have limited benefits on test outcomes. In Study 1, only the quantity of information recalled during test trials had a significant effect on total test performance. Results of Study 2 found that only quality of rewritten notes significantly affected total and memory item performance. Future research should continue to examine the testing effect in conjunction with lecture notetaking.

References

Austin, J. L., Lee, M., & Carr, J. P. (2004). The effects of guided notes on undergraduate students' recording of lecture content. *Journal of Instructional Psychology*, *31*(4), 314-320.

Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., III, & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology*, *22*(7), 861–876. doi:10.1002/acp.1391

Annis, L., & Davis, J. K. (1975). The effect of encoding and an external memory device on notetaking. *The Journal of Experimental Educational*, 44-46.

Bacchel, Rachna., & Thaman, Richa G. (2014). Effective use of pause procedure to enhance student engagement and learning. *Journal of Clinical and Diagnostic Research, 8*(1), 1-3.

Bacon, F. (2000). Novum organum (L. Jardine & M. Silverthorne, Trans.). Cambridge, England: Cambridge University Press. (Original work published 1620).

Baker, L., & Lombardi, B. (1985). Students' Lecture Notes and Their Relation to Test Performance. *Teaching of Psychology*, *12*(1), 28–32.

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*(2), 213–238. doi:10.3102/00346543061002213

Benton, S., Kiewra, K., Whitfill, J., & Dennison, R. (1993). Encoding and External-Storage Effects on Writing Processes. *Journal of Educational Psychology*, *85*(2), 267–280.

Bretzing, B. H., & Kulhavy, R. (1981). Note-Taking and Passage Style. *Journal of Educational Psychology*, *73*(2), 242–250.

Brown, A.L., Bransford, J.D., Ferrara, R.A., & Campione, J.C. (1983). Learning, remembering and understanding. In J.H. Flavell and E.M. Markman, (eds.), *Hand- book of Child Psychology, Vol III: Cognitive Development*(pp. 77-166). New York: Wiley.

Butler, A. C., & Roediger, H. L., III. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, *19*(4-5), 514–527. doi:10.1080/09541440701326097

Butler, A. C., Marsh, E. J., Goode, M. K., & Roediger, H. L. (2006). When additional multiple choice lures aid versus hinder later memory. *Applied Cognitive Psychology*, *20*(7), 941–956. doi:10.1002/acp.1239

Carpenter, S.K., & DeLosh, E.L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory*

& Cognition, *34*, 268–276.

Carrier, M. L. (2003). College students' choices of study strategies. *Perceptual and Motor Skills*, *96*, 54–56.

Carrier, C. A., & Titus, A. (1979). The effects of Note Taking: A Review of Studies. *Contemporary Educational Psychology*, *4*, 299–314.

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*(3), 354.

Cohn, E., Cohn, S., & Bradley, J. (1995). Notetaking, Working Memory, and Learning in Principles of Economics. *The Journal of Economic Education*, *26*(4), 291–307.

Cornelius, T. L., & Owen-DeSchryver, J. (2008). Differential Effects of Full and Partial Note on Learning Outcomes and Attendance. *Teaching of Psychology*, *35*(1), 6–12. doi:10.1080/00986280701818466

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268–294. doi:10.1037//0096-3445.104.3.268

DiVesta, F. J., & Gray, S. (1972). Listening and Note Taking. *Journal of Educational Psychology*, *63*(1), 8–11.

Duchastel, P.C., & Nungester, R.J. (1981). Long-term retention of prose following testing. *Psychological Reports, 49,* 470.

Duchastel, P.C., & Nungester, R.J. (1982). Testing effects measured with alternate test forms. *Journal of Educational Research, 75,* 309–313.

Dunkel, P., & Davy, S. (1989). The Heuristic of Lecture Notetaking: Perceptions of American International Students Regarding the Value & Practice of Notetaking. *English for Specific Purposes*, *8*, 33–50.

Fisher, J.L., & Harris, M.B. (1973). Effect of note taking and review on recall. *Journal of Educational Psychology, 65,* 321-325.

Fisher, J. L., & Harris, M. B. (1974). Effect of Note-Taking Preference and Type of Notes Taken on Memory. *Psychological Reports*, (35), 384–386.

Gates, A. I. (1917). Recitation as a Factor in Memorizing. *Archives of Psychology* (40th ed., pp. 1–104). New York: The Science Press. Retrieved from http://books.google.com/books/download/Recitation_as_a_factor_in_memorizing.pdf?id=YinQhb3ZS8YC&output=pdf&sig=ACfU3U0dCJ-oJYvoQ6ESH62kUpQQVGL2vQ

Goetz, E. T., & Palmer, D. J. (1991). Proportion of student's reporting strategy use. *Reading Psychology: An International Quarterly*, *12*, 203-204.

Hartwig, M. K., & Dunlosky, J. (2011). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review*, *19*(1), 126–134. doi:10.3758/s13423-011-0181-y

Howe, M. J. (1970). Repeated presentation and recall of meaningful prose. Journal of Educational Psychology, 61, 214–219.

Kang, S. H., McDermott, K. B., & Roediger III, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*(4-5), 528-558. doi:10.1080/09541440601056620

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval Practice Produces More Learning than Elaborative Studying with Concept Mapping. *Science*, *331*(6018), 772–775. doi:10.1126/science.1199327

Karpicke, J. D., & Grimaldi, P. J. (2012). Retrieval-Based Learning: A Perspective for Enhancing Meaningful Learning. *Educational Psychology Review*, *24*(3), 401–418. doi:10.1007/s10648-012-9202-2

Karpicke, J. D., & Roediger, H. L. (2008). The Critical Importance of Retrieval for Learning. *Science*, *319*(5865), 966–968. doi:10.1126/science.1152408

Karpicke, J. D., Butler, A. C., & Roediger, H. L., III. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory*, *17*(4), 471–479. doi:10.1080/09658210802647009

Karpicke, J., & Roediger, H. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*(2), 151–162. doi:10.1016/j.jml.2006.09.004

Kiewra, K. (1985a). Investigating Notetaking and Review: A Depth of Processing Alternative. *Educational Psychologist*, *20*(1), 23–32.

Kiewra, K. (1985b). Students' Note-Taking Behaviors and the Efficacy of Providing the Instructors Notes for Review. *Contemporary Educational Psychology*, (10), 378–386.

Kiewra, K. A. (1985c). Providing the Instructors Notes: An Effective Addition to Student Notetaking. *Educational Psychologist*, *20*(1), 33–39. doi:10.1207/s15326985ep2001_5

Kiewra, K. (1987). Notetaking and Review: the research and its implications. *International Sscience*, 233–294. doi: 10.1007/BF01326640

Kiewra, K. A. (1989). A review of note-taking: The encoding-storage paradigm and beyond. *Educational Psychology Review*, *1*(2), 147-172.

Kiewra, K.A., & Benton, S.L. (1985). The effects of higher-order review questions with feedback on achievement among learners who take notes or received the instructor's notes. Human Learning, 4 (1985), pp. 225–231

Kiewra, K. A., & Benton, S. L. (1988). The relationship between information-processing ability and notetaking. *Contemporary Educational Psychology*,*13*(1), 33-44. doi: 10.1016/0361-476X(88)90004-5

Kiewra, K., DuBois, N., Christian, D., & McShane, A. (1988). Providing Study Notes: Comparison of Three Types of Notes for Review. *Journal of Educational Psychology*, *80*(4), 595–597. doi: 10.1037/0022-0663.83.2.240

Kiewra, K., DuBois, N., Christian, D., McShane, A., Meyerhoffer, M., & Roskelley, D. (1991). Note-Taking Functions and Techniques. *Journal of Educational Psychology*, *83*(2), 240–245. doi: 10.1037/0022-0663.83.2.240

King, Allison. (1992). Comparison of self-questioning, summarizing, and notetaking review as strategies for learning from lectures. *American Educational Research Journal, 29*(2), 303-323.

Landrum, R. E. (2010). Faculty and Student Perceptions of Providing Instructor Lecture Notes to Students: Match or Mismatch? *Journal of Instructional Psychology*, *37*(3), 216–221.

Locke, E. A. (1977). An Empirical Study of Lecture Note Taking among College Students. *The Journal of Educational Research*, *71*(2), 93–99.

Lockhart, R.S. (1975). The facilitation of recognition by recall. *Journal of Verbal Learning and Verbal Behavior*, 14, 253–258.

Marton, F., & Säljö, R. (1976). On Qualitative Differences in Learning: I—Outcome and process. *British journal of educational psychology*, *46*(1), 4-11.

Maqsud, M. (1980). Effects of personal lecture notes and teacher-notes on recall of university students. *British Journal of Educational Psychology, 50,* 289-294.

McDaniel, M. A., & Fisher, R. P. (1991). Tests and Test Feedback as Learning Sources. *Contemporary Educational Psychology*, 192–201.

McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology*, *103*(2), 399–414. doi:10.1037/a0021782

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, *19*(4-5), 494–513. doi:10.1080/09541440701326154

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and instruction*, *14*(1), 1-43.

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior*, *9*(5), 596-606.

Morgan, C.H., Lilley, J.D., & Boreham, N.C. (1988). Learning from lectures: The effect of varying the detail in lecture handouts on note-taking and recall. *Applied Cognitive Psychology, 2*(2), 115-122.

Neef, N. A., McCord, B. E., & Ferreri, S. J. (2006). Effects of guided notes versus completed notes during lecture on college students quiz performance. *Journal of applied behavior analysis*, *39*(1), 123-130.

Nungester, R., & Duchastel, P. (2013). Testing Versus Review:, 1–5.

Nye, P., Crooks, T., Powley, M., & Tripp, G. (1984). Student Note-Taking Related to University Examination Performance. *Higher Education*, (13), 85–97.

O'Donnell, A., & Dansereau, D. (1993). Learning from Lectures: Effects of Cooperative Review. *The Journal of Experimental Education*, *61*(2), 116–125.

Palincsar, A.S., & Brown, A.L. (1984). Reciprocal teaching of comprehension-fostering and monitoring activities. *Cognition and Instruction, 1,* 117-175.

Palmatier, R. A., & Bennett, J. M. (1974). Notetaking Habits of College Students. *Journal of Reading*, *18*(3), 215–218.

Peverly, S. T., Brobst, K., Graham, M., & Shaw, R. (2003). College adults are not good at self regulation: A study on the relationship of self-regulation, note-taking, and test taking. *Journal of Educational Psychology, 95,* 335-346

Peverly, S. T., Ramaswamy, V., Brown, C., Sumowski, J., Alidoost, M., & Garner, J. (2007). What predicts skill in lecture note taking?. *Journal of Educational Psychology*, *99*(1), 167.

Peverly, S. T., Marcelin, G. E., & Kern, M. B. (2014). Tailoring interventions for students with difficulties in lecture note-taking. In J. T. Mascolo, D. P. Flanagan & V. C. Alfonso (Eds.). *Essentials of Planning, Selecting, and Tailoring Interventions for Unique Learners*. New York: John Wiley & Sons, Inc.

Peverly, S. T., & Sumowski, J. F. (2012). What variables predict quality of text notes and are text notes related to performance on different types of tests? *Applied Cognitive Psychology*, *26*(1), 104–117. doi:10.1002/acp.1802

Piolat, A., Olive, T., & Kellogg, R. T. (2005). Cognitive effort during note taking. *Applied Cognitive Psychology*, *19*(3), 291-312.

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, *60*(4), 437–447. doi:10.1016/j.jml.2009.01.004

Roediger, H. L., Finn, B., & Weinstein, Y. (2012). Applications of cognitive science to education. In S. D. Sala and M. Anderson (Ed.), *Neuroscience in education: The good, bad, and the ugly.* (pp. 128-151). New York, NY: Oxford University Press.

Roediger, H. L., & Karpicke, J. D. (2006a). The Power of Testing Memory . *Perspectives on Psychological Science*, *1*(3), 181–210. doi:10.1111/j.1745-6916.2006.00012.x

Roediger, H. L., & Karpicke, J. (2006b). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention, 1–8.

Roediger, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. Journal of Experimental Psychology: Applied, 17, 382–395.

Ruhl, K.L., Hughes, C.A., & Schloss, P.J. (1987). Using the pause procedure to enhance lecture recall. *Teacher Education and Special Education, 10,* 14-18.

Ruhl, K.L., Hughes, C.A., & Gajar, A.H. (1990). Efficacy of the pause procedure for enhancing learning disabled college students' long-and short-term recall of facts through lecture. *Learning Disability Quarterly, 13*, 55-64. doi:10.2307/1510392

Shimmerlick, S. M., & Nolan, J. D. (1976). Reorganizaation and the recall of prose. *Journal of Educational Psychology, 68*(6), 779.

Spitzer, H. (1939). Studies in Retention. *The Journal of Educational Psychology*, *30*, 1–16.

Szpunar, K. K., Khan, N. Y., & Schaacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Psychological and Cognitive Sciences, 110(*16), 6313-6317.

Titsworth, B. S., & Kiewra, K. A. (2004). Spoken organizational lecture cues and student notetaking as facilitators of student learning. *Contemporary Educational Psychology*, *29*(4), 447–461. doi:10.1016/j.cedpsych.2003.12.001

Weinstein, Y., McDermott, K. B., & Roediger, H. L. (2010). A comparison of study strategies for passages: Rereading, answering questions, and generating questions. *Journal of Experimental Psychology: Applied*, *16*(3), 308–16. doi:10.1037/a0020992

Wheeler, M., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory*, *11*(6), 571–580. doi:10.1080/09658210244000414

Williams, R. L., & Eggert, A. C. (2002). Notetaking in College Classes: Student Patterns and Instructional Strategies. *The Journal of General Education*, *51*(3), 173–199. doi:10.1353/jge.2003.0006

Wirt, J., Choy, S., Greald, D., Provasnik, S., Rooney, P., & Watanabe, S. (2001). The condition of education. Washington, DC: Government Printing Office (NCES Publication No. 2001072).

Wittrock, M.C. (1990). Generative processes of comprehension. *Educational Psychologist, 24,* 345-376.

Appendix A

Write your student identification number here: XXXX________

Please answer the following:

Gender: ___Female ___Male ___Transgender ___Other

Date of Birth: ___Month ___Day ___Year

Is English your first language? ___Yes ___No

I belong to the following group:

___Black/African-American
___Latina/Latino
___White American
___Asian-American/Pacific Islander;
___Native American/Alaskan Native;
___Other (specify: _____________)

Year in school (circle one): 1 2 3 4 other: ___

What is your major? _________________

What is your minor? _________________

Please estimate your overall academic average:

___ A+
___ A
___ A-
___ B+
___ B
___ B-
___ C+
___ C
___ C-
___ D+
___ D
___ D-
___ F+
___ F
___ F-
___ I prefer not to respond

Appendix B

Welcome back to Human Emotion. Today we will be answer the question of "What is an emotion?"

So William James said "To the psychologist along can such questions occur as: Why do we smile, when pleased, and not scowl? Why are we unable to talk to a crowd as we talk to a single friend? Why does a particular maiden turn our wits upside-down?" "The common man can only say: 'Of course we smile, of course our heart palpitates at the sight of the crowd, of course we love the maiden. And so, probably, does each animal feel about the particular things it tends to do in the presence of certain objects…"

So today we're going to answer this question that William James brought up long ago, of "what is an emotion?" As you can see here on our roadmap we'll do it as follows: we'll first discuss some early theories of emotion, then the components of what an emotion response is, followed by two different approaches to classify the landscape of emotion, followed by some take-away questions of the main points covered in our brief lecture today and finally an expert interview of an emotion scholar in the field.

So let's begin here with the theories of emotion. So what I'm going to do is approach this question by time traveling together to see how different scholars thought about what an emotion was, and how these theories evolved over time. Starting from the Ancient Greek, going through the Enlightenment, then to Darwin, William James, his predecessors James-Lange, Cannon-Bard, Schacter-Singer and concluding with some cognitive theories from Lazarus. So let's start first here with the Ancient Greek approaches to the question of "what is an emotion?" So as you can see here, theories about emotion stretched back as far as the Stoics of Ancient Greece. Some of the main points to take home when you think about how did they answer the question of "what is an emotion," is first looking at Hippocrates who said there were four humors. The four humors were black bile, yellow bile, phlegm and blood, and emotional health really consisted of a balance between all of these four humors. And when these four humors were out of balance, both mental and physical disease could take over. Aristotle really advocated this principle of Moderation. Essentially saying that part of what an emotion is, and part of what predicts emotional health, is finding a balance. Not having too little emotion or having too much emotion. So he stressed that this was a really important principle of emotional functioning. Both of these theories however, still saw emotions as being inferior to or in conflict with both reason and rationality.

Next, we move into the Enlightenment. So during the Enlightenment period we see this flourishing of intellectual attention devoted to really trying to pinpoint or understand what an emotion is. And here we see philosophers such as Descartes, Spinoza and Hume really trying to get in deep about what an emotion is. And for example, Descartes, he provided an intricate taxonomy of emotions, a description of their bodily causes, their effects and functions in his text *The Passions of the Soul*.

Perspectives on emotion from an evolutionary perspective were initiated in the 19th century by Charles Darwin in his book, *The Expression of Emotions in Man & Animals.* Darwin really advocated that emotions are not simply irrational, that they actually serve an important purpose for us as humans in both communication and also aiding in our survival. So he therefore argued that emotions did evolve via natural selection and that they were not specific to humans, and that we also see them in animals as well. So this work, thinking about this evolutionary approach to emotion really paved the way for future research, not only trying to understand the function of emotion across time, but also for more recent theories looking at the physical and neurobiological kind of underpinnings of emotion today.

If you go through time you're now back to William James who is the father of psychology in many ways. When he thought about the question of "what is an emotion," he advocated for what we think of now as the physiological approach. So he really argued that the essence of emotions is a physiological response, and that each emotion has a specific sort of profile in the body, certain patterns of heart rate, certain patterns of muscle tension. And in this way emotions are really secondary to bodily or physiological phenomena. So he would argue, for example, that you would have a stimulus in the environment, a snake as you can see here, followed by a certain physiological response in the body, such as your heart rate increasing, and that that is what would lead to the experience of an emotion such as fear. He worked with a student who sort of further pushed this idea of the physiological approach and together they wrote that "The perception of bodily states, as they occur, *is* the emotion."

Later theories by Cannon and Bard suggested that no, emotions are not simply physiological phenomena, and that, as you can see here, physiological responses alone cannot explain emotion experience. So this theory really argued that instead, physiological responses are too slow, it takes a while for our heart rate to increase, and often imperceptible. So this cannot account for our relatively rapid and intense emotions that often come online. So this theory went as follows: you would see a stimulus, just as you see a snake here, this would be followed by certain patterns of subcortical brain activation, sort of registering and getting that input that this is a snake in the environment, and this would be followed by the almost simultaneous experience of both your physiological responses kicking into gear, such as your heart rate increasing, as well as the experience or subjective quality of fear.

Schacter and Singer then came along and proposed what they called a 2-factor theory of emotion. So here they said the appraisal of the physiological experience defines and determines the emotion. So in other words they suggested that physiological reactions contribute to emotion experience by really facilitating a certain kind of cognitive appraisal or way that you evaluate what's going on in your body. And that evaluation is really what defines or gives rise to the emotion itself. So again they talked about two factors in emotion or two stages, and they go something like this. So in the first stage, you see physiological arousal in response to an evoking stimulus, such as the snake here. Then in the second you have what they thought of as a cognitive elaboration, or a way in which you're appraising the meaning of your heart rate increasing. And it's the way that you think about what that heart means that simultaneously gives rise the experience of emotion or fear. So if you interpret your heart rate as a sign that something is threatening, you're likely to feel fear. If you interpret your heart rate as something exciting in the environment, you're less likely to experience fear, perhaps more likely to experience something like happiness or joy.

More recent cognitive theories advocated originally by Lazarus here, really argued that cognitive activity in the form of judgments, evaluations or thoughts, is necessary for emotion to occur. So for example, again you might see a snake in the environment, and then what would happen next is you'd have some sort of cognitive evaluation that the snake is dangerous. You'd be thinking or evaluating the world as such. You would see your heart rate increasing, in response to sort of thinking and evaluating cognitively that the snake is dangerous, and that would lead to the fear experience. So without being able to judge, evaluate or think about a stimulus in the environment, such as a snake, that there is no emotion without cognition.

So we've sort of covered the landscape of some early theories of emotion dating all the way to more recent cognitive theories of emotion. Now I'd like to turn to thinking about, what are the necessary ingredients or components that make up an emotion response? So here, we'll turn now to the components of emotion.

So emotions are not one single phenomena, they have multiple pieces or multiple components. The first one that you can see here, is the idea that emotions have a valence. This simply means that emotions have a flavor to them, they can be positive, they can be negative or they can be neutral. Second, they have what people call an "aboutness." Emotions are about something, and this is referred to as an eliciting or an intentional object. The emotion can be about a snake, for example, as we discussed earlier. But emotions rarely occur without anything to which they're about that sort of gets them started or triggered. This third component really dates back to the components of Darwin who said that emotions serve a certain purpose or function, and in particular, emotions are vital to our survival. They enable us to pursue important goals. Without having emotions, we would have a lot of trouble pursuing things that are meaningful to us in our everyday life. Then finally here on the left, emotions have multiple components themselves. So when an emotion goes online or when it's started, there are different pieces to what make up or what characterize an emotion.

Let's turn here to this multi-component response to see exactly what it means. So when you think of an emotion, it actually has three kinds of core pieces or parts. The first is a subjective experience or qualia of what the emotion is. This is your internal representation or experience of what it feels like to have an emotion. Second is the sort of outward display of behavior. So like this woman here, when you're feeling a certain emotion like fear, you're likely to show it. Show it in your face and show it in your bodily actions and movements that go along with the emotion. And finally and importantly, there are physiological aspects to an emotion response. So you can see this in our brain, as you saw there, as well as in our autonomic nervous system. So things like changes in heart rate, changes in sweating, changes in your breathing.

So as much as we've talked about what an emotion is, we've talked about some theories and the components of an emotion. I think part of what helps students really understand what an emotion is, is also defining what an emotion is not. So let me tell you a few things that an emotion definitely is not. An emotion is not a mood. So a mood, as you can see here, is a more long-lasting state that can last days, to weeks, to even months. And unlike an emotion, it does not have an aboutness. It does not have an intentional or eliciting object. It's something more diffuse. So for example you may simply wake up in a bad mood for no particular reason, and you may feel that way for some time. That's more of a mood state, not an emotion. An emotion is also not a feeling, though in colloquial speak we often use the word feeling to refer to an emotion state, so saying I'm feeling happy, or I'm feeling sad. But really feeling refers to the subjective representation of an emotion, this sort of private or internal experience to the individual. Whereas an emotion also has a behavioral component, we express emotion on our face, it also has a physiological response in our bodies. As you can see here on the right, emotion is also not affect. So affect is this term that you'll hear a lot as you go through the course that I think of as a broader umbrella, so it's a broader, all-encompassing term. It refers to general topics of emotions, feelings and moods altogether. So under the broad umbrella of affect, emotion is just one particular component or instance.

An emotion is also not a personality trait. So personality traits are stable individual differences across different situations and times. Whereas emotions are a more brief response, typically to something external in the environment or an internal thought or feeling. And finally emotion is not our internal thoughts or cognitions. Although cognitions, as Lazarus said, can give rise to emotions, cognitions themselves are something quite distinct. Cognitions do not have facial expressions and they do not always have physiological changes in arousal that are accompanied by them.

So we're now going to turn to our third point on the roadmap, which is looking at different classification systems trying to organize these different components and pieces of emotion. And there's two main classification systems that we'll walk through today. The first as you can see here on the left is the classification system often referred to as the Basic/Discrete classification system for emotion. And here what you see is that the basic nut and bolts of the theory is that emotions are discrete or specific entities or categories. They're biologically fixed and they're universal to all humans. Some examples of emotions that fall under this perspective are what they think of as these basic hardwired emotions that we're evolutionarily equipped to experience and display. These include anger, disgust, fear, happiness, sadness and surprise. And it's these basic emotions that form the sort of primary colors of emotions landscape that give rise to more complex emotions. Things like guilt pride shame and these are thought to arise from combination of basic emotions or are culturally influenced and constructed. And as you can see at the bottom, and we have pictures at the top, some sample theorists that really sort of championed this Basic/ Discrete classification system were Paul Ekman, Rene Descartes and Silvan Tomkins. So this is in contrast to the Dimensional classification system of emotion. So here emotions are not specific or individual categories but rather they are thought of as a combination of several psychological dimensions. And you can see here a picture. This really refers to two different dimensions of an emotion response. So we see here pleasant and unpleasant. This is sort of thought of as the valence dimension of emotion, how pleasant and unpleasant are emotions. And this is contrasted with the arousal dimension, this has anxiety at the top and boredom at the bottom. Really we're thinking of how highly arousing are emotions versus how sort of lethargic or low arousing are they. You can plot these two dimensions sort of side-by-side or perpendicular. You can pick out an emotion sort of where it fits on this emotional landscape. And here you can also see three commonly associated theorists with the Dimensional account here. You can see the father of the first psychology laboratory here Wilhelm Wundt, James Russell and Lisa Feldman Barrett.

Appendix C

Notes:

Appendix D

Please take the next seven minutes to rewrite and organize your notes on the sheets of paper I have given you. For example, you can add in missing information from the lecture or re-organize the information to make your notes more coherent. You should rewrite and organize your notes so that they are good enough for someone who did not take notes to study for the test:

Appendix E

Please take the next seven minutes to write down everything you can recall from the lecture:

Appendix F

1. Walter Cannon and Phillip Bard believed
    a. That thoughts or judgments are needed to explain emotional experience
    b. Only physiological responses explained emotional experience
    c. Emotions are necessary for survival
    d. That physiological responses are not enough to explain emotional experience

2. After the James and James-Lange Theories of Emotion, how did the subsequent theories of emotion evolve?
    a. They continued to believe that only physiological reactions explained emotional experience
    b. Eventually theories also included physical reactions as an important piece of emotional experience
    c. They shifted from the sole focus on physiological reactions to explain emotion, to a stronger emphasis on cognition for interpreting either the stimulus or the physiological reaction
    d. They shifted from solely focusing on physiological reactions to also behavioral reactions to explain emotion

3. Who wrote *The Passion of the Soul*?
    a. Baruch Spinoza
    b. Charles Darwin
    c. Hippocrates
    d. Rene Descartes

4. Based on the Basic/Discrete system, embarrassment would be considered to be a
    a. Basic emotion
    b. Complex emotion
    c. Discrete emotion
    d. Universal emotion

5. According to Darwin's evolutionary approach, emotions
    a. Have evolved over time and serve an important purpose in communication and survival
    b. Are inferior to reason and rationality
    c. Have no important purpose
    d. Are only important for humans but not for animals

6. Professor Mood believes that all emotions have the same physiological reaction. Which of the following would most likely disagree?
    a. Charles Darwin
    b. Richard Lazarus
    c. William James
    d. The Stoics

7. Emotions
    a. Have multiple pieces and components
    b. Are one-dimensional
    c. Require a specific stimulus that elicits the emotion
    d. A and C

8. Which of the following individuals was a philosopher of the Enlightenment?
    a. Baruch Spinoza
    b. Aristotle
    c. William James
    d. Charles Darwin

9. Why is Darwin's theory more supportive of the importance of emotions?
    a. His theory argued that emotions were also seen in animals
    b. Emotions are important for adaptation to physical and social environments
    c. He believed that emotions are also important in communication
    d. Similar to Aristotle, he also believed in a balance of emotions

10. If an Ancient Greek were bedridden with a medical illness, what would Hippocrates probably say?
    a. He had angered the Greek gods
    b. He was feeling too much emotion
    c. One of his four humors was out of balance with the others
    d. He was feeling too little emotion

11. What is the difference(s) between mood and emotion?
    a. Mood does not have an eliciting stimulus
    b. Emotions are more long-lasting
    c. Mood is more long-lasting
    d. A and C

12. Based on the lecture, Hippocrates' theory of emotion is most similar to which later theory?
    a. Cognitive Theory
    b. James-Lange Theory
    c. Cannon-Bard Theory
    d. 2-Factor Theory

13. Julie just noticed the handsome man standing next to her on the train. According to Richard Lazarus, which of the below could occur?
    a. Fear, because of her fight or flight response
    b. Happiness, then her heart begins to race
    c. Sadness, because her heart is racing
    d. Anger, because her heart is racing

14. How were Hippocrates and Aristotle's theories similar?
    a. Both believed that rationality was superior to emotions
    b. Both theories thought emotions were unnecessary
    c. Both focused on positive emotions rather than negative ones
    d. Both believed that emotion was superior to rationality

15. According to William James, what is the correct sequence of events of an emotional response?
    a. Environment, emotion, physiological reaction
    b. Stimulus, physiological reaction, emotion
    c. Physiological reaction, cognitive appraisal, emotion
    d. Stimulus, emotion/physiological reaction

16. Mood is a sub-category of
    a. Emotion
    b. Cognition
    c. Affect
    d. Feeling

17. Based on the lecture, what are the different parts of a multi-component response?
    a. Physiological response, behavioral response
    b. Cognitive response, physiological response, behavioral response
    c. Subjective experience, physiological response, cognitive response
    d. Subjective experience, behavioral response, physiological response

18. According to the Cannon-Bard theory, what happens almost simultaneously during an emotional experience?
    a. The physical and physiological response to a stimulus
    b. The physiological response and cognitive appraisal
    c. The physiological response and subjective experience of a specific emotion
    d. Seeing the stimulus and physical reaction to it

19. Aaron receives a bad grade on his test and he rips up his paper. Ripping up his paper is an example of which component of emotion?
    a. Physiological
    b. Behavioral
    c. Subjective Experience
    d. Cognitive

20. What is the most likely reason Aristotle's theory of emotion was named the "Moderation Principle"?
    a. Individuals needed to have equal amounts of emotion and rationality
    b. Everything should occur in moderation
    c. Optimally, individuals should not have too little or too much emotion

d. Positive emotions provide balance to mental health functioning

21. What is a personality trait?
    a. Subjective representation of an emotion
    b. Thoughts or judgments
    c. Stable individual difference across different situations
    d. Feelings and moods

22. The goal directed nature of emotion is most closely associated with
    a. Stanley Schacter and Jerome Singer
    b. Charles Darwin
    c. Walter Cannon and Phillip Bard
    d. William James

23. John and Lee both appear to be sad. When their teacher asks them what is going on, John states that he does not know, but that he has felt this way for the last month. Lee says that he is sad because his girlfriend dumped him yesterday. Which interpretation is most likely true?
    a. They are both experiencing emotions
    b. Lee's sadness is greater because he was dumped
    c. John's sadness would be described as a mood, while Lee's would be described as an emotion.
    d. They are not experiencing emotions because there is no physiological reaction

24. Which of these four individuals most likely believed that emotions hindered rationality?
    a. David Hume
    b. Charles Darwin
    c. William James
    d. Carl Lange

25. Contemporary methods of moderating the effects of stress on well-being, such as therapy or mindfulness mediation, might most closely resemble the views of
    a. Hippocrates
    b. William James
    c. Aristotle
    d. David Hume

26. Schacter-Singer's theory of emotion is most similar to the theory of which of the following:
    a. William James
    b. Charles Darwin
    c. Walter Cannon and Phillip Bard
    d. Aristotle

27. In Cognitive Theory, what is necessary to experience an emotion when a stimulus is present in the environment?
    a. Physiological reaction
    b. Physical reaction
    c. Thoughts or judgments about the stimulus
    d. A subjective experience

28. According to Hippocrates, what were the four humors?
    a. Phlegm, blood, saliva, tears
    b. Happiness, sadness, anger, surprise
    c. Black bile, yellow bile, phlegm, and blood
    d. Black bile, blood, saliva, and yellow bile

29. Which of the theories below believed that only physiological reactions explained the emotional experience?
    a. Cognitive Theory
    b. James-Lange Theory
    c. 2-Factor Theory
    d. Cannon-Bard Theory

30. In the Dimensional system, which emotion would most likely fall high in the unpleasant dimension and high on the arousal dimension?
    a. Boredom
    b. Nervousness
    c. Frustration
    d. Anger

31. If Brian sees a bear and makes the judgment that it is dangerous, which leads to a physiological response and fear, this is an example of which theory?
    a. 2-Factor Theory
    b. Cognitive Theory
    c. James-Lange Theory
    d. Cannon-Bard Theory

32. An emotion is not a(n)
    a. Mood
    b. Feeling
    c. Affect
    d. All of the above

33. Which is an example of a basic emotion?
    a. Shame
    b. Guilt
    c. Pride
    d. Happiness

34. In the 2-Factor Theory of Emotion, what are the two factors or stages?
    a. Seeing the stimulus and physiologically reacting to it
    b. Physical reaction to the stimulus and then physiological arousal
    c. Physiological arousal to stimulus, and cognitive appraisal of the physiological response
    d. Seeing the stimulus and the emotional experience

35. Paul Ekman is a theorist associated with the
    a. Classification System
    b. Dimensional System
    c. Two Factor System
    d. Basic/Discrete System

36. What are the possible valences an emotion can have?
    a. Positive, negative, neutral
    b. Positive and negative
    c. Good, bad, neutral
    d. Good and bad

37. Who was the theorist who believed that certain emotions displayed characteristic physiological reactions, while other theorists believed that physiological reactions were subject to interpretation?
    a. William James
    b. Walter Cannon and Phillip Bard
    c. Richard Lazarus
    d. Stanley Schacter and Jerome Singer

Appendix G

Taking Notes:

In a moment, you will watch a lecture. Using the blank sheets of paper I have given you, please take the best possible notes. You will use them to review for a test later in the experiment.

1st Study Period:

Reread/review:

Please take the next seven minutes to reread and review your notes. Do not add any additional information or reorganize your notes.

Rewrite/review:

Please take the next seven minutes to rewrite and organize your notes on the sheets of paper I have given you. For example, you can add in missing information from the lecture or re-organize the information to make your notes more coherent. You should rewrite and organize your notes so that they are good enough for someone who did not take notes to study for the test:

Test/retrieve:

Please take the next seven minutes to reread and review your notes. Do not add any additional information or reorganize your notes.

2nd Study Period:

Reread/review:

Please take the next seven minutes to reread and review your notes. Remember, do not add any additional information or reorganize your notes.

Rewrite/review:

Please take the next seven minutes to reread and review the notes you have just rewritten. Do not make any additional changes to them.

Test/retrieve:

Please take the next seven minutes to write down everything you recall from the lecture on the provided sheets of paper.

Test Instructions:

In a moment, you will take a multiple-choice test on the lecture. Please answer all of the questions to the best of your ability. You will have 20 minutes to complete the test.

Word Search Instructions:

Please take the next two minutes to complete the following word search. You do not have to find all of the words.

Appendix H

A	K	Y	E	K	N	O	D	P	M	B	I	S	O	N	E	F	O	X
T	E	W	I	P	R	B	E	A	R	O	T	A	G	I	L	L	A	T
D	S	W	A	I	A	L	V	U	L	T	U	R	E	N	A	W	S	U
R	I	U	C	H	I	N	C	H	I	L	L	A	M	A	G	G	P	O
T	H	C	M	C	T	W	T	B	B	N	R	E	E	D	N	I	E	R
D	S	A	A	A	N	O	O	H	A	A	E	L	K	T	I	P	A	T
I	I	N	R	M	T	E	R	O	E	R	B	V	A	C	T	O	C	R
S	F	E	W	E	O	O	N	R	D	R	R	O	A	W	H	T	O	N
E	Y	A	C	A	T	O	P	I	A	P	G	A	O	R	G	T	C	E
V	L	G	O	D	C	S	S	O	R	P	E	R	C	N	I	E	K	R
D	L	L	B	O	C	A	M	E	P	E	C	C	H	U	N	R	O	W
D	E	E	R	W	A	O	M	A	Z	P	V	S	K	R	D	A	L	J
N	J	S	A	L	M	O	N	N	H	D	I	L	E	E	D	A	I	A
D	O	L	G	A	E	C	A	F	R	F	O	H	O	R	R	D	O	C
D	K	I	U	R	L	P	E	A	D	H	P	T	U	W	M	I	N	K
C	C	A	O	K	M	R	P	R	T	O	I	N	N	O	R	E	H	A
C	E	U	C	I	R	O	O	O	G	G	N	Y	E	K	R	U	T	L
A	G	Q	H	E	E	W	L	I	E	E	K	R	A	V	D	R	A	A
R	L	C	T	L	S	S	E	R	R	P	O	R	C	U	P	I	N	E

AARDVARK	DONKEY	HIPPOPOTAMUS	OTTER	SALMON
ALLIGATOR	DOVE	JACKAL	PANTHER	SLOTH
BABOON	EAGLE	JELLYFISH	PARROT	SWAN
BARRACUDA	ELK	KIWI	PEACOCK	SWORDFISH
BEAR	FERRET	LEOPARD	PELICAN	TIGER
BISON	FOX	LION	PIG	TORTOISE
CAMEL	GECKO	LLAMA	PORCUPINE	TROUT
CHIMPANZEE	GOAT	MACAW	QUAIL	TURKEY
CHINCHILLA	GOPHER	MEADOWLARK	RACCOON	VULTURE
COBRA	HAMSTER	MINK	RAVEN	WOLVERINE
COUGAR	HAWK	MOOSE	REINDEER	WOODPECKER
CROW	HERON	NIGHTINGALE	ROADRUNNER	WREN

B	G	O	L	D	E	N	R	A	I	N	A	S	O	B	C	M	L	E
I	I	D	T	T	E	E	T	S	C	Y	T	G	L	U	R	O	P	H
W	S	R	O	G	U	R	R	A	P	U	E	E	A	T	T	I	H	S
W	H	B	C	O	E	N	T	T	L	R	Z	N	L	T	N	O	E	A
P	L	A	I	H	W	A	L	I	R	A	U	O	D	E	M	N	M	Y
O	E	L	R	N	L	X	P	E	H	E	N	C	H	R	P	A	L	R
P	L	D	P	P	H	T	O	H	Z	D	B	C	E	N	A	C	O	R
L	N	C	A	T	R	I	C	B	O	A	E	B	T	U	T	E	C	E
A	C	Y	A	E	N	T	C	N	R	E	H	E	U	T	R	P	K	H
R	E	P	E	R	I	E	P	K	B	E	S	D	N	R	W	I	H	C
O	D	R	P	W	B	L	D	O	O	W	N	O	T	T	O	C	F	R
S	A	E	E	A	A	O	S	N	E	R	H	O	S	H	A	D	E	E
P	R	S	T	N	L	H	R	R	I	K	Y	W	E	E	C	D	Y	T
L	K	S	E	N	O	M	O	V	A	L	W	W	H	T	B	A	U	H
U	E	E	L	P	A	M	Y	O	I	S	H	O	C	U	P	N	E	A
M	E	N	I	R	A	T	C	E	N	T	L	L	D	P	L	L	N	P
P	E	A	R	C	O	G	K	N	I	G	A	L	L	A	S	P	E	N
E	V	E	Y	W	I	L	L	O	W	R	S	E	W	L	A	R	C	H
I	T	S	S	E	Q	U	O	I	A	I	E	Y	E	K	C	U	B	N

APPLE	CEDAR	LINDEN	POPLAR
APRICOT	CHERRY	LONDON PLANE	REDBUD
ASH	CHESTNUT	MAPLE	RUBBER TREE
ASPEN	COTTONWOOD	NECTARINE	SEQUOIA
ARBORVITAE	ELM	OAK	SPRUCE
BALD CYPRESS	FIR	PALM	SYCAMORE
BEECH	GINKGO	PEACH	TULIPTREE
BIRCH	GOLDENRAIN	PEAR	WALNUT
BOXWOOD	HAZELNUT	PECAN	WITCHHAZEL
BUCKEYE	HEMLOCK	PLUM	WILLOW
BUTTERNUT	HICKORY	PINE	YELLOWWOOD
CATALPA	LARCH		

A A D Y A W R O N I E T S N E T H C E I L W I
U T N H S P A I N O D E C A M R A L B A N I A
S E A D S F I N L A N D A Z E R B A I J A N M
T E L P O E D N A L E C I C T B U L G A R I A
R N R N S R U S S I A N F E D E R A T I O N L
I O E B E D R K R A M N E D C F T M O A I R T
A R Z Y I D N A E A S E A Z D U R H O V R B A
I T T N O A E A P L E L E I I N U A O N E S T
N H I A P R M W L H U C O S T N A G N L A L E
A E W M O M L O S R H X A V G A E L G C A C Y
M R S R L E I S N R E N E A E Z O I E G E T O
O N E E A N T D E T M H R M R N U R U R I A C
R I Y G N I H P N A E Y T E B M I T C C I I U
E R O L D A U G R A E N H E A O R A N S S K K
C E T N A B A I E N L A E I N O U A D U S A R
E L S U L T N I G O I T N G P M C R R R E V A
E A A I R O I L V N R O O L R I L A G P L O I
R N C E S K A T S T T G C C T O L O N Y A L N
G D T I N N E O E S A N I A S E T I N C W S E
T H E W D O B Y E R L L V A B A V O D L O M D

ALBANIA
ANDORRA
ARMENIA
AUSTRIA
AZERBAIJAN
BELARUS
BELGIUM
BOSNIA-HERZEGOVINA
BULGARIA
CROATIA
CYPRUS
CZECH REPUBLIC
DENMARK
ENGLAND
ESTONIA
FINLAND
FRANCE
GEORGIA
GERMANY
GREECE
HUNGARY
ICELAND
IRELAND
ITALY
LATVIA
LIECHTENSTEIN
LITHUANIA
LUXEMBOURG
MACEDONIA
MALTA
MOLDOVA
MONACO
NETHERLANDS
NORTHERN IRELAND
NORWAY
POLAND
PORTUGAL
ROMANIA
RUSSIAN FEDERATION
SAN MARINO
SCOTLAND
SLOVAKIA
SLOVENIA
SPAIN
SWEDEN
SWITZERLAND
TURKEY
UKRAINE
SERBIA/MONTENEGRO
VATICAN CITY
WALES