

UCLA

UCLA Electronic Theses and Dissertations

Title

Using Selective Redundancy and Testing to Optimize Learning from Multimedia Lessons

Permalink

<https://escholarship.org/uc/item/7z12s6xz>

Author

Yue, Carole Leigh

Publication Date

2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Using Selective Redundancy and Testing to Optimize
Learning from Multimedia Lessons

A dissertation submitted in partial satisfaction of the requirements for the
degree Doctor of Philosophy in Psychology

by

Carole Leigh Yue

2014

© Copyright by
Carole Leigh Yue
2014

ABSTRACT OF THE DISSERTATION

Using Selective Redundancy and Testing to Optimize Learning

from Multimedia Lessons

by

Carole Leigh Yue

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2014

Professor Elizabeth Ligon Bjork, Chair

Multimedia learning refers to learning from a combination of words and images. In the present dissertation, a multimedia lesson is defined as an animated, narrated educational video that depicts a scientific process—a format of instructional material becoming increasingly common in online, hybrid, and traditional classrooms. The overarching goal of the present research was to investigate how to optimize learning from multimedia lessons using two related theories of multimedia learning (the cognitive theory of multimedia learning [CTML], Mayer, 2009; the cognitive-affective theory of learning with multimedia [CATLM], Moreno, 2006) as guiding frameworks. A secondary goal was to explore students' metacognitive beliefs regarding the efficacy of various types of multimedia presentation conditions for learning.

In one body of research, we attempted to strike a balance between the positive and negative effects of including seductive details in a multimedia presentation (i.e., the tendency of tangentially related facts to engage interest but at the cost of impairing learning) by varying the level of redundancy between the narration and on-screen text. While inclusion of identical on-

screen text has been shown to impair multimedia learning, we previously found that inclusion of a certain level of non-identical on-screen text could actually increase learning (Yue, Bjork, & Bjork, 2013). We utilized this finding by presenting seductive details with identical on-screen text and key points with non-identical on-screen text, thereby inducing shallow processing of the former information and deeper processing of the latter information. Our results indicate that such a strategy leads to an optimal presentation scheme—that is, one in which seductive details can be presented during the lesson to engage students’ interest but without a cost to their learning of key information.

In another body of research, we focused on optimizing processing during restudy, rather than during initial encoding. After watching a multimedia lesson for the first time, participants practiced retrieving either the auditory modality while experiencing the visual modality, or vice versa. Although not leading to greater learning than re-presenting the intact lesson, this type of retrieval practice did result in equivalent learning, indicating that students can engage in such practice without harm to their performance on a later exam when needed (e.g., should they only have access to an instructor’s slides but not the accompanying narration from a lecture).

In a third body of research, we explored potential indirect benefits of testing on learning from multimedia lessons. After experiencing a lesson, students either read half the facts from the lesson or took a test on those facts, after which they studied either the same lesson again or an entirely different multimedia lesson. Taking the intervening test, as compared to restudying facts, led students to make more effective use of a second study opportunity (as reflected in superior final test performance) whether the multimedia lesson presented in that second study opportunity was the same or different from the first. That is, the test-taking experience generalized to better study of even totally different material. Interestingly, however, it was the

students who studied facts—not those who took a test on those facts—who had the more unrealistic sense of confidence about how well they would do on the final test.

In summary, the present research offers new insights—of both a theoretical and applied nature—for our understanding of multimedia learning. Our findings reveal new methods for inducing deep processing of key information and, importantly, ones that can be easily implemented in the construction and delivery of multimedia instructional materials for a variety of educational domains.

The dissertation of Carole Leigh Yue is approved.

Alan Dan Castel

Aimee Drolet Rossi

Daniel M. Oppenheimer

James W. Stigler

Elizabeth Ligon Bjork, Chair

University of California, Los Angeles

2014

In memory of Bob Miller, a good man and a great dad

TABLE OF CONTENTS

List of figures	ix
Acknowledgments	xii
Vita	xiv
CHAPTER 1: Introduction and Overview	1
Theories of Multimedia Learning	1
Metacognition and Testing	6
Outline of the Present Dissertation	8
CHAPTER 2: Using Varied Redundancy to Minimize the Impact of Seductive Details	10
Method	15
Results	18
Discussion	22
CHAPTER 3: Single-modality Retrieval Practice in Multimedia Learning	25
Experiment 1	27
Method	27
Results	29
Discussion	34
Experiment 2	35
Method	36
Results	39
Discussion	44
General Discussion	45
CHAPTER 4: Cued-recall Tests Can Potentiate Subsequent Study of a Multimedia Lesson ...	49

Experiment 1a.....	52
Method.....	52
Results and Discussion.....	55
Experiment 1b.....	60
Method.....	60
Results and Discussion.....	60
Experiment 2.....	63
Method.....	64
Results and Discussion.....	65
General Discussion.....	71
CHAPTER 5: General Discussion.....	76
Theoretical Implications.....	79
Practical Implications.....	80
Limitations and Future Directions.....	82
Concluding Remarks.....	83
Appendix A.....	84
Appendix B.....	86
Appendix C.....	87
Appendix D.....	88
References.....	89

LIST OF FIGURES

<i>Figure 1.1.</i> The cognitive theory of multimedia learning. Arrows represent cognitive processes, and boxes represent memory stores. Reprinted with permission from “Cognitive Constraints on Multimedia Learning: When Presenting More Material Results in Less Understanding,” R. E. Mayer, H. Heiser, & S. Lonn, 2001, <i>Journal of Educational Psychology</i> , 93, Figure 2, p. 190. Copyright 2001 by the American Psychological Association.	2
<i>Figure 1.2.</i> The cognitive-affective theory of learning with multimedia. Reprinted with permission from “Does the Modality Principle Hold for Different Media? A Test of the Method-Affects-Learning Hypothesis,” by R. Moreno, 2006, <i>Journal of Computer Assisted Learning</i> , 22, p. 151. Copyright 2006 by John Wiley and Sons	5
<i>Figure 2.1.</i> Average proportions of idea-units correctly recalled on the final free-recall test as a function of presentation condition. SDs = seductive details. Error bars represent the standard error of the mean in all figures	19
<i>Figure 2.2.</i> Average correct performance on the transfer test as a function of presentation condition	21
<i>Figure 3.1.</i> Average proportions of idea-units correctly recalled on the final free recall test for each condition in Experiment 1	31
<i>Figure 3.2.</i> Average correct proportions on the transfer test for each condition in Experiment 1	32
<i>Figure 3.3.</i> Stated condition preferences by participants in Experiment 1 shown as a function of the condition they experienced	34

<i>Figure 3.4.</i> Proportion correct on immediate and delayed recall tests for each condition in Experiment 2	42
<i>Figure 3.5.</i> Proportion correct on immediate and delayed transfer tests for each condition in Experiment 2	42
<i>Figure 3.6.</i> Stated condition preferences by participants in Experiment 2 shown as a function of the condition they experienced	43
<i>Figure 4.1.</i> Schematic of the basic experimental procedures for Experiments 1a and 1b (top half) and Experiment 2 (bottom half)	53
<i>Figure 4.2.</i> Average proportions of correct responses on the final cued recall test as a function of question type in Experiment 1a.....	56
<i>Figure 4.3.</i> Average confidence ratings in each condition as a function of time of judgment in Experiment 1a	58
<i>Figure 4.4.</i> Average proportions of correct responses on the final cued recall test as a function of question type in Experiment 1b.....	61
<i>Figure 4.5.</i> Average confidence ratings in each condition as a function of time of judgment in Experiment 1b	62
<i>Figure 4.6.</i> Average proportions correct on the final cued recall test for the first presented lesson as a function of question type in Experiment 2	67
<i>Figure 4.7.</i> Average proportions correct on the final cued recall test for the second presented lesson for the read and test conditions in Experiment 2	68
<i>Figure 4.8.</i> Average confidence ratings for first lesson material for the read and test conditions as a function of time of judgment in Experiment 2	69

Figure 4.9. Average confidence ratings for second lesson material for the read and test conditions as a function of time of judgment in Experiment 2 70

ACKNOWLEDGMENTS

Some of the work published in this dissertation was presented at Annual Meetings of the Psychonomic Society and the Association for Psychological Science Conventions. The James S. McDonnell Foundation funded this research. Dr. Nick Soderstrom collaborated on the experiments presented in Chapter 4.

I am indebted to a great many people who have helped me get to this point. Dr. Elizabeth Bjork, thank you for revising draft upon draft, asking insightful questions, giving excellent suggestions, always being pleasant when I popped by your office, and for setting an example of a kind, brilliant, successful woman in academia. Dr. Robert Bjork, thank you for your encouragement, your wide-open office door, and all your helpful questions and comments over the last five years. Dr. Alan Castel, thank you for your absolutely incredible support, wisdom, and kindness. Dr. Danny Oppenheimer, thank you for helping me to be a better researcher and teacher. Dr. Jim Stigler, Dr. Aimee Drolet Rossi, and Dr. Barbara Knowlton, thank you for your support and feedback. Dr. Stephen Chew, thank you for getting me interested in cognitive psychology in the first place and continuing to encourage me.

I am also grateful to the many current and former CogFoggers who have offered encouragement, insight, and laughter over the last few years: Courtney, Mikey, Doe, Veronica, Erin, Saskia, Alison, Nick, Adam, Catherine, Laura, Tara, Toshi, Monica, Kou, and many others. Jeri, thank you for being an excellent mentor and friend. KP, thanks for being my buddy since recruitment weekend. John, thank you for your openness and helpfulness in all things research. Thank you to the many research assistants who helped collect and score data: Amira, Kate, Iris, Katie, Mo, Hayley, Tai, Emiline, Daniela, Devon, Kaila, and Mariam, I can't wait to see what's in store for you. Cheryl Polfus, Lindsay Kovner, and Steve Herbert, thank you for doing

probably much more than I even realize to help keep me funded, teaching, and on track throughout the program.

CG friends: Heather, Hougant, Ki-Young, Donald, Nina, Cathy, Jen, Greg, Jinsy, Carl, Celina, Randy, Sebastian, Dan, Giovanni, and Aiko, thank you for your support and lots of fun times throughout this process. Thanks also to dear friends Jennifer, Rachel, Molly, and Allison for staying connected even though we're far apart.

Thank you, Mom and Dad, for always encouraging my excitement for learning and supporting me in my endeavors. To Mom especially, thank you for instilling in me a true love of reading. Zill, as in so many things, thanks for blazing the trail and doing this Ph.D. thing first, and for being an awesome sister.

Michael, your love, sense of humor, and support have been invaluable. I love you.

Thank you, God, for providing friends, family, challenges, ability, and opportunity throughout my life.

VITA

2009	B.A., Psychology and Spanish Samford University Birmingham, Alabama
2010	M.A., Psychology University of California, Los Angeles
2010-2012	Teaching Assistant and Teaching Associate Department of Psychology University of California, Los Angeles
2012	C. Phil, Psychology University of California, Los Angeles
2013-2014	Teaching Fellow Department of Psychology University of California, Los Angeles
2010-2014	Writing Consultant and Program Leader Graduate Writing Center University of California, Los Angeles
2013-2014	Content Creator Khan Academy Mountain View, California

PUBLICATIONS AND PRESENTATIONS

- Yue, C. L., Bjork, E. L., & Bjork, R. A. (2013). Reducing verbal redundancy in multimedia learning: An undesired desirable difficulty? *Journal of Educational Psychology, 105*, 266-277. doi: 10.1037/a0031971
- Yue, C., Kim, J., Ogawa, R., Stark, E., & Kim, S. (2013). Applying the cognitive theory of multimedia learning: An analysis of medical animations. *Medical Education, 47*, 375-387. doi: 10.1111/medu.12090
- Yue, C. L., Castel, A. D., & Bjork, R. A. (2013). When disfluency is—and is not—a desirable difficulty: The influence of typeface clarity on metacognitive judgments and memory. *Memory & Cognition, 41*, 229-241. doi: 10.3758/s13421-012-0255-8
- Yue, C. L. (2014, March). Optimizing learning from multimedia. Invited talk at the Center for Education Innovation in Life Sciences, UCLA, Los Angeles, CA.

- Yue, C. L., & Bjork, E. L. (2013, November). Modality-specific retrieval practice in multimedia learning. Poster at the 54th annual meeting of the Psychonomic Society Conference, Toronto, Ontario, Canada.
- Yue, C. L., & Bjork, E. L. (2013, November). Modality-specific retrieval practice in multimedia learning. Poster at the 54th annual meeting of the Psychonomic Society Conference, Toronto, Ontario, Canada.
- Yue, C. L., Oppenheimer, D. M., & Castel, A. D. (2013, November). Highlighting successful metacognition: Students optimize quantity of highlighting when studying. Poster at the 54th annual meeting of the Psychonomic Society Conference, Toronto, Ontario, Canada.
- Yue, C. L., & Bjork, E. L. (2013, May). Retrieval practice in multimedia learning. Poster at the 25th annual American Psychological Society convention, Washington, D.C.
- Yue, C. L., Bjork, E. L., & Bjork, R. A. (2012, November). What level of redundancy facilitates learning from multimedia lessons? Poster at the 4th annual Science of Learning in Medical Education Symposium, Los Angeles, CA, and the 53rd annual meeting of the Psychonomic Society Conference, Minneapolis, MN.
- Kim, S., Kim, J., Yue, C., Ogawa, R., & Stark, E. (2012, August). Multimedia learning principles in animation design: A review of medical animations. Poster at the annual Conference of the Association of Medical Education in Europe, Lyon, France.
- Yue, C. L., Bjork, E. L., & Bjork, R. A. (2012, May). Redundancy in multimedia learning: Effects on performance and metacognitive judgments. Poster at the 24th annual American Psychological Society convention, Chicago, IL.
- Yue, C. L., Bjork, R. A., Castel, A. D., & Bjork, E. L. (2011, November). When disfluency is not a desirable difficulty: The influence of typeface clarity on metacognitive judgments and memory. Poster at the 52nd annual meeting of the Psychonomic Society Conference, Seattle, WA.
- Yue, C. L., Bjork, R. A., & Bjork, E. L. (2011, October). How much is too much? The role of explanatory captions in technology-assisted educational materials. Poster at the Symposium on the Science of Learning in Medical Education, Los Angeles, CA.
- Yue, C. L., Bjork, R. A., & Castel, A. D. (2011, April). The effect of perceptual disfluency on metacognitive judgments and memory. Poster at the Western Psychological Association conference, Los Angeles, CA.
- Miller, C., & Chew, S. (2009, August). Cued processing, cognitive load and student learning. Poster at the American Psychological Association convention, Toronto, Ontario, Canada.
- Miller, C., & Chew, S. (2009, April). Cued processing, cognitive load and student learning. Talk at the Southwestern Psychological Association conference, San Antonio, TX.

CHAPTER 1: INTRODUCTION AND OVERVIEW

In its most literal interpretation, multimedia instruction was taking place long before computers became common fixtures in the classroom; a simple chalk illustration corresponding to an instructor's lecture presents information in both visual and verbal modalities, and thus constitutes a multimedia lesson. With advances in computer technology, however, more elaborate multimedia educational materials have become increasingly prominent, and multimedia instruction is now often synonymous with computer-assisted education. The goal of multimedia instruction is to help learners construct mental representations from input in two or more modalities. Multimedia lessons are composed of information presented in both verbal (e.g., spoken or written text) and pictorial (e.g., static illustrations, animations, graphs, diagrams) forms (Mayer, 2005, 2009). Ideally, these lessons are based on empirically derived principles of multimedia learning, such as the multimedia principle, which states that people learn better from words and pictures than from words alone (Mayer & Anderson 1991, 1992). This foundational principle, along with the increased demand for technology in the classroom, has provided the impetus for studying multimedia learning in more depth over the last several decades.

Theories of Multimedia Learning

The Cognitive Theory of Multimedia Learning

The primary theoretical framework for research on multimedia learning is the cognitive theory of multimedia learning (CTML; Mayer, 2009). Underlying this theory is the assertion that instructional design should be based on the human mind's structure and functions. CTML assumes three characteristics of human information processing: (1) people process verbal and visual information in separate channels in working memory; (2) working memory capacity in each of those channels is limited, and (3) active processing is vital for learning to occur. Figure

1.1 depicts this theory, with the boxes representing memory stores, and the arrows indicating cognitive processes.

As can be seen in Figure 1.1, visual and verbal information proceed through different channels in working memory. Visual information enters through the visual sensory modality and proceeds along that channel for further processing. Verbal information can enter through either the visual sensory modality (in the form of written text) or the auditory sensory modality (in the form of spoken words). After it passes through sensory memory, however, the representation of that information may shift to the verbal channel if necessary. For example, on-screen text may be initially processed in the visual channel, but that visual stimuli may be transferred over to the verbal channel in working memory.

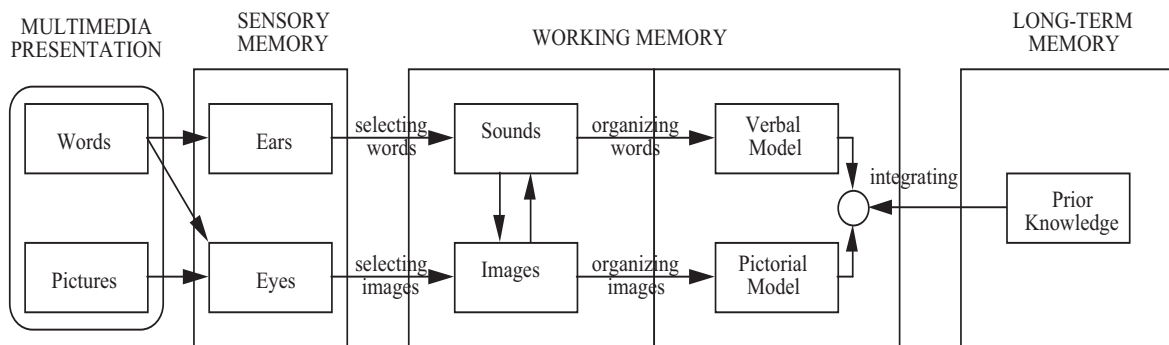


Figure 1.1. The cognitive theory of multimedia learning. Arrows represent cognitive processes, and boxes represent memory stores. Reprinted with permission from “Cognitive Constraints on Multimedia Learning: When Presenting More Material Results in Less Understanding,” R. E. Mayer, H. Heiser, & S. Lonn, 2001, *Journal of Educational Psychology*, 93, Figure 2, p. 190. Copyright 2001 by the American Psychological Association.

Because each channel has a limited capacity, it is not possible for learners to process all the information that initially comes in through sensory memory. Therefore, the job of working memory is to select relevant images and words and organize them into coherent visual and verbal models. At this point, the learner may retrieve related information from long-term memory and

integrate it with the information in each channel of working memory to create a single, coherent mental model of the concept at hand.

Within this basic structure, CTML proposes that learners can engage in three types of processing: essential processing, extraneous processing, and germane processing. Together, these types of processing create the total cognitive load that the learner experiences when engaging in a task (Sweller, Ayres, & Kalyuga, 2011). Essential processing is the basic amount of cognitive activity required by the task. For example, reading a textbook requires a minimal level of attention and text processing effort. Clearly, some essential processing is necessary in any kind of learning task; the goal of instructional design, however, is to manage the amount of essential processing. To illustrate, a novice learning about how a car brake works is facing several challenging tasks: learn the names of the relevant parts, know where they are located, and understand how they all work together. To manage essential processing, an instructor might first present a diagram that allows the student to learn the names of the parts and where they are, without being concerned yet about what they do. In a second step, the student would be expected to learn how the parts work together to create a functional brake system. This segmentation lessens the basic demand of each task and allows the student to learn more effectively than if they saw the entire lesson twice (Pollock, Chandler, & Sweller, 2002; Mayer, Mathias, & Wetzell, 2002)

Extraneous processing is any cognitive activity that uses cognitive resources but does not help the learner build an accurate mental model. Instructors, for example, often add “fun facts” to lessons with the intent of increasing student interest and engagement. Unfortunately, lessons that contain these so-called *seductive details* tend to lead to worse performance on later tests than lessons without such details (e.g., Mayer, Heiser, & Lonn, 2001). The reason for this impairment

lies in the limited capacity of working memory—if seductive details successfully engage interest, then students’ processing resources become devoted to the fun facts instead of the key lesson content, leading to worse learning.

Germane, or generative, processing involves cognitive activity that helps the learner create an accurate mental model around the key lesson content, such as engaging in self-explanation. Importantly, germane processing can only occur if there are enough cognitive resources remaining after essential and extraneous processing—and even then, there is no guarantee that learners will engage in germane processing. Elements of instructional design can promote such processing by allowing learners to pause, replay, or change the speed of a multimedia lesson (Schwan & Riempp, 2004) or encouraging learners to generate their own explanations during learning (Neuman & Schwarz, 2000).

In summary, CTML predicts that instructional design that promotes germane processing, manages essential processing, and/or minimizes extraneous will enhance learning from multimedia lessons. This theory has spawned decades of empirical research and the development of many principles of multimedia instructional design. The present dissertation is to further explore the notion of promoting germane processing by drawing in other theories of learning.

The Cognitive-affective Theory of Learning with Multimedia

Building on CTML, the cognitive-affective theory of learning with multimedia (CATLM) incorporates learner motivation, metacognition, and individual differences into the original CTML framework (Moreno, 2006). Figure 1.2 illustrates this theory in full, but the present dissertation focuses on multimedia learning involving visual and auditory stimuli. In addition to the existing assumptions of CTML (discussed above), CATLM adds two other empirically

supported assumptions that are particularly relevant to the present dissertation; namely, that (a) motivational factors mediate learning by affecting cognitive engagement and (b) metacognitive factors mediate learning by regulating cognitive processing.

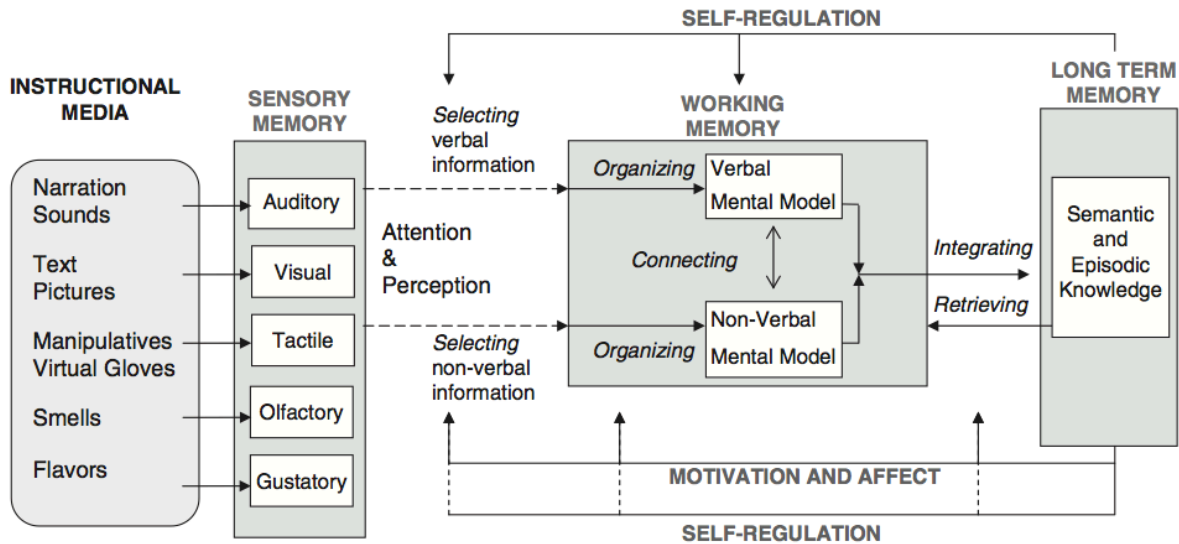


Figure 1.2. The cognitive-affective theory of learning with multimedia. Reprinted with permission from “Does the Modality Principle Hold for Different Media? A Test of the Method-Affects-Learning Hypothesis,” by R. Moreno, 2006, *Journal of Computer Assisted Learning*, 22, Figure 1, p. 151. Copyright 2006 by John Wiley and Sons.

Interest is one such motivational factor that may be a particularly important consideration in lessons that are designed to be self-regulated. Learners persist longer in reading text on a topic that interests them than on one that does not (Ainley, Hidi, & Berndorff, 2002). Similarly, people might be expected to watch more of an interesting video than an uninteresting one. When adding interest, however, instructors must take care to avoid introducing too much extraneous processing (Mayer et al., 2001). Previous research has examined how prior knowledge (Magner, Schwonke, Aleven, Popescu, & Renkl, 2014) and working memory capacity (Park, Moreno, Seufert, & Brünken, 2011) might mediate the impairment typically caused by seductive details. Because those factors are not under an instructor’s control, Chapter 2 of the present dissertation

extended this work by exploring the possibility of manipulating aspects of the lesson to increase germane processing of key information, despite the presence of seductive details.

In addition to motivational factors, CATLM addresses metacognitive factors, or people's thoughts about their own cognitive processes. Metacognition is important for comprehension, communication, memory, and many other mental processes (Flavell, 1979). For example, people's metacognitive judgments about various study techniques and their own ease of learning influence subsequent study behaviors (McCabe, 2011; Rawson, O'Neil, & Dunlosky, 2011). Metacognitive processes can be divided into two major components: monitoring and control (Nelson & Narens, 1994). These processes arise from distinctions and interactions between the object-level (i.e., the actual mental process) and the meta-level (i.e., the mental model or cognitive interpretation of the actual process). For example, an object-level action could involve a learner comprehending the narration of a video, whereas the corresponding meta-level action would be the learner's realization realizing that he or she had understood what the narrator said (Nelson & Narens, 1994). This meta-level awareness could then affect whether the learner chooses to continue with the video or replay a portion of it. This example illustrates a central concept in the study of metacognition: the object-level informs the meta-level, and the meta-level modifies the object level. When new information is present in the object-level, the learner's meta-level mental model either changes or notes that there is no need to change. As the mental model changes, control processes can be activated to initiate, maintain, or terminate an action, as in the example above.

Metacognition and Testing

Unfortunately, misleading surface characteristics of the learning material or the subjective ease of processing during learning or retrieval can lead to inaccurate metacognitive

monitoring, which can lead to ineffective metacognitive control (e.g., Rhodes & Castel, 2008; Benjamin, Bjork, & Schwartz, 1998; McCabe, 2011). Students often prefer learning strategies that appear effective during initial knowledge acquisition, without consideration for long-term benefits. For example, many studies have shown that testing (sometimes called retrieval practice) results in better long-term learning than re-reading (see Roediger & Butler, 2011, for a review). One possible reason for this benefit is that retrieving a memory strengthens a person's mental representation of that memory and engages generative cognitive processes (Bjork, 1975; McDaniel, Roediger, & McDermott, 2007).

Many students, however, may be forgoing this potentially useful learning tool: When asked about their study habits, students prefer re-reading text to taking a test on that information (Karpicke, 2009; Kornell & Bjork, 2007). Chapter 3 in the present dissertation investigated the potential benefits of testing—specifically, of testing one modality from a multimedia lesson—as a novel way to enhance germane processing during study to promote long-term learning from multimedia lessons. This situation would parallel a student who had been present in lecture, but only had lecture slides (visual modality) or an audio podcast (auditory modality) available for later study.

Beyond the direct effects of testing, recent research has begun to revisit the indirect benefits of testing—namely, the finding that taking a test can improve later restudy, a phenomenon known as *test-potentiated learning* (Izawa, 1966). Previous studies have demonstrated that testing can improve restudy of the same, previously studied material (e.g., Arnold & McDermott, 2013; Pyc & Rawson, 2010, 2012; Soderstrom & Bjork, 2014) as well as new, previously unstudied material (e.g., Szpunar, McDermott, & Roediger, 2008; Wissman, Rawson, & Pyc, 2011). Metacognitive awareness may be at the root of this phenomenon; tests

may help learners identify what they do not yet know or realize that they need to change encoding strategies. These two possibilities are not mutually exclusive, but they do suggest slightly different predictions for what would happen with a test over only some of the information. If testing only helps learners realize what they do not know, then untested information would be unaffected during restudy, as learners would focus primarily on previously tested information. If testing causes some change in encoding strategy, however, then untested information might well benefit as much as tested information during restudy.

This phenomenon has potentially strong educational implications, as students often take quizzes over a subset of material and then restudy all the material before the final exam. Little is known, however, about how testing a subset of information affects restudy of all the information in a lesson. Chapter 4 in the present dissertation explored this question, as well as how a partial test on some information might affect later study of new information. All experiments related to this issue also examined the role learner confidence might play in promoting more effective encoding strategies after a test.

Outline of the Present Dissertation

The overarching goal of the present research was to investigate how to optimize learning from multimedia lessons. More specifically, we investigated how to promote generative processing during encoding and restudy in order to maximize retention and transfer from a multimedia lesson. In Chapter 2, we explored whether varying redundancy between the narration and on-screen text could help focus learners' processing on relevant information and minimize the negative effect of interesting, but irrelevant, details—thus, allowing the inclusion of such details to keep the learner's interest high while not impairing his or her learning of relevant information. In Chapter 3, we investigated the potential direct benefits of single-

modality retrieval practice on recall and transfer, with the idea that such practice might lead to greater integration of the two modalities and thus construction of a better mental model of the topics to be learned. In Chapter 4, we report research that explored the potential indirect benefits of testing when learners were given the opportunity to restudy the original lesson, as well as an opportunity to study a new lesson. In Chapter 5, we summarize our findings and discuss implications, limitations, and future directions for this line of research.

CHAPTER 2: USING VARIED REDUNDANCY TO MINIMIZE THE IMPACT OF SEDUCTIVE DETAILS

In designing instructional materials, instructors often include intriguing or attention-grabbing details as a way of increasing student interest in a topic, even if those details are only tangentially related to the main point of the lesson. In a lesson about the life cycle of a star, for example, it may be tempting to include a photograph of a nebula that resembles an animal and to describe the constellation where that nebula can be located. Unfortunately, students sometimes remember such interesting details vividly, but at the cost of being able to remember less of the key information from the lesson. In such cases, these extraneous tidbits are known as *seductive details*, and the *seductive details effect* is said to occur when students recall fewer key points from a lesson containing seductive details than from the same lesson without seductive details (Garner, Gillingham, & White, 1989; Lehman, Schraw, McCrudden, & Hartley, 2007; Mayer, Griffith, Jurkowitz, & Rothman, 2008; Wade, Schraw, Buxton, & Hayes, 1993).

One type of instruction that may be particularly susceptible to the seductive details effect is multimedia instruction, especially given the increasing ease of providing static and/or dynamic visualizations to such lessons. Although both students and instructors may think that throwing in a cartoon or short video clip helps—or, at least, does not harm—student learning, research indicates that such interesting, but extraneous, images or video clips can actually impair learning (Mayer et al., 2008; Mayer, Heiser, & Lonn, 2001; Rey, 2011). The present research explored whether an ideal balance between the positive and negative effects of presenting such interesting details could be obtained. That is, is there a way to include such interesting extraneous details—thereby increasing students’ overall interest level in the lesson—but without impairing learning of the important facts of the lesson?

Background on the Seductive Details Effect

Much of the research on the seductive details effect is grounded in the relationship between interest and recall. Interest, once deemed by Hidi and Baird (1986) as a “neglected variable” in the literature on discourse processing, has been shown to impact recall in both positive and negative ways. On the positive side, interest in a topic can engage readers’ attention and promote useful encoding strategies (Hidi & Baird, 1986; Kintsch, 1980; see also Moreno, 2006). Conversely, however, interesting details can divert readers away from the main point of a passage, particularly if those details are not centrally related to the topic at hand (Hidi, Baird, & Hildyard, 1982; Wade & Adams, 1990). In these cases, the interesting information is said to “seduce” the learner’s cognitive resources away from the important information.

Our understanding of how seductive details do their damage in a multimedia environment is based on cognitive load theory (CLT) and the cognitive theory of multimedia learning (CTML) according to which there are three types of load that learners can experience: intrinsic, extraneous, and germane. Intrinsic load is the cognitive processing inherently required by the nature of the task to mentally represent the lesson content. For example, understanding a single spoken word imposes minimal intrinsic load because the listener only needs to know vocabulary, but understanding an entire sentence requires the listener to incorporate that knowledge of vocabulary with knowledge of grammatical structure, thus imposing greater intrinsic load. Extraneous load involves the engagement of mental activities in response to aspects of the learning situation that do not help the learner create an accurate mental model (e.g., distracting sound effects, irrelevant information). Germane load occurs when learners engage in generative processing and create coherent mental models of the subject to be learned or comprehended—that is, an organized conceptual framework that learners can recall and use to solve new

problems in a related context. This type of processing, although effortful, is necessary for the learner to understand the topic at hand (Mayer, 2005; Sweller, Ayres, & Kalyuga, 2011). Seductive details, then, increase intrinsic and extraneous load, but do not contribute to germane load. More specifically, seductive details are assumed to divert the learner's attention and schema activation processes (Harp & Mayer, 1998; Lehman et al., 2007); thus, because it is also assumed that learners' cognitive resources are limited, this diversion results in seductive details being processed at the expense of, rather than in addition to, key facts.

Within this framework, two hypotheses regarding the exact mechanisms underlying the seductive details effect—the attention distraction hypothesis and the schema diversion hypothesis—have been proposed, and both informed the present research. The attention distraction hypothesis states that seductive details draw the learner's attention away from key information and toward the seductive details. Lehman et al. (2007) found evidence for this explanation by measuring reading time of seductive detail sentences versus key fact sentences in a text passage. In addition to the expected impairment to retention, participants spent less time reading key fact sentences when seductive details were present than when they were absent, indicating that seductive details drew attention away from key facts.

The schema diversion hypothesis states that seductive details activate prior knowledge related to that information, leading the learner to form a mental model around the seductive details instead of around the key facts. In support of this hypothesis, Harp and Mayer (1998) found that the location of seductive details' presentation in a lesson played a determining role in the occurrence of the seductive detail effect. When, for example, they were presented at the end of a lesson (instead of at the beginning or interspersed throughout), retention of key points and transfer were the same as in lessons without seductive details. Harp and Mayer theorized,

therefore, that presenting seductive details at the end minimized the degree to which any activation of prior knowledge specifically related to the seductive details could detract from learners' construction of conceptual models based on the main points.

The Harp and Mayer (1998) manipulation thus provided theoretically important findings, but we do not see their way of eliminating the seductive detail effect as providing an effective means for achieving the goal of the present research. That is, we do not see it as addressing the goal of finding a way to be able to include interesting extraneous facts (or seductive details) in a lesson in order to spark and/or maintain the learner's interest throughout a lesson, but without harming learning of the key points of the lesson. Clearly, having all of the seductive details presented at the end of the lesson would not serve the purpose of capturing or maintaining interest throughout a lesson.

Other previous efforts to eliminate the seductive details effect have included use of typographical signals (e.g., bolded sentences) or organizational signals (e.g., preview sentences and numbered steps), but these were unsuccessful (Harp & Mayer, 1998). Although these techniques may often serve as useful guides for text processing (Dixon & Glover, 1990; Loman & Mayer, 1983; see also Gaddy, van den Broek, & Sung, 2001), they require the learner to make a high-level decision to follow such guides, and apparently the temptation of seductive details is stronger than the intent to follow these guides. Or, possibly, the guides merely created a feeling of fluency, making learners feel that they could understand the main text with little effort.

The Present Research

Given these previous unsuccessful attempts to eliminate the seductive detail effect, we sought a manipulation that would not require such high-level decisions on the part of learners. Rather, we focused on one that would more or less automatically force them to engage in more

effective processing of key information during its presentation than they would use when their attention and processing resources were diverted away by seductive details. For developing such a technique, we turned to recent findings on the redundancy principle of CTML, which states that when an animation is accompanied by a narration and on-screen text identical to that narration, learning is impaired (Kalyuga, Chandler, & Sweller, 1999; Mayer et al., 2001; Mayer & Johnson, 2008). When on-screen text is not identical, however, learning can be fostered (Mayer & Johnson, 2008; Yue, Bjork, & Bjork, 2013; see also Adesope & Nesbit, 2012, for a review).

The impairment caused by identical information is said to occur because completely identical information presented in two modalities requires extraneous cognitive processing to ensure that the two sources of verbal information match. When they do match, learners do not need to do any additional processing of the information; in short, they only have to process the verbal information at a superficial level to ensure that the visually and verbally presented words are the same. In contrast, benefits for non-identical information may occur because reconciling two different sources of verbal information that have the same meaning requires not only additional attentional resources, but also deeper, semantic-level processing—thus increasing germane load and, in turn, learning.

If non-identical text does increase attention and induce germane processing, it may be possible to mitigate the seductive details effect by (a) accompanying presentation of the key information in a lesson with non-identical text to make processing of that information occur at a deeper and more effective level, while (b) accompanying presentation of any seductive details with identical text to minimize deeper processing of them. Such an arrangement might thus allow

us to increase students' interest in the overall lesson with the inclusion of interesting extraneous details, but without a cost to the students' learning of the key information.

To test this idea in the present study, we had different participants experience different types of multi-media lessons, including one containing no seductive details and two containing seductive details. In a critical version of the lesson including seductive details, we provided identical text in conjunction with the presentation of seductive details and non-identical text in conjunction with the presentation of key information in an effort to selectively induce germane processing of key information and shallow processing of seductive details. While we expected participants to prefer lessons with seductive details, our hope was that those presented with this latter type of lesson would not demonstrate a seductive detail effect, as measured by a recall test. If so, then we would have found a way to maintain high student interest while optimizing processing for key information and minimizing processing for seductive details. Because non-identical text tends to have a stronger benefit for recall than transfer (Mayer & Johnson, 2008; Yue et al., 2013), we did not expect any differences in transfer scores among presentation conditions.

Method

Participants and Design

A total of 75 participants were recruited through Amazon Mechanical Turk and were paid \$1 for their participation. Three participants were excluded from data analysis because they reported not hearing sound during the video, and another three were excluded because they demonstrated high prior knowledge on the pretest (i.e., 8 or higher out of 13). Thus, 69 participants (33 female, average age: 37.4 years) remained in the final analysis. There were three presentation conditions to which participants were randomly assigned: (a) a control condition in

which the lesson was presented with non-identical captions and no seductive details (all non-identical/no seductive details); (b) a condition in which the lesson was presented with non-identical captions but also included seductive details (all non-identical/with seductive details); and (c) a condition in which the key facts were presented with non-identical captions and the seductive details were presented with identical captions (non-identical key facts/identical seductive details). Totals of 24 participants served in the all non-identical/no seductive details condition, 23 in the all non-identical/with seductive details condition, and 22 in the non-identical key facts/identical seductive details condition.

Materials and Procedure

Upon beginning the experiment, participants took a pretest to determine their level of prior knowledge of astronomy. This pretest was based on similar instruments used in previous studies (e.g., Mayer & Johnson, 2008) and contained one question asking participants to rate their overall level of knowledge of astronomy on a scale from 1-7, with 1 being “very low” and 7 being “very high.” The rest of the pretest consisted of six statements with which participants clicked “yes” to indicate that the statement described them or “no” to indicate that the statement did not describe them: (1) I regularly keep up with NASA or other news about space travel; (2) I know the difference between the main sequence and red giant phases of a star; (3) I know what a neutron star is; (4) I can name all the elements a star uses for fuel; (5) I can explain why a star goes supernova; (6) I can explain what causes high luminosity of a star.

After the pretest, participants listened to a brief clip of music (Mozart’s Serenade No. 13 in G) to make sure that the computer volume was adjusted to an appropriate level. They then watched the lesson, which was an animated, narrated video describing the life cycle of a star. The presentation times for the video without seductive details and the video with seductive

details were 254 and 312 seconds, respectively. The key facts in the narration consisted of 24 idea units (IUs); in addition, 7 interesting, but irrelevant, facts were included in the lessons as seductive details (see Appendix A for the complete narration and seductive details). To ensure that these experimenter-determined seductive details were, in fact, more interesting than the key facts, 22 independent participants rated each statement for interest and relevance to understanding the life cycle of a star. Each rating was on a scale from 1-7, with 1 being not at all interesting/relevant and 7 being highly interesting/relevant. The experimenter-determined seductive details were rated as significantly more interesting ($M = 5.0$, $SE = .2$) than key facts ($M = 4.4$, $SE = .3$), $t(21) = 3.5$, $p < .01$, $d = .5$, as well as less important ($M = 3.7$, $SE = .4$) than the key facts ($M = 5.0$, $SE = .2$), $p < .01$, $d = .9$.

The non-identical captions were created by slightly re-wording the narration so that it would convey the same information using generally the same sentence structure and exactly the same number of words. To illustrate, “Stars are born out of nebulae, which are clouds in space made up of dust and gas” was reworded as “Stars are created from nebulae, which are clouds made up of dust and gas in outer space.”

After watching the lesson, participants were given five minutes to take a free recall test, in which they were asked to type an explanation of the life cycle of a star in as much detail as they could remember. After five minutes, the screen moved on to the first transfer question. Participants then had two minutes to answer each of five transfer questions. Transfer questions were designed to require participants to use the information they learned in the video to make inferences about astronomy. To illustrate, one question was: “Suppose two stars of the same mass entered the main sequence phase at the same time. What could cause those two stars to enter the red giant phase at different times?” To answer this question successfully, participants

had to understand why a star moved from the main sequence to the red giant phase and what factors could alter the rate of that process (see Appendix B for all transfer questions and possible answers). For both the free recall and transfer questions, participants had to wait until the screen moved on, even if they finished typing their response before the time was up.

After the transfer test, participants completed a post-experiment questionnaire that included metacognitive and demographic questions. The metacognitive questions asked participants to rate their confidence in their ability to explain the life cycle of a star; what type of on-screen text, if any, they preferred when watching video lessons; and if they generally preferred lessons with “just the facts” or with interesting, but irrelevant, details.

We scored the pretests by summing participants’ self-rating of their knowledge of astronomy and the number of “yes” responses to the yes/no questions. Scores of 8 or higher out of a possible 13 indicated high prior knowledge of astronomy. One rater scored the free-recall test by determining how many of the 24 IUs were contained in participants’ responses. Any seductive details they typed were recorded, but not included in the total score. The same rater also scored the transfer test by assigning one point for a correct response per transfer question. To test for interrater reliability, a second rater scored 10% of the free-recall and transfer-test questions, and the correlation between the two raters was .96. Because of the high degree of interrater reliability, the first rater’s scores were used for all analyses.

Results

Retention

Free-recall test performance as a function of presentation condition are shown in Figure 2.1, and a one-way analysis of variance (ANOVA) revealed the apparent main effect of presentation condition on free recall scores to be significant, $F(2,66) = 8.19$, $MSE = .12$, $p =$

.001. To examine this effect more closely, we conducted planned comparisons with a Fisher's LSD correction for multiple comparisons. As indicated in Figure 2.1, we replicated prior findings on the seductive details effect: Overall higher performance on the retention test when the lesson containing all non-identical text did not contain seductive details ($M = .29$, $SE = .02$) than when it did ($M = .15$, $SE = .03$), $p < .001$, $d = 1.15$. Thus, as indicated would be the case by the ratings of the 22 independent participants reported in the Materials and Procedure subsection above, the experimenter-determined seductive details did, in fact, function as seductive details.

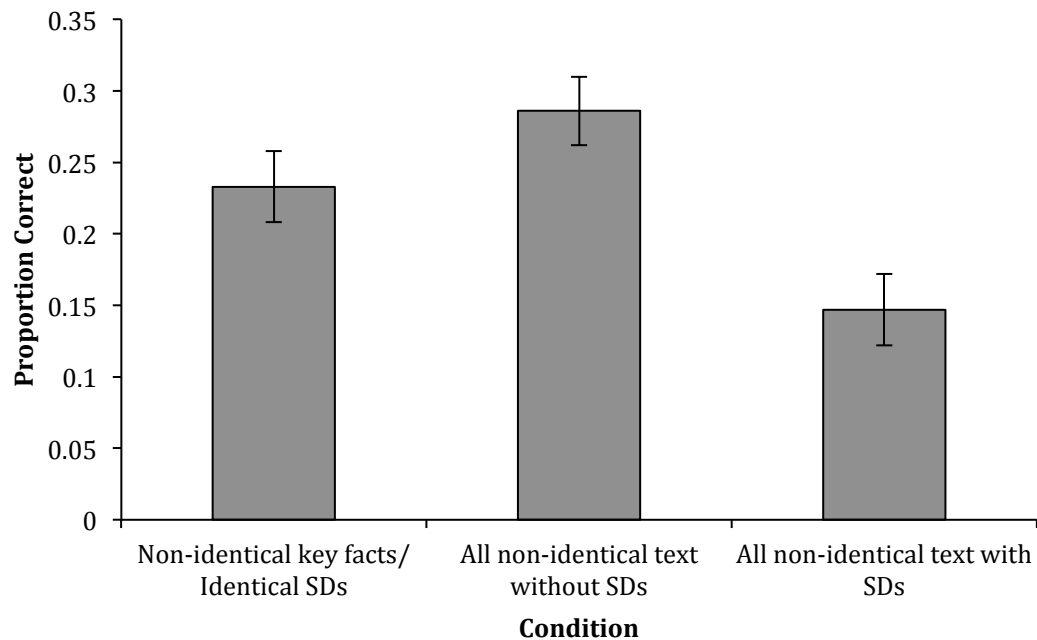


Figure 2.1. Average proportions of idea-units correctly recalled on the final free-recall test as a function of presentation condition. SDs = seductive details. Error bars represent the standard error of the mean in all figures.

The primary comparisons of interest in the present study, however, were between the performance of participants in the non-identical key facts/identical seductive details condition versus that of the participants in the other two presentation conditions. If, as we predicted, combined presentation of non-identical text with key facts and identical text with seductive

details increased effortful and deep processing of the key facts while leading to more passive processing of the seductive details, then the non-identical key facts/identical seductive details group should have scored at least as well as the group with no seductive details. And, in fact, it did—there was no difference in recall between the group without seductive details and the group that received seductive details accompanied by identical text ($M = .23$, $SE = .03$), $p = .29$. Furthermore, recall performance in the non-identical key facts/identical seductive details group was higher than that in the all non-identical text group with seductive details, $p = .02$, $d = .76$. Together then, these two comparisons suggest that the selective presentation of non-identical text with key information helps learners to overcome the typical detrimental effects of seductive details.

Because the free recall prompt asked participants to describe the life cycle of a star, rather than to type everything they remembered from the lesson, we did not expect the number of seductive details reported to differ between the groups that experienced seductive details. Altogether, 11 participants in the non-identical key facts/identical seductive details group reported any seductive details, with an average of 2.0 ($SE = .43$) seductive details per person, while 12 participants in the all non-identical/with seductive details group reported any seductive details, with an average of 1.6 ($SE = .23$) seductive details per person. This difference was not significant, $t(15) = .86$, $p = .40$. The highest number of seductive details reported by any single participant was four.

Transfer

Correct performance on the transfer test as a function of presentation condition is presented in Figure 2.2. As with the free-recall scores, we analyzed these scores using a one-way ANOVA and planned comparisons with Tukey HSD corrections. As predicted, the main

effect of presentation condition was not significant, $F(2,66) = .96$, $MSE = .09$, $p = .39$, nor did the planned comparisons revealed any significant differences between groups, all $p > .05$.

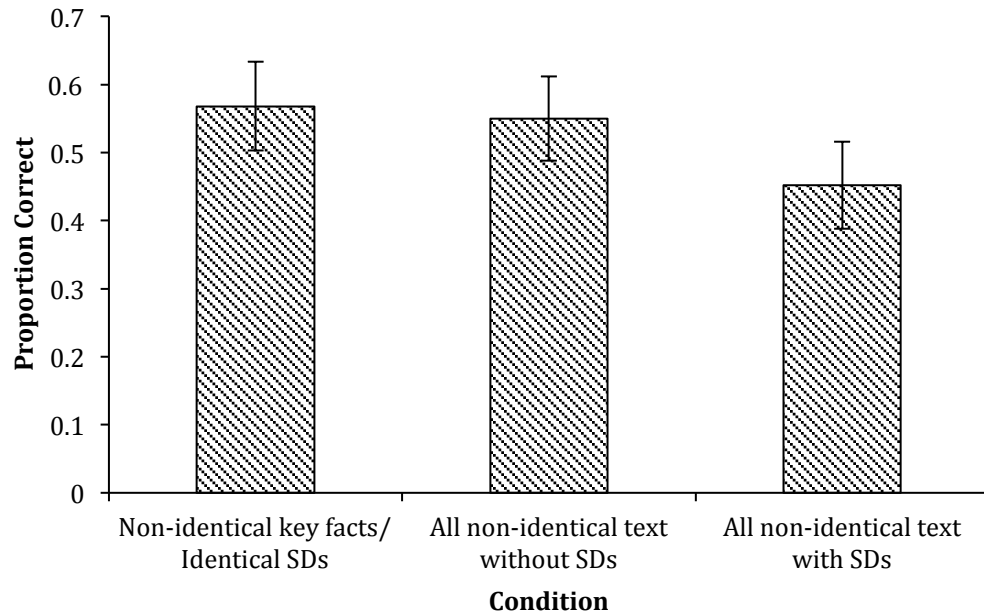


Figure 2.2. Average correct performance on the transfer test as a function of presentation condition.

Metacognitive Questions

In our post-experiment questionnaire, we asked participants to rate their confidence that they could explain the life cycle of a star to someone else. Participants who experienced the lesson without seductive details were significantly more confident ($M = 4.46$, $SE = .31$) about their ability to do so than were participants who experienced the lesson with seductive details accompanied by non-identical text (i.e., the all non-identical/with seductive details group, $M = 3.13$, $SE = .31$), $p = .01$, $d = .84$, but not significantly more confident than those participants who experienced the lesson with seductive details accompanied by identical text (i.e., the non-identical key facts/identical seductive details group, $M = 3.55$, $SE = .32$), $p = .11$.

We also asked participants what type of multimedia lesson they would prefer—one with “just the facts” or one that contained interesting, but irrelevant information. Interestingly, despite being the group with the least confidence in their ability to explain the life cycle of a star to someone else, 16 of the 23 participants experiencing the lesson with seductive details accompanied by non-identical text (or essentially 70%) stated that they would prefer a lesson with interesting, but irrelevant, information included. Overall, most participants (55 out of 69 or essentially 80%) shared this opinion. A chi square test of independence revealed that this preference was consistent across conditions, $X^2(2, N = 69) = 2.26, p = .32$.

Our final metacognitive question asked participants what type of on-screen text they would prefer if they were watching a lesson with animation and narration. Their options were: “Simultaneous on-screen text identical to the narration,” “Simultaneous on-screen text with the same information as the narration, but worded differently,” and “No on-screen text—I prefer images and narration only.” Replicating Yue et al. (2013), most participants in the present study (48 out of 69 or essentially 70%) reported that they would prefer a lesson with identical on-screen text, compared to 9 who preferred non-identical text and 12 who preferred no text at all. Even for participants who experienced both types of on-screen text (i.e., those in the non-identical key facts/identical seductive details condition), 16 preferred identical text, 5 preferred non-identical text, and 1 preferred no text. A chi-square test of independence again revealed that participants’ preferences did not depend on presentation condition, $X^2(4, N = 69) = 7.37, p = .12$.

Discussion

The goal of the present research was to determine if we could find a presentation method that would mitigate the seductive details effect while still allowing instructional materials to include such information. Our approach was to see if by using different types of on-screen text,

or different levels of redundancy of on-screen text, we could selectively induce germane processing for the key information in a multimedia lesson, but not for the processing of seductive details. We based this approach on the attention distraction and schema diversion explanations of how seductive details impair learning, as well as the explanation for the redundancy principle proposed by CLT and CTML. Measuring learning in terms of free-recall performance, our approach was successful: Presenting key facts with non-identical text and seductive details with identical text boosted recall in that condition to the level of participants who did not experience any seductive details.

One major contribution of the present research is that it illustrates that mixing levels of redundancy of on-screen text within a lesson can have differential effects on the processing of facts presented throughout that lesson. Most prior studies on the redundancy effect have employed between-subjects designs, but this study suggests that manipulating redundancy within subjects can result in variable learning outcomes. We must be cautious, however, in interpreting this result, as we only tested the key information. We designed our test to be similar to one students might encounter in a real class—that is, one that tested only key information and not irrelevant details. These types of questions may have prompted students to withhold seductive details during the free recall test, even if they did remember them; therefore, even though our results are consistent with prior findings on the redundancy effect (e.g., Kalyuga et al., 1999; Mayer et al., 2001), they do not warrant strong claims regarding how identical text affects processing of interesting extraneous details.

In addition, the present results provide instructors with an alternative to omitting all interesting extraneous information from a lesson. Both popular belief and empirical data indicate that interest can be an important motivational factor that influences student engagement and, as a

result, learning (Ainley, Hidi, & Berndorff, 2002; Moreno, 2006). Given that students do prefer a lesson with seductive details instead of a more straightforward one, instructors now have a way to present those irrelevant details without impairing retention of key information. The methods demonstrated to be useful in the present study may be a very important step, particularly in online lessons where students select what to watch and when to stop. If students will keep watching a lesson with seductive details, but not one without them, the present results provide a way to keep such details in the lesson but in a way that minimizes their negative effects on learning of the main points of the lesson.

CHAPTER 3: SINGLE-MODALITY RETRIEVAL PRACTICE IN MULTIMEDIA LEARNING

The benefits of retrieval practice are well documented with text materials (Bjork, 1988; Carpenter & DeLosh, 2006; Roediger & Karpicke, 2006a), but little research has explored the potential benefits of retrieval practice for learning materials presented in a multimedia, the focus of the present research. In the typical retrieval practice paradigm, participants initially study some material, such as a set of cue-target word pairs (e.g., cabin: mountain). Then, during a second phase of the paradigm, participants are given the cue and asked to recall the target (e.g., cabin: ?), or they are given the complete cue and target pair again (e.g., cabin: mountain). Typically, participants have better long-term memory for the items they practiced retrieving than items they only read during the second study phase (see Roediger & Butler, 2011, for a review).

The retrieval practice paradigm has been applied to educational materials as well, with similar benefits emerging (Butler & Roediger, 2007; Karpicke & Blunt, 2011; Lyle & Crawford, 2011). In the one study using a retrieval-practice paradigm in the context of multimedia learning, Johnson and Mayer (2009) showed participants an animated, narrated lesson about lightning formation after which they either re-watched the lesson, took a practice recall test on the lesson, or took a practice transfer test on the lesson. Participants who took practice tests in the second phase of the study scored better on similar delayed tests than participants who merely re-watched the lesson during that time.

Johnson and Mayer's (2009) results point to one potential explanation for the benefits of retrieval practice: transfer-appropriate processing. When the second phase requires participants to use skills that will be needed on a later criterion test, the experience that participants gain from practicing those skills benefits them on similar future tests (Roediger & Karpicke, 2006a).

Another potential explanation for the benefits of retrieval practice comes from the framework of desirable difficulties (Bjork, 1994); although taking a test is more challenging than re-reading or re-watching the to-be-learned material, being required to try to produce an answer engages cognitive processes that enhance long-term retention (McDaniel, Roediger, & McDermott, 2007).

Additionally, some evidence suggests that not only the successful retrieval of a response, but also just the search for a response, can benefit learning (Kornell, Hays, & Bjork, 2009). In order to separate attempted retrieval from successful retrieval, Kornell et al. asked participants to generate answers to real, as well as impossible, trivia questions, thus ensuring that participants could not correctly retrieve the answer for a subset of questions. They found that on a final test, participants who had tried to generate an answer to the impossible questions recalled more than participants who had only read the answer, even though participants never came up with the correct answer. These results suggest that even merely attempting retrieval engages processes that promote learning.

The dual-channel input provided by multimedia materials offers the possibility for a unique type of retrieval practice; namely, learners could re-experience either the audio or the visual portion of a presentation and simultaneously retrieve the non-presented modality. Furthermore, the act of retrieving the non-presented information and integrating it with the presented information could provide a significant benefit over simply re-studying the original, complete presentation.

Typically, the benefits of multimedia learning over single-modality learning are said to arise from the integration of audio and visual information into a single, coherent mental model (Mayer, 2005; Schnotz & Bannert, 2003). In fact, several studies have shown that separating the

initial presentation of audio and visual material results in poorer learning than presenting the audio and visual material together (Mayer & Anderson, 1991, 1992). No studies, however, have examined the effects of separately presenting the audio and visual material after an initial, integrated presentation. If we consider that the act of integrating audio and visual information from a multimedia lesson results in deep processing, and that retrieval practice also engages cognitive process that promote learning, then incorporating retrieval practice into the act of integration could very well provide a benefit to long-term retention above that of simply re-watching the lesson.

Experiment 1

Method

Participants. A total of 125 undergraduate students from the University of California, Los Angeles participated in the experiment in exchange for course credit. Four participants were eliminated due to prior knowledge, and 11 were eliminated for self-reporting that they did not comply with the retrieval practice instructions. Thus, data from 110 participants (81 female, average age: 20.4 years) are included in the data analyses reported here.

Materials. The pretest, administered to measure participants' level of prior knowledge, was composed of two parts: (1) four specific questions about astronomy, such as "What causes high luminosity of a star?" and (2) a list of 13 astronomical terms (e.g. "nebula" and "brown dwarf"), with participants asked to indicate which ones they recognized and which ones they could define. Because the present study was designed to assess how novices learn, participants who correctly answered three or four questions in the first part, or who indicated that they could provide definitions for seven or more terms in the second part, were eliminated from data analysis.

The lesson consisted of a narrated, animated PowerPoint slide show illustrating 24 facts or idea units about the life cycle of a star, which are listed in Appendix A. Participants listened to the narration via headphones while they viewed the lesson on a 21.5 in iMac computer screen, with the presentation of the slides paced by the computer.

The final test was also delivered via the computer, and participants typed their answers on the keyboard. First, participants were given a free-recall prompt (“Please describe the life cycle of a star in as much detail as you can remember.”). After completion of that task, six open-ended transfer questions were asked to assess learners’ abilities to apply their knowledge to novel situations related to the lesson (e.g., “Why is it possible for medium- and high-mass stars to go supernova?”). All transfer questions, along with samples of acceptable responses, are listed in Appendix B. Finally, in a post-experiment questionnaire, participants answered metacognitive questions assessing their perceived interest, understanding, and preferred learning conditions, and they also answered questions about the overall interestingness and clarity of the lesson.

Design and procedure. Participants were randomly assigned to one of four between-subjects conditions: study-once, study-twice, retrieve-narration, and retrieve-animation. First, all participants were given the pretest, after which they put on headphones and watched and listened to one complete presentation of the lesson on the computer. After watching the same presentation once, the groups then engaged in different tasks. The study-once group was immediately given the final test. The study-twice group was instructed to keep their headphones on and to watch the same presentation again, after which this group was immediately given the final test. The retrieve-narration group was told that they would see only the animation again, but as they watched, they should try to remember what the narration had been at each point in the

animation. Then, after watching the animation again, this group was immediately given the final test. Similarly, the retrieve-animation group was told that they would listen to only the narration again, but they should try to vividly recall the animation that went along with the words as they listened. Then, after listening to the narration again, this group was immediately given the final test. As part of the post-experiment questionnaire, participants in the retrieval practice conditions were asked if they had complied with the retrieval instructions (“If you were asked to remember the narration or envision the animation, did you do so?”), as well as indicating how successful they had been at recalling the narration or animation (“How successful were you at remembering the narration/envisioning the animation during the second presentation?”) on a scale from 1-5. They were encouraged to answer honestly and were reminded that their participation credit was not dependent on their responses. Participants also reported demographic data (i.e., age, gender, major, and primary language).

Results

Scoring. Participants with high prior knowledge, as measured by their score on the pretest, were eliminated. Two raters independently scored participants’ responses to the free recall and transfer questions. The free recall question was scored out of 24 points, with participants receiving one point for each fact or idea unit they correctly recalled. The idea unit did not have to be recalled verbatim; rather, any response that indicated the participant understood the main idea of the fact was counted as being correct. To illustrate, for the idea unit, “Stars are born in nebulae, which are clouds in space made up of dust and gas,” a correct answer might be: “Stars start as clouds of dust and gas.” Each transfer question received one point for a correct answer, and zero points for an incorrect answer. In scoring these questions, no partial credit was given nor any extra credit for providing more than one correct response, although

there were multiple possible correct answers to each question. Interrater reliability was .88, and discrepancies were discussed and resolved.

Compliance with retrieval practice instruction. Based on participants' response to the question of whether they recalled the animation or narration when they were asked to do so, we were able to determine which participants engaged in retrieval practice. Out of the 46 participants taking part in the two retrieval practice conditions, 11 reported that they had not followed instructions to recall the absent modality. Thus, to determine if retrieval practice had affected learning, we conducted two separate t-tests comparing the recall and transfer scores of participants who complied and participants who did not comply. Participants who complied with the retrieval practice instructions performed significantly better ($M = .37, SE = .03$) than those who did not ($M = .16, SE = .04$) on the recall test, $t(44) = 3.58, p = .001, d = 1.3$. Similarly, participants who complied with the retrieval practice instructions performed significantly better ($M = .46, SE = .05$) than those who did not ($M = .24, SE = .07$) on the transfer test, $t(44) = 2.18, p = .04, d = .8$.

Because complying with the retrieval practice instruction had a clear effect on learning in these conditions, we eliminated participants who reported that they did not attempt retrieval practice and ran replacement subjects until the number of participants in each retrieval practice group was approximately the same as that in each study group, thus allowing a more valid comparison of the effects of single-modality retrieval practice with the effects of re-studying.

Pretest. After eliminating participants with high prior knowledge in accordance with the preset criteria for the pretest (i.e., earning more than 2 points on Part 1 or more than 7 points on Part 2), the average score on Part 1 of the pretest was 0.37 out of a possible 4 points. A one-way ANOVA with condition (study-once vs. study-twice vs. retrieve-narration vs. retrieve-animation)

as the independent variable showed that prior knowledge did not statistically differ across the conditions, $F(3,106) = .71$, $MSE = .26$, $p = .55$. The average score on Part 2 of the pretest was 1.4, indicating that participants were, on average, prepared to provide a definition for fewer than 2 of the 13 terms listed. As expected, a one-way ANOVA revealed that there were no significant differences in the Part-2 scores among conditions, $F(3,106) = .23$, $MSE = .62$, $p = .87$. In sum, then, participants had generally the same level of prior knowledge across conditions.

Recall. As indicated in Figure 3.1, which illustrates the correct recall proportions for each condition, the retrieve-animation group appeared to perform best on the recall test ($M = .38$, $SE = .03$), followed by the study-twice group ($M = .35$, $SE = .03$), the retrieve-narration group ($M = .33$, $SE = .04$), and the study-once group ($M = .30$, $SE = .03$), respectively. Despite these numerical differences, however, a one-way ANOVA with condition (study-once vs. study-twice vs. retrieve-narration vs. retrieve-animation) as the between-subjects factor did not reveal a significant main effect of condition on recall, $F(3,106) = 1.12$, $MSE = .04$, $p = .35$.

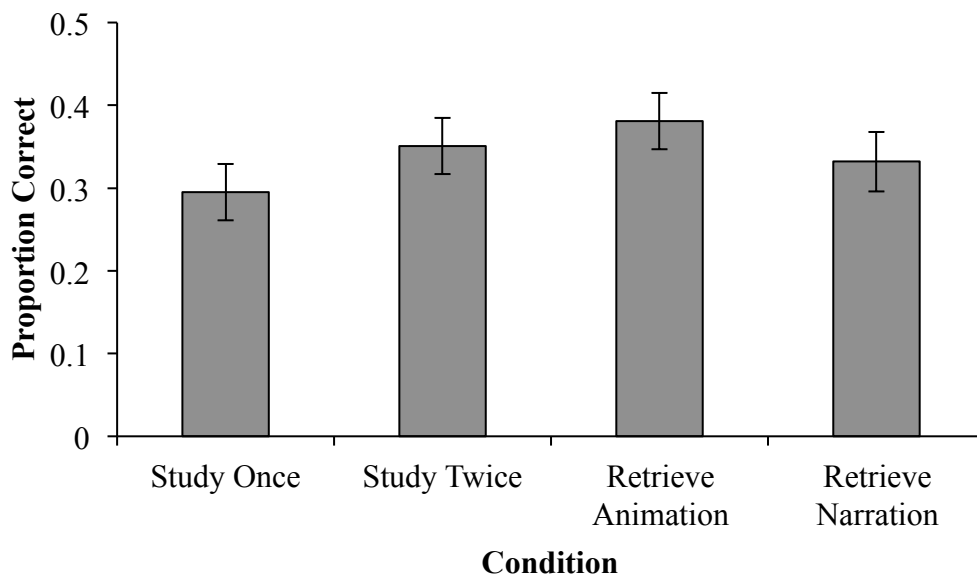


Figure 3.1. Average proportions of idea-units correctly recalled on the final free recall test for each condition in Experiment 1.

Transfer. As indicated in Figure 3.2, the retrieve-narration group appeared to perform best on the transfer test ($M = .51$, $SE = .06$), followed closely by the study-twice group ($M = .48$, $SE = .06$), the retrieve-animation group ($M = .42$, $SE = .06$), and the study-once group ($M = .35$, $SE = .06$), respectively. Again, despite these numerical differences, a one-way ANOVA did not reveal a significant main effect of condition on transfer performance, $F(3,106) = 1.47$, $MSE = .13$, $p = .23$.

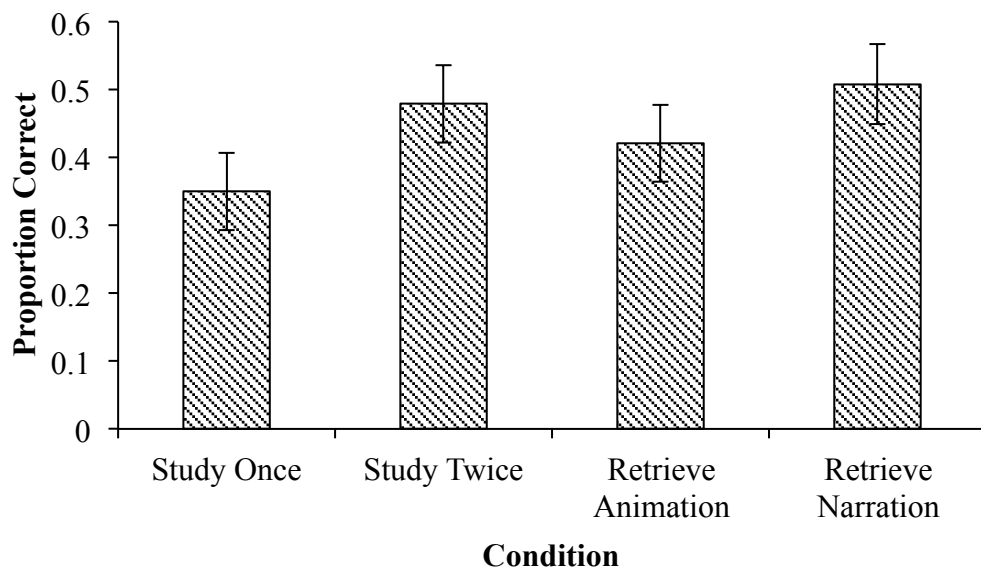


Figure 3.2. Average correct proportions on the transfer test for each condition in Experiment 1.

One of the questions in the post-experiment questionnaire for participants in the retrieval conditions had been to rate how vividly their recall had been on a scale from 1-5, with 1 being “Not at all” and 5 being “Extremely vivid.” We thus used these ratings to explore a possible relation between strength of retrieval practice and learning for those participants who reported having complied with the retrieval practice instructions. Unfortunately, we did not have this information for all of the retrieval-practice participants, as we only began to include this question

about half way through data collection. For participants who had complied with instructions and who had been asked to make this rating, however, we conducted a correlational analysis between their vividness ratings and their recall and transfer performance. In calculating this correlation, we combined the retrieve-animation and retrieve-narration participants into one group because a *t*-test comparison had shown that the ratings between these two groups did not differ, $t(31) = .84$, $p = .41$. Our expectation was that stronger retrieval practice would positively correlate with test scores. And, in fact, it did—there was a significant positive correlation between the self-reported strength of participants' retrieval and their recall, $r = .57$, $p = .001$, as well as between retrieval strength and transfer performance, $r = .52$, $p = .002$.

Metacognitive judgments. As part of the post-experiment questionnaire, participants were asked to state which condition of learning they would have preferred. These judgments were collected by first reminding each participant of what had happened in the condition they had experienced and then describing to them what had happened in the other three conditions of the experiment. As can be seen in Figure 3.3, which shows the stated preferences as a function of the condition that participants had actually experienced, most participants (74 out of 110) reported that they believed that the study-twice condition would be better for learning. The retrieve-animation condition was the next most preferred, with 30 participants selecting that option. Only six participants believed that the retrieved-narration condition would be best, and no participants selected the study-once condition. A chi-square test of independence indicated that preference did not vary by group, $X^2(6, N = 110) = 10.18$, $p = .12$.

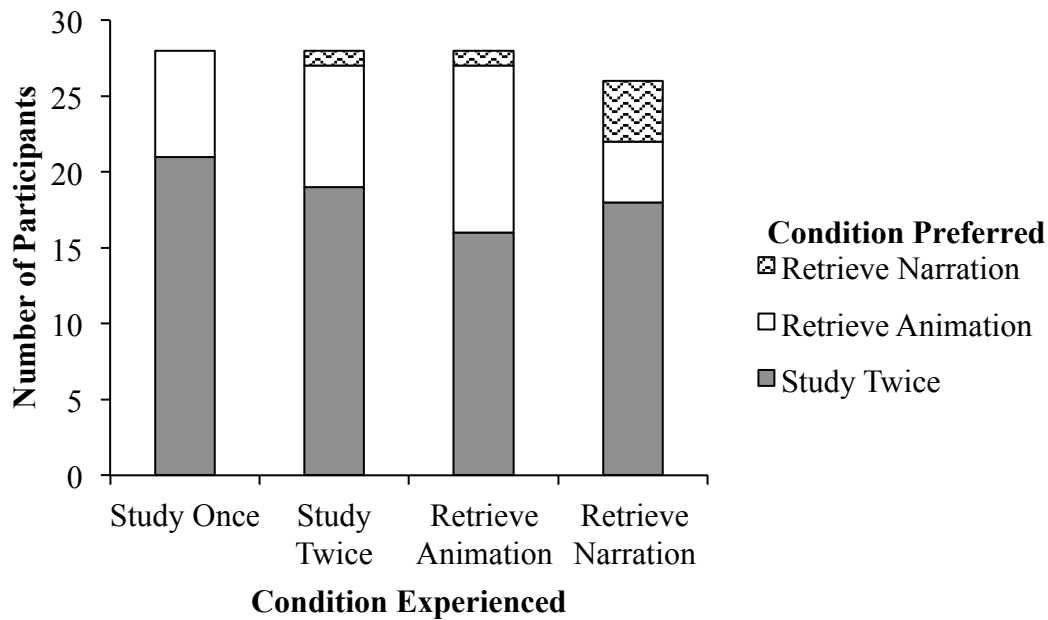


Figure 3.3. Stated condition preferences by participants in Experiment 1 shown as a function of the condition they experienced.

Discussion

We compared the effects of two types of single-modality retrieval practice on learning to those obtained by studying a multimedia lesson in its entirety either once or twice. Contrary to our expectations, we did not find that single-modality retrieval practice had a learning benefit over studying an entire multimedia lesson twice as measured by final recall and transfer test performance. Surprisingly, we also did not see a significant benefit for studying the presentation twice over studying it once. The similarity between the scores in the study-once and study-twice conditions, however, may be due, in part, to the immediacy of the restudy—the study-twice condition was a type of massed practice, which has been shown with text material to have no advantage over studying that material once (e.g., Callender & McDaniel, 2009).

Although we had expected retrieval practice to benefit learning, another possible outcome could have been that experiencing only one modality of the multimedia lesson would lead to worse learning in those conditions. It could, for example, be argued that participants in the

retrieval practice conditions had fewer opportunities to study the material, so they were at a disadvantage compared to the study-twice group. Interestingly, however, re-exposure to only one modality of the multimedia presentation did not harm learning compared to studying the entire presentation twice. It is possible, then, that participants who engaged in some manner of retrieval practice when they were instructed to do so did experience a learning benefit as a result of that activity, but it was not enough of a benefit to emerge as a significant improvement in their performance.

Related to this possibility, it is important to note that participants in the present study only engaged in covert retrieval, and the strength of that retrieval may have varied a great deal across participants in the retrieval practice groups. It may be that if retrieval practice could be strengthened across participants, any mnemonic benefits of that practice would be more evident on the final test. The positive correlations observed between self-reported retrieval strength and test scores provide some support for this conjecture. Therefore, in an effort to bring all participants up to a strong level of retrieval, we asked participants in Experiment 2 to engage in overt retrieval practice of the non-presented modality.

Experiment 2

In Experiment 1, we demonstrated that covert retrieval practice of a non-presented modality of a multimedia lesson did not differentially affect learning compared to studying the whole presentation either once or twice. Although some prior research has indicated that covert and overt retrieval practice lead to equal mnemonic benefits (e.g., Smith, Roediger, & Karpicke, 2013), those studies were conducted with simple word lists. It may be that more complex educational materials, such as a multimedia lesson about the life cycle of a star, would be more sensitive to differences created by overt and covert retrieval practice. In addition, the present

Experiment 1 provided evidence of a positive correlation between strength of retrieval practice and final test score, suggesting that strengthening retrieval practice overall might lead to a boost in learning. Therefore, in Experiment 2, participants overtly retrieved the non-presented modality by writing (in the retrieve-narration condition) or drawing (in the retrieve-animation condition). Additionally, we removed the study-once condition and added a second lesson with a week-delayed test in order to assess possible effects of retrieval practice after a longer retention interval, as benefits of retrieval practice tend to emerge after longer delays (e.g., Roediger & Karpicke, 2006a).

Method

Participants. A total of 53 undergraduates from the University of California, Los Angeles participated in exchange for extra credit in psychology courses. Five participants were eliminated due to high prior knowledge, leaving 48 participants (35 female, average age: 20.6) of whom 18 were in the study-twice condition, 13 in the retrieve-animation condition, and 17 in the retrieve-narration condition.

Materials. The same lesson about the life cycle of a star was used in Experiment 2. In addition, a similar lesson about lightning formation (170 s in duration) was included. Similar to the astronomy lesson, the lightning lesson was created in PowerPoint, but the narration and images were modeled after the lesson used in Mayer, Heiser, and Lonn (2001), and the narration in the present lightning lesson was identical to the base narration in that study. The pretest was modified to assess knowledge of meteorology in addition to astronomy based on similar instruments used in previous studies. The pretest contained one question asking participants to rate their overall level of knowledge of astronomy on a scale from 1-7, with 1 being “very low” and 7 being “very high,” and one question asking them to rate their overall level of knowledge of

meteorology on the same scale. The rest of the pretest consisted of twelve statements (six about astronomy, six about meteorology) to which participants clicked “yes” or “no” to indicate whether the statement did or did not, respectively, describe them. The six astronomy statements were: (1) I regularly keep up with NASA or other news about space travel, (2) I know the difference between the main sequence and red giant phases of a star, (3) I know what a neutron star is, (4) I can name all the elements a star uses for fuel, (5) I can explain why a star goes supernova, and (6) I can explain what causes high luminosity of a star; and the six meteorology questions were: (1) I regularly read weather maps in the newspaper or online, (2) I can distinguish between cumulus and nimbus clouds, (3) I know what a low pressure system is, (4) I can explain what makes the wind blow, (5) I can draw the symbol for a cold front, and (6) I can draw the symbol for a warm front. Each “yes” response received one point and was added to the participant’s overall rating of their knowledge of that subject. Scores of 8 or higher (out of a maximum of 13) were considered high prior knowledge, and subjects earning those scores were eliminated from analyses.

The recall and transfer tests for the lightning lesson were similar to the recall and transfer tests for the star lesson. Appendix C contains the narration of the lightning lesson, and Appendix D contains the transfer questions and possible answers for that lesson.

Design and procedure. The procedure was similar to that of Experiment 1. Participants first took a pretest to assess their knowledge of astronomy and meteorology and then watched and listened to one of the lessons. Next, participants assigned to the study-twice condition watched and listened to the entire lesson again. Participants in the retrieve-animation group were given a pencil and a packet of paper divided into three sections per page with each section corresponding to a different slide. The participants were instructed that they would be listening

to the same narration again, but would be asked to draw the animation that they remembered seeing during that narration in the appropriate section. Each slide number appeared on the computer screen while the narration was playing so that participants would know the appropriate section in which to draw in the accompanying animation. After the narration for each slide played once, the presentation paused and the following instruction appeared on the screen: “Please hit the spacebar when you are ready to continue.” By allowing participants to move through this phase at their own pace, we intended to avoid rushing them and to encourage them to engage in detailed retrieval practice.

Participants in the retrieve-narration condition received an identical packet, but they were instructed that they would be seeing the animation again, and they should write the narration they remembered hearing with that animation in as close to the original words as they could recall. While the animation was playing, each slide number appeared in the top-right corner of the screen to help participants stay on track. After the animation was completed, the screen went black and, just as in the retrieve-animation condition, participants were allowed to proceed to the next slide at their own pace.

After the retrieval phase (or restudy phase, in the case of the study-twice group), all participants were given a recall and transfer test on the first lesson. They then watched and listened to the second lesson and engaged in the same activity for the second phase of this lesson. For example, if they had studied the first lesson twice, they also studied the second lesson twice; if they had engaged in retrieval practice for the first lesson, they engaged in the same type of retrieval practice for the second lesson. After this phase, participants were released and instructed to come back the following week. We sent a reminder email to participants the day before they were supposed to return, and when they did so, they were given the recall and

transfer tests on the second lesson. As in Experiment 1, all participants were given the same type of post-experiment questionnaire at the end of their participation in the experiment. Which lesson appeared as the first and second lesson was counterbalanced across participants. Thus, Experiment 2 employed a 2 (test time: immediate vs. delayed) x 3 (condition: study-twice vs. retrieve-animation vs. retrieve-narration) mixed-subjects design, with test time as the within-subjects factor and condition as the between subjects factor.

Results

Scoring. The recall and transfer test for the star materials were scored in a similar manner as in Experiment 1. Two raters scored data from the first 23 subjects. Interrater reliability was .93. Discrepancies were discussed and resolved, and one rater scored the remainder of the data.

The recall and transfer tests for the lightning materials were scored in a similar manner. Participants received one point for each idea unit produced on the free recall test, with a highest possible score of 16. Participants received one point for providing a correct answer for each transfer question, with a highest possible score of 4. Interrater reliability was .93, and discrepancies were discussed and resolved.

Retrieval practice performance was based on the number of idea units recalled or the number of correct images drawn for the retrieve-narration and retrieve-animation conditions, respectively. Much like the final recall test, retrieval practice was scored leniently, meaning that participants only had to recall the general idea of the narrated segment to receive credit. To measure recall of the animation, we counted the number of images in each segment. For example, the lightning lesson contained a total of 131 possible images across 16 slides. The first slide pictured a clipart-style image of a house, grass, water, and wavy lines depicting a cold

front. Each of these images was worth one point. Participants were also expected to draw some indication that the cold front moved from over the water to over the ground, which was another point. Thus, there were five possible points to earn on the first slide of the lightning lesson. The number of possible images to draw varied across slides.

The star drawings were scored similarly. Because the animation was simpler than that in the lightning lesson, there were only 42 possible points. For example, one slide contained a star, a set of inward arrows to indicate gravity, and a set of outward arrows to indicate gas pressure. Since each outward arrow represented the same thing, participants received one point if they drew any number of outward arrows—i.e., four arrows meant one point, not four.

Pretest. First, we eliminated participants with high prior knowledge in accordance with the preset criteria (i.e., earning more than 8 out of 13 points on either the meteorology or astronomy pretest). The average meteorology pretest score was 2.25 ($SE = .21$), and the average astronomy pretest score was 3.27 ($SE = .28$). A one-way ANOVA with condition (study-twice vs. retrieve-narration vs. retrieve-animation) as the independent variable revealed that prior knowledge of meteorology was not statistically different across conditions, $F(2,45) = .02$, $MSE = .03$, $p = .99$. Similarly, prior knowledge of astronomy was also not found to differ across conditions, $F(2,45) = .11$, $MSE = .41$, $p = .90$.

Retrieval practice performance. To ensure that participants were equally successful on the retrieval practice portion of the lessons, and to ensure that neither retrieval practice condition differed in the amount of success participants had in their practice, we conducted a 2 (condition: retrieve-animation vs. retrieve-narration) x 2 (lesson content: star vs. lightning) mixed-subjects ANOVA, with condition as the between-subjects factor and lesson content as the within-subjects factor. This analysis revealed the following outcomes: (a) no significant main effect of condition

on retrieval practice performance, $F(1,28) = .26$, $MSE = .01$, $p = .61$, indicating that participants were equally able to retrieve the narration ($M = .81$, $SE = .03$) as they were the animation ($M = .79$, $SE = .03$); (b) a significant main effect of lesson content, $F(1,28) = 5.01$, $MSE = .08$, $p = .03$, with participants successfully retrieving significantly more from the lightning lesson ($M = .83$, $SE = .02$) than from the star lesson ($M = .76$, $SE = .03$); and (c) no significant interaction between lesson content and condition, $F(1,28) = 1.14$, $MSE = .02$, $p = .30$. Because no interaction occurred, and the order of the lessons was counterbalanced, it seems highly unlikely that the advantage to the lightning retrieval practice affected other patterns of results.

Recall. We performed a 2 (test time: immediate vs. delayed) x 3 (condition: study-twice vs. retrieve-animation vs. retrieve-narration) mixed subjects ANOVA, with test time as the within-subjects factor. As revealed by this analysis and indicated in Figure 3.4, which shows correct recall proportions on the immediate and delayed tests for each condition, there was no interaction between test time and condition, $F(2,45) = .55$, $MSE = .02$, $p = .58$. Unsurprisingly, participants performed significantly better on the immediate test ($M = .60$, $SE = .04$) than on the delayed test ($M = .36$, $SE = .03$), $F(1,45) = 39.77$, $MSE = 1.39$, $p < .001$, $d = .98$. Furthermore, consistent with the results of Experiment 1, but contrary to our expectations, there was no effect of condition, $F(1,45) = .04$, $MSE = .01$, $p = .96$.

Transfer. Just as with the recall data, we performed a 2 (test time: immediate vs. delayed) x 3 (condition: study-twice vs. retrieve-animation vs. retrieve-narration) mixed subjects ANOVA, with test time as the within-subjects factor. As revealed by this analysis and indicated in Figure 3.5, which illustrates the correct performance on the transfer tests for each condition, a pattern of results very similar to that seen in the recall data was obtained. Namely, there was no interaction between condition and test time, $F(2,45) = .25$, $MSE = .03$, $p = .78$, and participants

performed significantly better on the immediate test ($M = .45, SE = .05$) than on the delayed test ($M = .30, SE = .04$), $F(1,45) = 4.86, MSE = .49, p < .03, d = .48$. Just as with the recall results, there was no effect of condition on transfer performance, $F(1,45) = .48, MSE = .03, p = .62$.

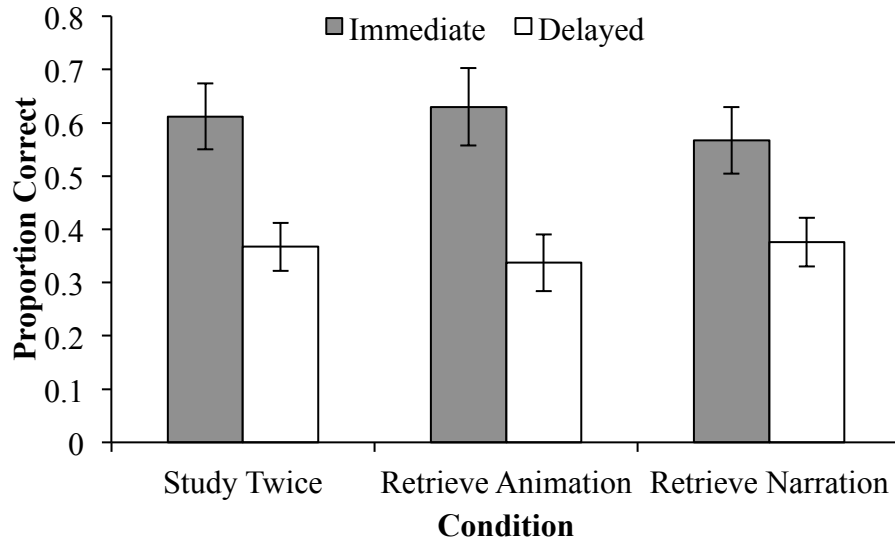


Figure 3.4. Proportion correct on immediate and delayed recall tests for each condition in Experiment 2.

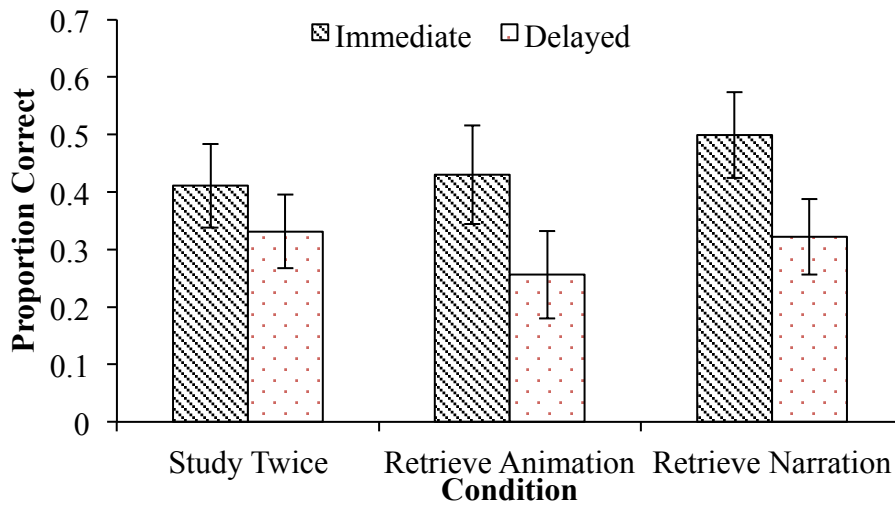


Figure 3.5. Proportion correct on immediate and delayed transfer tests for each condition in Experiment 2.

Metacognitive judgments. As in Experiment 1, participants were asked to state in which condition they would have preferred to participate following completion of their participation in Experiment 2, and Figure 3.6 shows these stated preferences as a function of the condition experienced by the participant. Most people (31 out of 48) stated that they would prefer to study the whole lesson twice, rather than have to try to recall part of it during the second presentation. Eight people said they would prefer the retrieve-animation condition, and nine people said they would prefer the retrieve-narration condition.

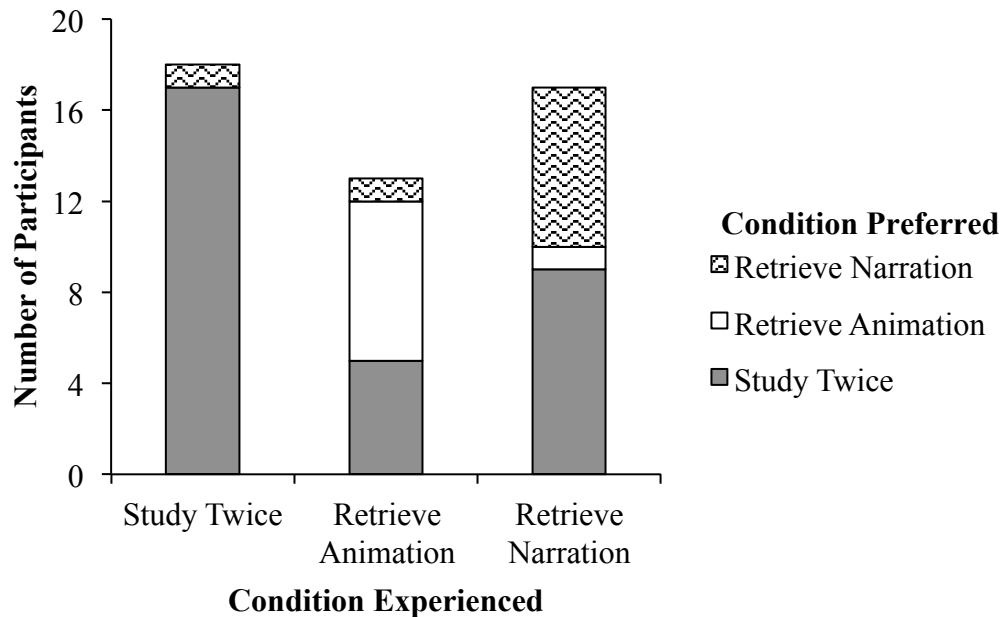


Figure 3.6. Stated condition preferences by participants in Experiment 2 shown as a function of the condition they experienced.

As in Experiment 1, we conducted a chi square test of independence to see if condition experienced affected condition preferred. And, contrary to our findings in Experiment 1, it did—participants in either retrieval practice condition were more likely to endorse the condition they had experienced than the other retrieval practice condition, $X^2(4, N = 48) = 26.26, p < .001$. Seven out of the eight people who preferred retrieve-animation had been in that condition, and

seven out of the nine people who preferred retrieve-narration had been in that condition. Thus, it would seem that experiencing some form of retrieval practice led students to believe it had been somewhat helpful, even if test scores did not reflect a difference between conditions.

Discussion

In Experiment 2, we expected that requiring participants to engage in overt retrieval practice would improve their performance on the final test. Similar to our findings in Experiment 1, however, we did not find a difference in performance between the conditions requiring retrieval practice and the study-twice condition on either immediate or delayed tests. It is particularly surprising that participants in the retrieve-narration condition did not experience a greater benefit on the recall test than their study-twice or even retrieve-animation counterparts, as the act of retrieving the narration so closely paralleled the requirements of the final recall test. It is possible, however, that by segmenting the retrieval practice, we inadvertently encouraged participants to think of the information from the lesson as a set of isolated pieces, rather than a single, integrated lesson. This type of separation may have made it less likely that we would see a benefit on the final free recall test or, especially, the transfer test.

For example, during the retrieval practice phase, participants may have engaged in item-level processing in an effort to recall the verbal or visual elements of the lesson corresponding to each slide. Although we did not give participants specific instructions regarding the level of detail they should strive for during retrieval practice, an examination of their retrieval packets indicates that they attempted to provide as much detail as they could remember. For instance, participants who recalled the narration tended to do so in sentence form, rather than bullet points (although there were certainly exceptions). Participants who retrieved the animation tended to draw more detail than was strictly necessary; for example, they would draw the grass on the

ground, even when the primary lesson content was occurring in the sky (e.g., water droplets forming into a cloud). Possibly, then, rather than helping participants to create an enriched mental model of the situation, trying to recall such details distracted them from integrating and understanding the key components and progression of the lesson.

Similar to Experiment 1, most participants stated that they would prefer to see an entire lesson twice. This finding is consistent with prior work indicating that participants prefer to reread rather than take a test (Kornell & Bjork, 2007). Interestingly, we also found that participants in the retrieve-animation or retrieve-narration conditions were more likely to endorse retrieval practice—specifically, the type of retrieval practice they had experienced—as a learning tool than those who watched the lesson twice in its entirety. This preference indicates that some participants believed that retrieval practice had helped them on the final test, even though their test scores did not coincide with that belief.

General Discussion

The primary motivation for the present series of experiments was to investigate whether single-modality retrieval practice would facilitate the integration of auditory and visual information from a multimedia lesson, thus improving recall and transfer. Previous work has demonstrated that taking a test, compared to restudying material, results in better performance on a later, similar test (e.g., Butler & Roediger, 2007; Johnson & Mayer, 2009). In addition, the effortful integration of visual and auditory information is thought to be one of the reasons that multimedia materials often result in greater learning than purely visual or purely verbal materials (Schnotz & Bannert, 2003). For these reasons, we conjectured that asking participants to generate the non-presented modality during retrieval practice would encourage the type of deeper processing that aids long-term retention and transfer.

In contrast to these expectations, however, we found that neither covert nor overt retrieval practice of the non-presented resulted in improved performance compared to restudying the entire lesson. One possible reason for our results is that our retrieval practice phase was too similar to restudy. Previous research has demonstrated that more difficult retrieval, as measured by response latency on an initial test, tends to lead to greater long-term learning than easier retrieval (e.g., Benjamin, Bjork, & Schwartz, 1998; Gardiner, Craik, & Bleasdale, 1973). The high scores on the retrieval practice phase in Experiment 2—approximately 80% for both animation and narration—suggest that the retrieval practice required may have been too easy. Future studies could test this explanation in two ways: (1) by increasing the interval between initial learning and retrieval practice, or (2) by removing the cue.

Delaying initial retrieval has been shown to be an important factor in promoting long-term learning via retrieval practice (Karpicke & Roediger, 2007). It is possible, then, that asking participants to engage in retrieval practice immediately after the initial learning phase did not encourage the type of processing that would have occurred with a longer delay.

On the other hand, it could be argued that Johnson and Mayer (2009) did find long-term benefits of retrieval practice when they used an immediate initial test. In that study, participants performed better on a week-delayed recall test if they had taken a recall test right after watching a multimedia lesson about lightning formation. This paradigm would be most similar to our retrieve-narration condition, for which we did not see a benefit. The main difference between these two paradigms is that Johnson and Mayer used a free recall test during the retrieval practice phase, whereas our recall test provided a cue in the form of the non-presented modality. If participants in the present study had already succeeded in integrating the visual and verbal information, then providing the animation during retrieval practice may have simplified retrieval

to the level of restudy. Therefore, future studies may wish to require participants to engage in retrieval practice without the aid of a cue, or increase the delay between initial learning and retrieval practice, or both.

We also investigated participants' preferences regarding restudy options. Consistent with our expectations, most participants in both experiments stated that they would prefer to study the lesson with both animation and narration twice, rather than engage in retrieval practice for one modality during the second study period. This finding is consistent with many other studies suggesting that learners prefer more fluent restudy options, such as rereading compared to testing (e.g., Kornell & Bjork, 2007). Interestingly, we also found that overt retrieval practice did lead some participants to believe that it had been helpful, and to make the judgment that they would prefer to use that type of study technique if given the option. This apparent shift in opinion about the benefits of retrieval practice indicates that participants were sensitive to the potential benefits of retrieval practice, although those benefits did not emerge in the present research. Possibly, participants felt that retrieval practice offered them a practice test of sorts, leading them to think that they were thus more confident when taking the final test than they would have been had they participated in the study-twice condition instead.

Although we did not observe benefits owing to single-modality retrieval practice on the recall and transfer tests administered in the present research as we had expected, it is still interesting that performance was not harmed, as some might have expected, when only one modality was presented during restudy. Furthermore, that most participants who were required to overtly engage in retrieval practice stated that they would prefer doing so compared to watching and listening to the complete lesson again suggests that experience with retrieval practice may make learners aware of the possible benefits of such a strategy, even with complex

materials. The present pattern of results thus encourages further exploration of conditions under which single-modality retrieval would have some learning advantages over complete representations of multimedia lessons, and how learners can gain awareness of beneficial study conditions.

CHAPTER 4: CUED-RECALL TESTS CAN POTENTIATE SUBSEQUENT STUDY OF A MULTIMEDIA LESSON

For the typical student and educator, the primary function of tests is to serve as tools for assessing learning. Tests, however, can also serve as powerful tools for learning (for reviews, see Roediger & Butler, 2011; Roediger & Karpicke, 2006b). The act of retrieving an item from memory (e.g., answering a test question) increases the probability that the item can be recalled on a later test, thus conferring a direct benefit of testing (Bjork, 1975). In addition, retrieval—even attempted retrieval—can enhance subsequent learning, an effect known as *test-potentiated learning*. Izawa (1966) was the first to note this phenomenon, and she and other researchers have examined this effect in more detail (e.g., Arnold & McDermott, 2013; Izawa, 1969, 1970, 1971; LaPorte & Voss, 1974; Soderstrom & Bjork, 2014; Wissman, Rawson, & Pyc, 2011).

Test-potentiated learning is especially relevant in educational contexts. Throughout a course, for example, students may take quizzes that cover a subset of the information that will be tested on a final exam. After taking these quizzes, students most likely restudy the quizzed material, as well as related information (e.g., material that was not on the initial test, but was included in the same chapter as the tested material) before taking the final exam. In such cases, it is not only the direct benefits of retrieval stemming from taking the initial quizzes that are relevant for learning, but also how that retrieval influences the effectiveness of subsequent restudy.

Several studies have shown that testing potentiates restudy of the same material (e.g., Arnold & McDermott, 2013; Pyc & Rawson, 2010, 2012; Soderstrom & Bjork, 2014) as well as subsequent study of new material (e.g., Szpunar, McDermott, & Roediger, 2008; Szpunar, Khan, & Schacter, 2013; Wissman et al., 2011). Arnold and McDermott, for example, had participants

study a set of Russian-English word pairs. In between restudy periods, participants took either one or five interim tests, without feedback, on the word pairs. Their results showed that participants learned more new (i.e., previously unrecalled) pairs during restudy when they had taken five tests than when they had taken only one.

Although Arnold and McDermott's (2013) primary goal was to disambiguate the indirect benefits of test potentiation from the direct benefits of testing, other researchers have speculated on the means by which test potentiation confers its benefits. One possibility is that interim tests reduce mind-wandering in subsequent learning situations (Szpunar et al., 2013). Of more direct relevance to the present study is the argument that tests improve metacognitive awareness; that is, tests give learners insight into what they do and do not know—knowledge they may use as a guide for future study (e.g., Finn & Metcalfe, 2007; Soderstrom & Bjork, 2014). Similarly, Wissman et al. (2011) suggested that the experience of a test encourages learners to engage in more effective encoding strategies in their subsequent study opportunities, which would imply that testing allows learners to gain metacognitive awareness of their previous encoding strategies and the necessity to improve them.

Beyond the experience of a test, the composition of the test may be an important consideration as well. Most previous studies examining test potentiation have used an initial test that covered all the material to be restudied and recalled on the final test (e.g., Arnold & McDermott, 2013; Izawa, 1969; Soderstrom & Bjork, 2014). Initial tests that include only some of the information that will be on the final test, however, appear to have mixed effects. Importantly, there is a large body of literature indicating that this type of partial testing can impair later retrieval of untested information, a phenomenon known as *retrieval-induced forgetting* (see Anderson, Bjork, & Bjork, 1994). Studies using more integrated educational

materials, however, have shown the opposite effect—namely, that testing can enhance the later retrieval of untested, related information (Chan, 2009; Little, Bjork, Bjork, & Angello, 2012; but see Little, Storm, & Bjork, 2011). This latter phenomenon is known as *retrieval-induced facilitation* (Chan, McDermott, & Roediger, 2006), and one proposed explanation for its occurrence is that when the to-be-learned information is highly integrated, retrieving multiple pieces of that information leads to automatic spreading activation or strategic retrieval of untested, related pieces of information (Chan, 2009).

Studies demonstrating retrieval-induced facilitation, however, have not examined what happens to the untested, related information when participants are able to restudy the original material after the initial test, but before the final test—as they would, presumably, be likely to do in an educational setting. In the context of learning classroom material, for example, it seems possible and even likely that after taking a test on some of the information, students would be encouraged to focus primarily on the previously tested information during restudy; that is, learners may believe that because information was on the initial test, it is the most important information from the lesson. Any shift to an improved encoding strategy, therefore, may only benefit retention of the previously tested information. On the other hand, even a test on some of the information could inform learners of their state of knowledge of the lesson content more broadly, thus encouraging them to restudy all the information more effectively, rather than only the previously tested information.

The present study

The goal of the present study was to explore how taking a test on some of the material from a multimedia lesson, versus being presented with the same information as a list of facts to study, would affect restudy, particularly of untested information, and also learner confidence. To

our knowledge, our study is the first to use a partial-testing procedure with multimedia materials to examine test-potentiated restudy of the same material. We chose to use a multimedia lesson because multimedia materials are highly integrated (Mayer, 2005; Schnotz & Bannert, 2003) and commonly used in educational settings. We predicted that while experiencing a test between first and second study opportunities of a lesson would lead to lower confidence than reading facts during that same interval, in support of prior research indicating the relationship between fluency and metacognitive judgments (e.g., Benjamin, Bjork, & Schwartz, 1998; King, Zechmeister, & Shaughnessy, 1980), more learning would actually occur during a second study for those participants experiencing a test (Experiments 1a and 1b). We also predicted that the benefits of taking a test would transfer to subsequent study of new, unrelated material (Experiment 2).

Experiment 1a

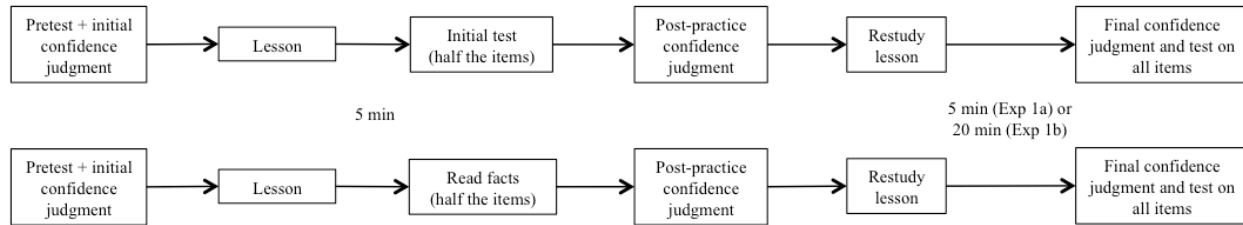
In Experiment 1a, we examined how studying a subset of facts from an educational multimedia lesson, versus taking a quiz on those same facts, would affect learning during restudy of that lesson. Participants watched and listened to a multimedia lesson and then either read sentences (read condition) or answered questions (test condition) about half the facts from that lesson. They then watched and listened to the multimedia presentation again before taking a final cued-recall test on all the facts from the lesson. A diagram illustrating this sequence of events is shown in the top half of Figure 4.1.

Method

Participants and design. A total of 45 undergraduates from the University of California, Los Angeles participated in exchange for course credit. Four participants were excluded from analyses due to high prior knowledge of the lesson content as determined by the

pretest, leaving 41 participants (30 female; average age, 20.5 years) in the study. Of these remaining participants, 22 served in the test-condition and 19 in the read-condition.

Experiments 1a and 1b



Experiment 2

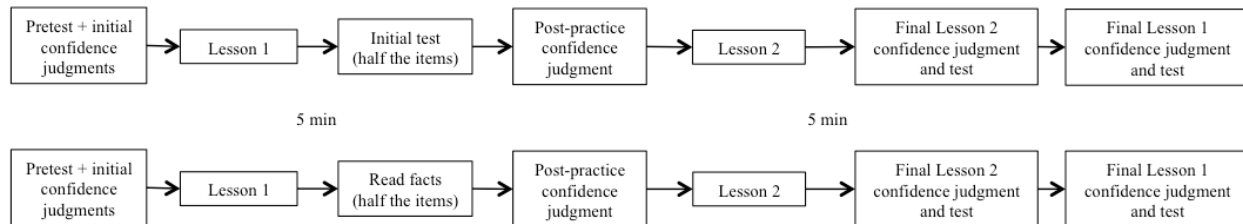


Figure 4.1. Schematic of the basic experimental procedures for Experiments 1a and 1b (top half) and Experiment 2 (bottom half).

Materials and procedure. As indicated in the top half of Figure 4.1, after consenting to participate, participants first took a pretest to assess their level of prior knowledge of astronomy. This pretest was based on similar instruments used in previous studies (e.g., Mayer & Johnson, 2008) and contained one question asking participants to rate their overall level of knowledge of astronomy on a scale from 1-7, with 1 being “very low” and 7 being “very high.” The rest of the pretest consisted of six statements to which participants responded either “yes” or “no” to indicate, respectively, whether that statement did or did not describe them: (1) I regularly keep up with NASA or other news about space travel, (2) I know the difference between the main sequence and red giant phases of a star, (3) I know what a neutron star is, (4) I can name all the

elements a star uses for fuel, (5) I can explain why a star goes supernova, (6) I can explain what causes high luminosity of a star. The total number of statements to which a participant replied “yes” was then added to the participant’s self-rating number to arrive at an overall pretest score for each participant. Scores of 8 or higher were considered to reflect high prior knowledge and data from the corresponding participants were thus excluded from analyses. On the pretest, we also asked participants to rate their confidence in their ability to explain the life cycle of a star on a scale from 1-7, with 1 indicating “very low confidence” and 7 indicating “very high confidence.” Answers to this question served as our initial measure of learner confidence.

After the pretest, participants listened to a brief clip of music (Mozart’s Serenade No. 13 in G) to ensure that the computer volume was adjusted to an appropriate level. They then watched and listened to the multimedia lesson, a 254-s long, animated and narrated video describing the life cycle of a star. In total, the lesson contained 24 idea units (listed in Appendix A). After the lesson ended, participants played Tetris during a 5-min distractor interval before beginning the practice phase. During the practice phase, those participants randomly assigned to the read condition were told that they were going to study some of the information from the lesson. Twelve facts then appeared on the computer screen, one at a time, for 10 s each (e.g., “The first stage in the life cycle of a star is the protostar stage.”). Participants randomly assigned to the test condition were told that they were going to answer questions about some of the information from the lesson. Subsequently, twelve questions appeared on the screen, one at a time, for 10 s each (e.g., “What is the first stage in the life cycle of a star?”). Participants had the full time to read and answer each question, which they did by typing their responses using the computer keyboard, and no feedback was provided during the test.

The facts or questions appearing on the screen during the practice phase were determined as follows. First, the 24 idea units in the lesson were consecutively numbered according to their appearance in the lesson, and then a given participant received either the odd- or even-numbered ones (counterbalanced across subjects) in the same order in which they had appeared in the lesson. This selection procedure was used in order to present an equal amount of information from throughout the presentation. Also, some of the later questions gave away answers to earlier ones, so we wanted to ensure that any correct responses reflected information participants had learned during presentation of the lesson and not information gleaned from prior questions.

After the practice phase, participants again rated their confidence in their ability to explain the life cycle of a star, and these responses were used as our measure of the learners' post-practice confidence. Participants then watched and listened to the multimedia lesson again. After this second presentation of the lesson was completed, participants again played Tetris during another 5-min distractor interval. After this distractor phase, participants again rated their confidence in their ability to explain the life cycle of a star, and these responses were used as our final confidence measure. Participants were then given the final test, which consisted of 24 cued-recall questions (one for each fact in the lesson). The questions on the final test were made up of the questions from both counterbalanced sets used for the initial test. After completing the test, participants answered demographic questions, following which they were debriefed and thanked for their participation.

Results and Discussion

Initial cued recall. Participants in the test condition answered, on average, just fewer than 4 out of 12 questions correctly on the initial test ($M = .30$, $SE = .04$). On both the initial and final tests, answers were scored leniently, meaning that a response was counted as correct if it

was clearly recognizable as the correct answer, but merely misspelled or incomplete (e.g., if “nebula” was the correct answer, “nebul” or “nubula” were scored as correct).

Final cued recall. Figure 4.2 shows correct recall proportions on the final cued recall test as a function of question type (i.e., information asked about in a final test question that had either been re-exposed or not as a studied or tested fact in the intervening practice phase between the first and second presentations of the multimedia lesson).

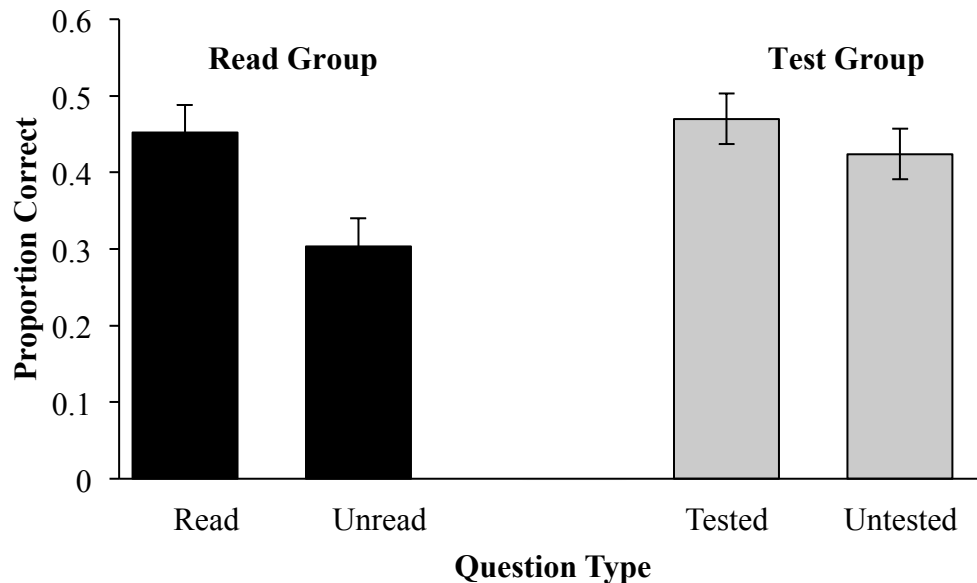


Figure 4.2. Average proportions of correct responses on the final cued recall test as a function of question type in Experiment 1a.

We conducted a 2 (exposure: seen in intervening practice phase vs. not seen in intervening practice phase) x 2 (condition: read vs. test) mixed-subjects ANOVA, with exposure as the within-subjects factor and condition as the between subjects factor, which revealed a marginal effect of condition, $F(1,39) = 2.88$, $MSE = .10$, $p = .10$, a significant main effect of exposure, $F(1,39) = 12.00$, $MSE = .19$, $p = .001$, and a marginal interaction between exposure

and condition, $F(1,39) = 3.41$, $MSE = .06$, $p = .07$. Additionally, a series of planned comparisons with Fisher's LSD adjustment for multiple comparisons revealed that the untested material ($M = .42$, $SE = .03$) was recalled just as well as the tested material ($M = .47$, $SE = .03$) in the test condition, $p = .24$, whereas unread material ($M = .30$, $SE = .04$) was recalled significantly less well than read material in the read condition ($M = .45$, $SE = .04$), $p = .001$, $d = .87$. Moreover, the untested material was recalled better than the unread material, $p = .02$, $d = .75$, whereas recall for the tested and read material did not differ ($p = .72$). Interestingly, the read group's score on the unread material was not different than the test group's score on the initial cued recall test, $p = 1.00$, suggesting that, for the read group, the second presentation of the lesson did not enhance learning of the unread material beyond what would have been its level of learning from a single presentation of the lesson.

Confidence. Figure 4.3 illustrates participants' confidence ratings in each condition at three points throughout the study: on the pretest (initial), right before the second presentation of the lesson (post-practice), and right before the final test (final). To examine the differences in confidence ratings between groups and how those ratings changed over the course of the experiment, we performed a 2 (condition: read vs. test) x 3 (time of judgment: initial vs. post-practice vs. final) mixed factors ANOVA, with condition as the between subjects factor. The ANOVA revealed reliable main effects of condition, $F(1,39)$, $MSE = 27.17$, $p = .02$, and time of judgment, $F(2,78) = 40.82$, $MSE = 29.56$, $p < .001$, and, importantly, a significant interaction between condition and time of judgment, $F(2,78) = 7.50$, $MSE = 5.43$, $p = .001$.

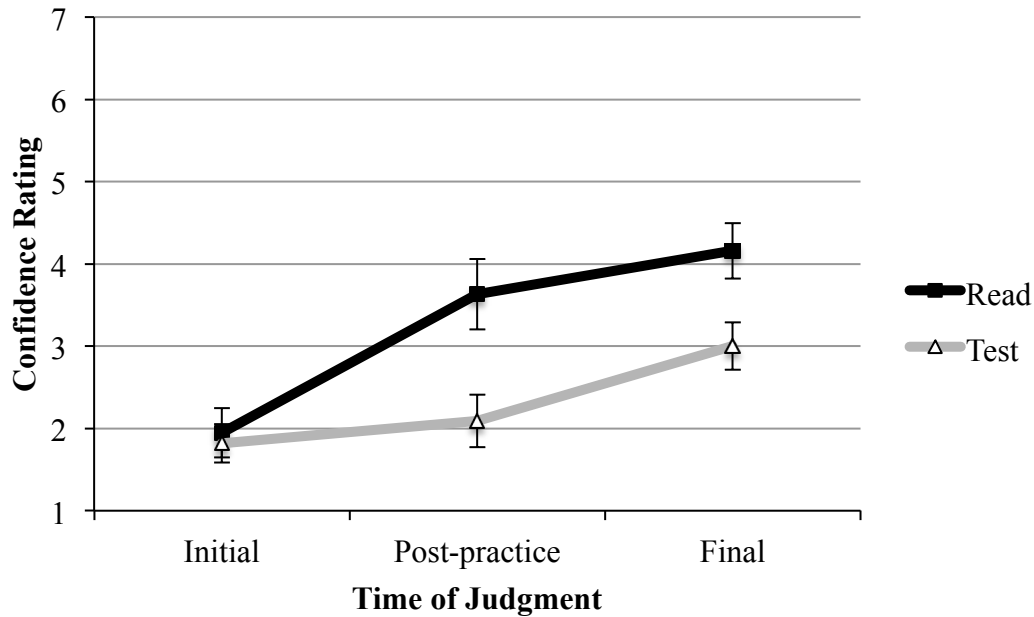


Figure 4.3. Average confidence ratings in each condition as a function of time of judgment in Experiment 1a. A confidence rating of 1 = “very low confidence”; a rating of 7 = “very high confidence.”

Follow-up comparisons indicated that, as expected, initial confidence ratings of the read ($M = 1.95, SE = .30$) and test ($M = 1.82, SE = 2.09$) groups were not statistically different, $p = .73$. Comparisons of post-practice and final confidence ratings, however, indicated that the read group was significantly more confident than the test group immediately before the second presentation of the lesson (read $M = 3.63, SE = .43$; test $M = 2.09, SE = .32, p = .006, d = .91$) and immediately before the final test (read $M = 4.16, SE = .34$; test $M = 3.00, SE = .29, p = .01, d = .83$). Interestingly, although the read group’s confidence increased significantly between the initial and the post-practice judgments ($p < .001, d = 1.05$) and between the post-practice and final judgments ($p = .03, d = .32$), the test group’s confidence increased only between the post-practice and final judgments ($p < .001, d = .64$). In other words, after seeing the lesson and taking a cued-recall test on half the items, participants were no more confident about their ability to explain the life cycle of a star than before they had been given the lesson ($p = 1.00$). If

instead, however, participants had read those facts instead of trying to answer questions about them, then they were more confident in their knowledge of the lesson content. Furthermore, this higher confidence for the read participants than the test participants remained after restudy of the lesson even though, as shown by their final-test performance, they had learned marginally fewer of the total facts about the life cycle of a star than had the test participants, thus demonstrating a pattern consistent with what is known in the literature as an illusion of competence (e.g., Koriat & Bjork, 2005).

To recap, Experiment 1a revealed several interesting findings. First, we found that taking a test on half the material from a multimedia lesson led to enhanced learning of the untested material during a second presentation of the lesson: The untested material was recalled better than the unread material and just as well as the tested material on a later final test of all the information. In contrast, however, reading that same material stated as facts did not lead to enhanced learning of the unread facts during a second presentation of the lesson, as reflected in their recall being significantly poorer than that for the read facts on the later final test.

We also found that participants who read half the facts from a lesson exhibited higher confidence than participants who took a quiz on those facts. This finding is consistent with previous work showing that learners tend to confuse the fluency, or ease, with which information is processed with long-term learning (Benjamin et al., 1998). Interestingly, even after the opportunity to restudy the lesson during which the test-condition participants learned more new information than the read-condition participants, the test participants remained less confident in their ability to explain the life cycle of a star than the read participants—although based on their final test performance, the tested participants would likely have been able to give a better explanation than their read counterparts.

Experiment 1b

In Experiment 1a, we showed that testing potentiated restudy of a multimedia lesson, with such benefits extending to the learning of the untested material, even though participants taking a test between the first and second presentations of the lesson were less confident than those who read facts during that intervening period. Experiment 1b was conducted as a close replication of Experiment 1a to see if this initial pattern of findings would occur again even after a longer delay between the restudy phase and final test.

Method

Participants. A total of 44 undergraduates participated in the study in exchange for course credit. Two participants were excluded because they exhibited high prior knowledge on the pretest, leaving 42 participants (26 female; average age: 20.6 years) of whom 20 had been randomly assigned to the read-condition and 22 to the test-condition.

Design, materials, and procedure. Experiment 1b was an exact replication of Experiment 1a, with the exception that participants engaged in a 20-min unrelated task rather than 5-min unrelated task, between the second presentation of the lesson and taking the final test.

Results and Discussion

Initial cued recall. The initial and final tests were scored in the same manner as in Experiment 1a. Similar to Experiment 1a, participants in Experiment 1b correctly recalled approximately 4 out of the 12 idea units on the initial test ($M = .34$, $SE = .03$).

Final cued recall. Figure 4.4 shows correct recall proportions on the final cued recall test as a function of question type obtained in Experiment 1b. A 2 (condition: read vs. test) \times 2 (exposure: seen in practice phase vs. not seen in practice phase) mixed factor ANOVA revealed a main effect of exposure, $F(1,40) = 4.90$, $MSE = .07$, $p = .03$, no effect of condition, $F(1,40) =$

2.30, $MSE = .13$, $p = .14$, and, critically, a significant interaction, $F(1,40) = 7.73$, $MSE = .115$, $p = .008$. To examine this interaction in greater detail, we performed the same planned comparisons as conducted in Experiment 1a, using Fisher's LSD correction for multiple comparisons. As in Experiment 1a, no difference was observed in the final recall performance for the previously untested material ($M = .49$, $SE = .04$) versus the previously tested material ($M = .47$, $SE = .04$), $p = .68$, but questions on previously read information ($M = .47$, $SE = .04$) were answered correctly more often than were questions on unread information ($M = .33$, $SE = .04$), $p = .001$, $d = .72$. Also consistent with Experiment 1a, recall performance for previously untested material by participants in the test condition was significantly higher than the recall performance for previously unread material by participants in the read condition, $p = .02$, $d = .79$. Together, these comparisons suggest that taking a test over a subset of the information presented during a multimedia lesson leads participants to restudy the material more effectively during a second presentation of the lesson than does the opportunity to read that same subset of information before the second presentation.

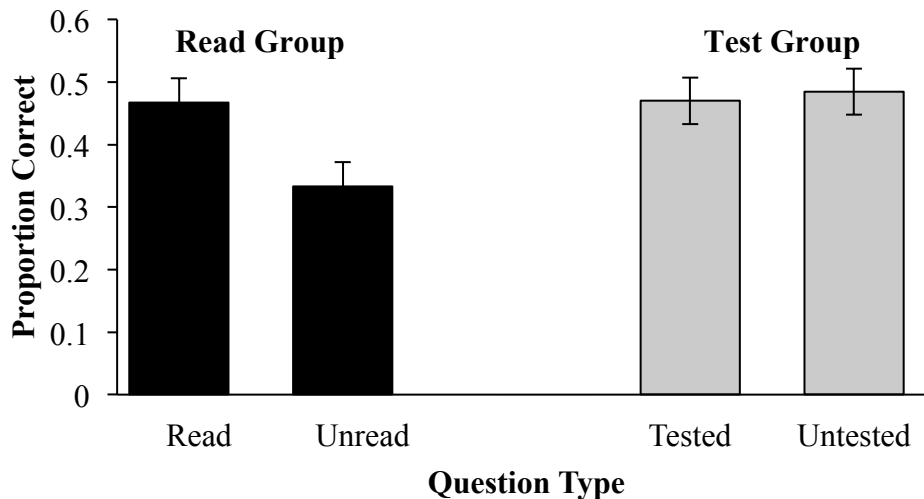


Figure 4.4. Average proportions of correct responses on the final cued recall test as a function of question type in Experiment 1b.

Confidence. Participants' initial, post-practice, and final confidence ratings are displayed in Figure 4.5. Just as in Experiment 1a, we performed a 2 (condition: read vs. test) x 3 (time of judgment: initial vs. post-practice vs. final) mixed factors ANOVA, with condition as the between subjects factor, which revealed main effects of condition, $F(1,40) = 10.48$, $MSE = 24.36$, $p = .002$, and time of judgment, $F(2,80) = 82.56$, $MSE = 62.23$, $p < .001$, and, again, a significant interaction between condition and time of judgment, $F(2,80) = 7.83$, $MSE = 5.90$, $p = .001$.

Follow-up comparisons revealed that, as in Experiment 1a, initial confidence ratings in the read group ($M = 1.70$, $SE = .25$) were equal to those in the test group ($M = 1.68$, $SE = .15$), $p = .95$. After that point, however, differences between the two groups emerged. As before, the read group was more confident than the test group both immediately before restudying the lesson (read $M = 4.25$, $SE = .23$; test $M = 2.86$, $SE = .27$), $p < .001$, $d = 1.22$, and immediately before the final test (read $M = 4.6$, $SE = .27$; test $M = 3.36$, $SE = .29$), $p = .003$, $d = .97$.

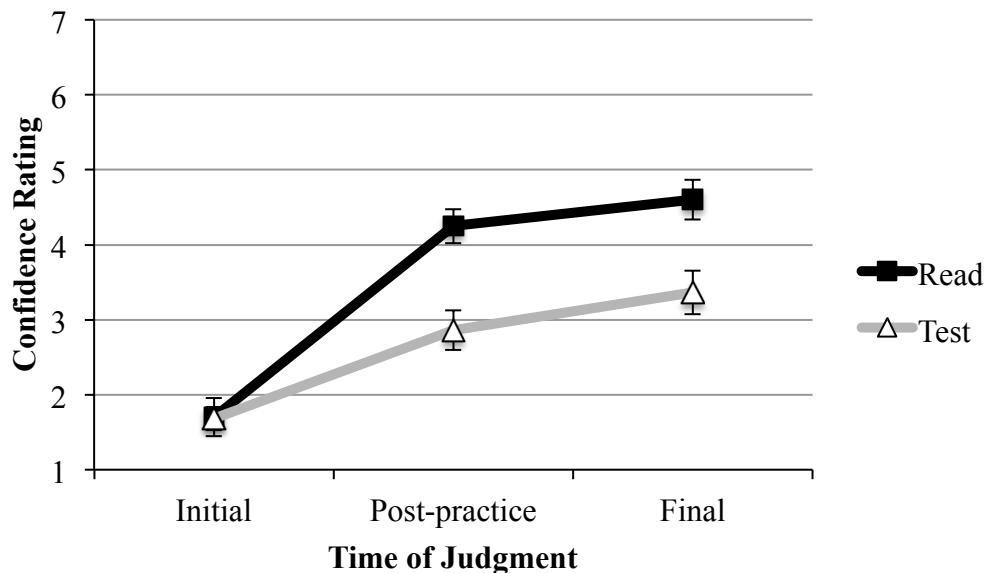


Figure 4.5. Average confidence ratings in each condition as a function of time of judgment in Experiment 1b. A confidence rating of 1 = “very low confidence”; a rating of 7 = “very high confidence.”

Similar to our findings in Experiment 1a, confidence ratings tended to increase at each time of judgment. In Experiment 1b, however, the test group increased significantly each time (initial to post-practice: $p < .001$, $d = 1.16$; post-practice to final, $p = .02$, $d = .38$, while the read group increased significantly only between initial and post-practice confidence judgments, $p < .001$, $d = 2.37$; thus, for the read participants, a second exposure to the entire lesson did not significantly increase their confidence above the level they had after a single presentation followed by reading half the facts from the lesson, $p = .12$.

In sum, Experiment 1b replicated the key findings of Experiment 1a. Taking a test on half the information from a multimedia lesson potentiated learning during restudy of that lesson. Furthermore, although reading facts led to a significantly greater increase in learner confidence, testing remained the more effective learning condition, especially for material not re-exposed to participants either in the form of facts or questions between lesson presentations.

Experiment 2

In Experiments 1a and 1b, we found that taking a test on a subset of information following presentation of a multimedia lesson, compared to reading that information as a set of facts, enhanced learning of the information not presented during that practice interval. It remains unclear, however, if the benefit to untested items occurred during restudy (owing to test-potentiated study) or during the initial test (owing to retrieval-induced facilitation). Thus, one goal of Experiment 2 was to try to determine which of these factors was responsible for the observed enhanced recall of the untested material on the final recall test in Experiments 1a and 1b. An additional goal was to extend previous work suggesting that testing can enhance subsequent study of new, unrelated material. Experiment 2 addressed both of these issues by replacing the restudy period of Experiments 1a and 1b (i.e., the second presentation of the same

lesson in those studies), with a new lesson. All participants were then tested on both lessons. We reasoned that such a design, a schematic illustration of which is presented in the bottom half of Figure 1, would allow us to see whether (a) test-potentiated study would generalize to the learning of new, unrelated material and whether (b) retrieval-induced facilitation could explain our previous results (specifically, the enhanced recall of untested items on the final cued-recall test).

Method

Participants. Sixty-three participants were recruited on Amazon Mechanical Turk, a website that allows users to complete tasks in exchange for money, and from the subject pool at the University of California, Los Angeles.¹ Studies using Mechanical Turk have replicated the results of many laboratory-based methods of data collection (see e.g., Mason & Suri, 2012). Online participants received \$2.00 for their participation, and UCLA students received partial course credit. Sixteen were eliminated due to high pretest scores, leaving 12 participants from the university's subject pool and 35 from Amazon Mechanical Turk (16 female; average age: 33.07 years). From this set, 23 randomly assigned participants served in the test condition and 24 in the read condition.

Materials, design, and procedure. The design of Experiment 2 required the use of two multimedia lessons. One of these was the same lesson about the life cycle of a star as used in Experiments 1a and 1b. The second was a similar, 170-s animated, narrated video about lightning formation. The images were modeled after the lesson used in Mayer, Heiser, and Lonn (2001), and the narration in the present lightning lesson was identical to the base narration in that

¹ We first ran subjects in the lab, but the term ended before we could complete data collection; therefore, we ran the remainder of the subjects online in order to complete data collection in a timely manner. We compared both subject pools' performance on multiple measures to ensure that the patterns were similar.

study. Questions were added to the pretest to assess knowledge of both meteorology and astronomy. As in Experiments 1a and 1b, initial confidence ratings for each lesson were also taken as part of the pretest. The lightning lesson was divided into 16 idea units, which are listed in Appendix C.

As illustrated in the bottom half of Figure 4.1, the procedure employed in Experiment 2 was very similar to that of Experiments 1a and 1b, except that after participants went through the practice phase and made their post-practice confidence ratings for that lesson, they were then presented with a multimedia lesson about lightning formation instead of being given a second presentation of the lesson about the life cycle of a star. Then, following a 5-min interval of playing Tetris, participants were asked to report how confident they were that they could explain how lightning formed, with this activity followed by a final cued-recall test on all 16 idea units from the lightning lesson. Next, they were asked to report how confident they were that they could explain the life cycle of a star (the subject matter of the first lesson they had experienced), and then, lastly, they were given the same final cued-recall test on that material as used in Experiments 1a and 1b. Following completion of this second final test, participants were debriefed either on the computer (Amazon Mechanical Turk participants) or in person (the lab participants), thanked, and excused.

Results and Discussion

Initial cued recall. The initial and final tests were scored in a similar manner as in Experiments 1a and 1b. The average score on the initial cued recall test was .27 ($SE = .03$), with no difference in performance between lab and online participants, $t(22) = .18, p = .86$.

Final cued recall: First lesson. One goal of Experiment 2 was to explore how participants would perform on the final cued recall test for a given lesson when not offered the

opportunity to experience a second presentation of that lesson following the practice period. Because we hypothesized that the critical improvement observed in Experiments 1a and 1b arose from processes occurring during the second presentation of the same lesson—not during the practice period—we did not anticipate seeing a benefit for untested information in the present paradigm as we did in Experiments 1a and 1b. We did, however, expect items that had been read or tested (i.e., exposed) during the practice period to be recalled better on the final test than items that were not presented in any form during the practice period.

The actual pattern of correct performance obtained on the final cued-recall test for each type of item from the first lesson is shown in Figure 4.6. As in Experiments 1a and 1b, we performed a 2 (exposure) x 2 (condition) mixed ANOVA to test our hypotheses. Unlike in those previous experiments, however, no indication of an interaction was revealed by the ANOVA, $F(1,45) = 0$, $MSE < .01$, $p = .99$. No effect of condition emerged, $F(1,45) = .01$, $MSE < .01$, $p = .93$, but there was an effect of exposure, $F(1,45) = 8.58$, $MSE = .15$, $p = .005$, with participants recalling exposed items ($M = .31$, $SE = .03$) more often than unexposed items ($M = .23$, $SE = .02$) on the final cued recall test. These findings thus support our hypothesis that without the restudy period, testing would provide no benefit to the untested material.

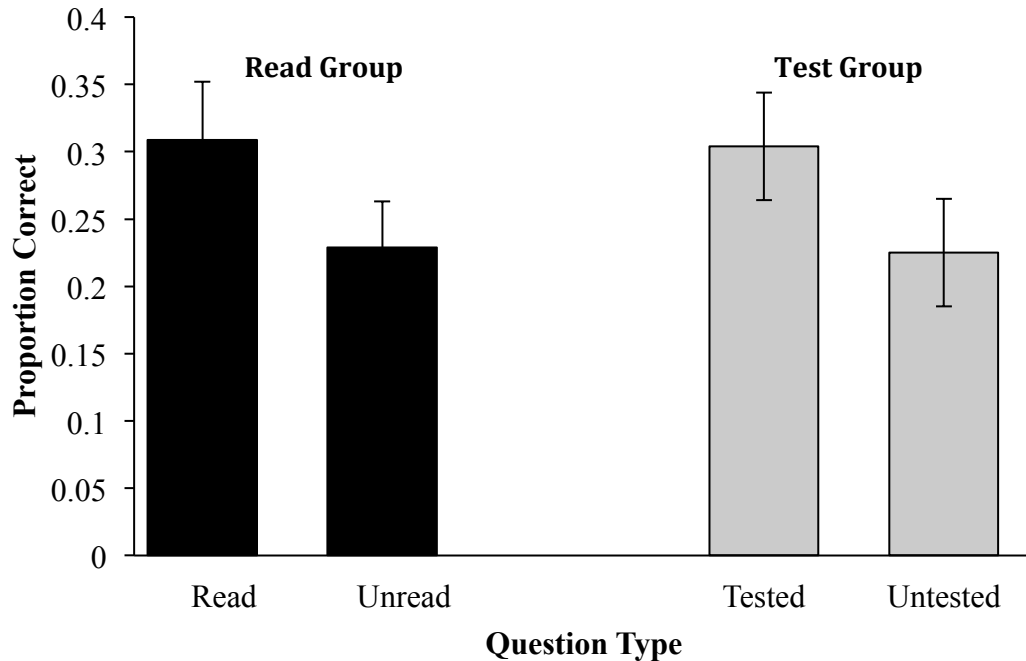


Figure 4.6. Average proportions correct on the final cued recall test for the first presented lesson as a function of question type in Experiment 2.

Final cued recall: New lesson. An additional goal of Experiment 2 was to explore how an initial test on previously studied information would affect learning of new information. We predicted that if the experience of a test, even if on different material, encourages learners to engage in more effective encoding strategies in general, rather than ones specific to the original lesson content, then test participants would show better overall performance than read participants on the test covering the second lesson. The results relevant to this prediction are shown in Figure 4.7; the test group did indeed perform significantly better ($M = .44$, $SE = .04$) on the cued-recall test for the new lesson than did the read group ($M = .33$, $SE = .04$), $t(45) = 2.07$, $p = .04$, $d = .61$. This finding suggests that the experience of a test leads learners to use more effective encoding strategies, or, possibly, to pay more attention during a subsequent lesson even when that lesson is on a topic different from the one covered by the initial test.

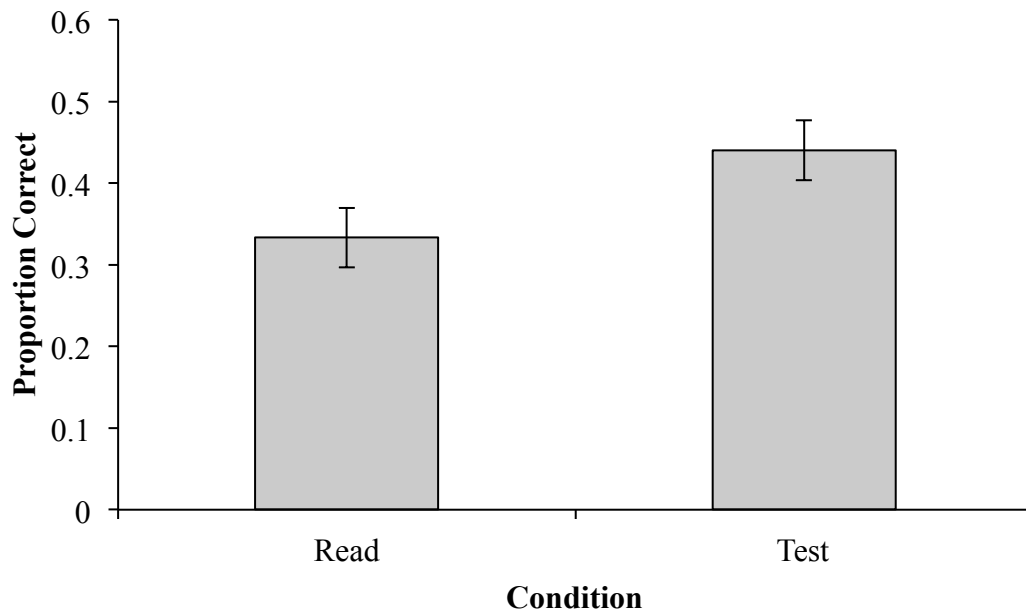


Figure 4.7. Average proportions correct on the final cued recall test for the second presented lesson for the read and test conditions in Experiment 2.

Confidence judgments. As in Experiments 1a and 1b, we performed a 3 (time of judgment) x 2 (condition) mixed ANOVA on participants' ratings of confidence in their ability to explain the life cycle of a star. As indicated in Figure 4.8, these judgments followed a similar pattern even when the restudy period for the first lesson was eliminated. As before, we observed a significant interaction between condition and time of judgment, $F(2,90) = 8.02$, $MSE = 5.39$, $p = .001$. As expected, participants in the read ($M = 1.50$, $SE = .17$) and test ($M = 1.52$, $SE = .17$) groups did not differ in their initial confidence ratings, $p = .80$. Additionally, as before, the read group was significantly more confident than the test group immediately following the practice phase (read: $M = 3.54$, $SE = .27$; test: $M = 2.30$, $SE = .28$), $p = .003$, $d = .92$, and immediately before the final test (read: $M = 3.04$, $SE = .29$; test: $M = 2.00$, $SE = .29$), $p = .02$, $d = .74$.

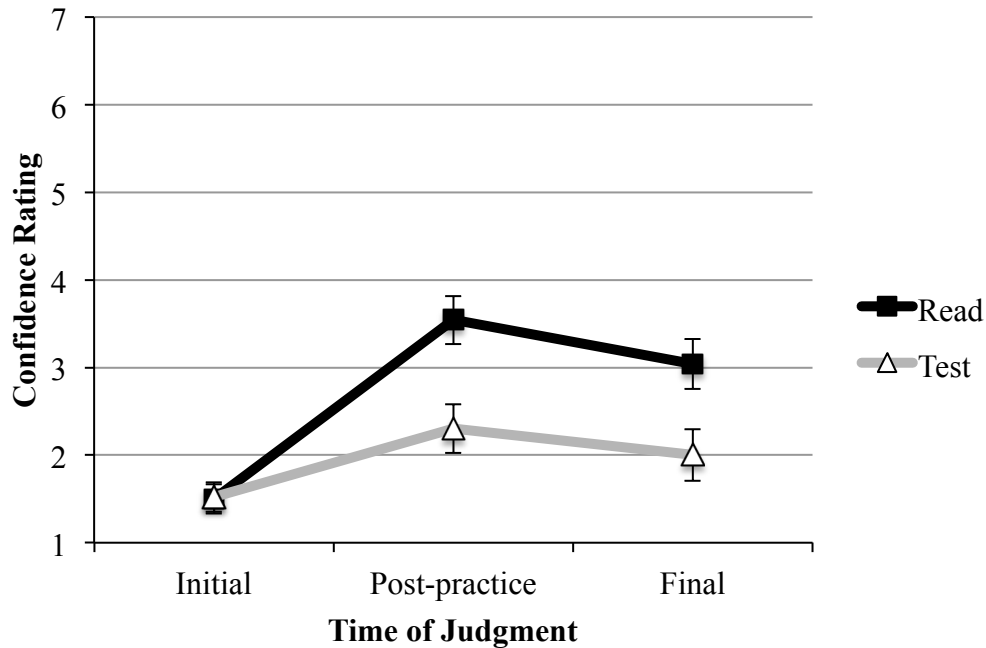


Figure 4.8. Average confidence ratings for first lesson material for the read and test conditions as a function of time of judgment in Experiment 2. A confidence rating of 1 = “very low confidence”; a rating of 7 = “very high confidence.”

Interestingly, even though the amount of time elapsing between post-practice and pre-final test was similar to that in Experiment 1b, we observed a somewhat different pattern in terms of how confidence judgments changed across the experiment. While judgments increased from the pretest measure to the post-practice measure for both groups (read $p < .001$, $d = 1.82$; test $p = .006$, $d = .71$), they actually decreased from post-practice to pre-final test in the read condition ($p = .004$, $d = .35$) and marginally so in the test condition ($p = .08$, $d = .23$). This drop in confidence between post-practice and pre-final test could have occurred because all participants experienced an intervening test on the second lesson. Even though that test covered different material, simply the act of taking a test could have shaken participants’ confidence in their ability to answer questions about another topic. This explanation would be consistent with prior work suggesting that experiencing a test can decrease overconfidence (e.g., Castel, 2008).

We asked participants to rate their confidence in their ability to explain lightning formation at only two points—on the pretest and right before they took the final test. Therefore, we performed a 2 (time of judgment) x 2 (condition) mixed ANOVA on confidence judgments about the lightning material, which revealed a significant interaction, $F(1,45) = 7.22$, $MSE = 6.53$, $p = .01$. As can be seen in Figure 4.9, similar to their confidence for the astronomy material, participants in the read ($M = 1.96$, $SE = .23$) and test ($M = 1.96$, $SE = .23$) groups reported similar levels of confidence in their ability to explain lightning formation on the pretest, $p = .99$. Despite undergoing the same experience with the lightning lesson, however, participants who had read facts about the star lesson ($M = 3.71$, $SE = .30$) were more confident than participants who had taken a test on the star lesson ($M = 2.65$, $SE = .30$), $p = .02$, $d = .73$. This finding is particularly interesting, given that read participants actually performed less well on the final test than did the test participants.

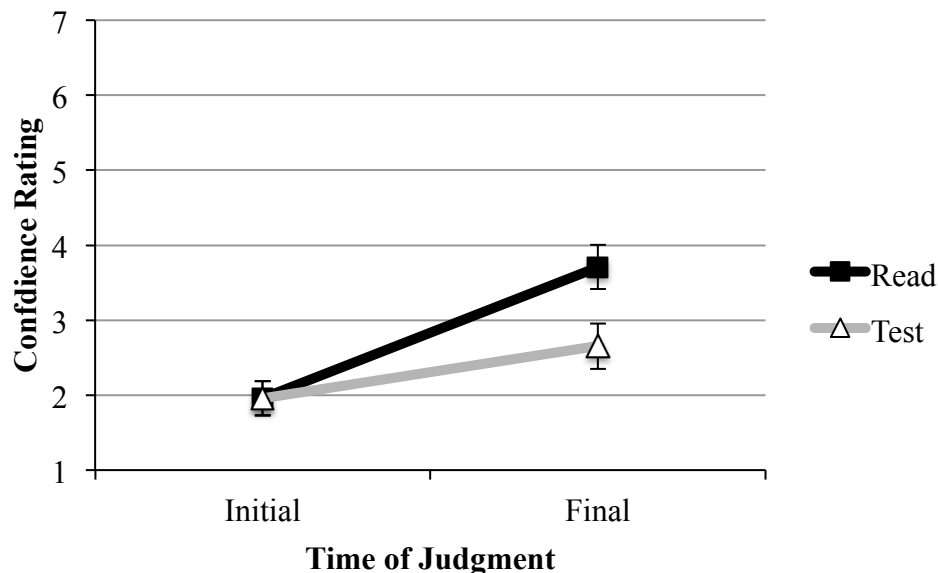


Figure 4.9. Average confidence ratings for second lesson material for the read and test conditions as a function of time of judgment in Experiment 2. A confidence rating of 1 = “very low confidence”; a rating of 7 = “very high confidence.”

General Discussion

The overarching goal of the present series of experiments was to explore whether testing a subset of information from a given multimedia lesson, as compared to presenting that same subset of information as facts to be read, could potentiate later restudy of that lesson, and possibly even, the study of subsequent lessons on different topics. In Experiments 1a and 1b, we found that taking a test potentiated restudy of the lesson such that material not tested during the practice phase in the test condition was recalled just as well as the material actually tested, and better than the material not presented as facts during the practice phase of the read condition. In other words, reading benefitted retention of only the presented items, whereas testing benefitted retention of all items in the lesson. Participants' confidence in the degree to which they had learned the material presented in the lesson, however, did not match their performance and, in particular, demonstrated that participants in the read condition were suffering from an illusion of competence. Namely, after the practice phase, participants who read facts were significantly more confident than participants who had taken a test in their ability to explain the content of the lesson even though their performance on the final cued-recall test revealed the opposite trend

This difference in confidence could well be one of the mechanisms leading to the observation of test potentiated learning in the present study. Previous research has shown that individuals often confused the fluency with which they can process currently available information with how well they have learned such information, making experienced fluency a misleading indicator of the probability of future retrieval (Benjamin et al., 1998; Bjork, Dunlosky, & Kornell, 2013; Soderstrom & Bjork, 2013). In the present research, reading facts would have been a very fluent activity, while taking a test on those facts would not have been (cf. Bjork et al., 2013). The taking of that test, however, presumably helped to inform learners of

what they did and did not know, and may have provided an impetus to use a more effective encoding strategy when given the opportunity to experience the lesson a second time (e.g., Pyc & Rawson, 2012; Soderstrom & Bjork, 2014; Wissman et al., 2011).

Soderstrom and Bjork (2014), for example, presented participants with 36 word pairs composed of forward associates, backward associates, and unrelated items. After an initial study period in which participants studied all the word pairs, students either restudied or took an interim test on those pairs. They then engaged in self-paced study of all the pairs before a final cued-recall test. Finally, participants reported what study strategies they used during the experiment by marking options on a checklist. Soderstrom and Bjork found that during self-regulated study, participants who took an interim test studied unrelated and backward pairs (i.e., the more difficult items) longer than forward pairs, while participants who had merely restudied did not spend more time on the more difficult items. In addition, tested participants marked a greater number of effective study strategies on the checklist, whereas restudy participants tended to mark more ineffective strategies.

Similarly, in the present studies, the experience of struggling to answer many of the questions on the initial test may have helped students realize they needed to shift to more effective encoding strategies when given the opportunity to restudy. This explanation is consistent with the mediator-shift hypothesis (Pyc & Rawson 2010, 2012), which states that failed retrieval attempts motivate learners to use better mediators (i.e., more effective encoding strategies) during subsequent study of the same material. It should be noted that we did not observe a retention benefit on the final cued recall test for facts that were tested over facts that were read during the practice period in Experiments 1a and 1b, but this finding is not entirely surprising. The benefits of testing over restudy are often found only at longer delays, especially

for educational material (e.g., Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008; Butler & Roediger, 2007; Roediger & Karpicke, 2006a, 2006b); therefore, our findings are consistent with prior literature on the direct benefits of testing.

Importantly, we also found that testing potentiated subsequent study of a new, unrelated multimedia lesson. These results thus add to the relatively sparse amount of previous work demonstrating such generalization effects of testing or that tests can potentiate study of new material (e.g., Szpunar et al., 2013; Wissman et al., 2011). Wissman et al. showed that taking free recall tests on initial expository text passages enhanced learning of the final critical passage. Similarly, Szpunar et al. demonstrated that interim tests during an online statistics lecture improved performance on the test over the final segment of that lecture. The present Experiment 2 adds several new findings in this domain: First, our results show that only one test is necessary to potentiate subsequent study, and second, test potentiated study of new material occurs with unrelated, multimedia lessons.

Limitations and Future Directions

Despite the integrated nature of our materials, we did not find retrieval-induced facilitation—that is, testing a subset of information in Experiment 2 (without subsequent restudy) did not enhance memory for the non-tested material. It could be that our test questions did not induce spreading activation, as was surmised to occur in Chan et al. (2006). In addition, the type of test could impact whether or not retrieval-induced facilitation occurs. Little et al. (2012), for example, found that a test comprised of multiple-choice questions with competitive alternatives resulted in retrieval-induced facilitation for untested, related information, whereas a cued recall test did not. Given that our initial test was a cued recall test, and each question tested a very

specific piece of information, participants may not have needed to bring to mind other information from the lesson to answer the question.

Also, if students are capable of changing to more effective encoding strategies upon restudy, the question remains as to why students do not engage in these better strategies from the start. Hypothetically, by the time students are in or beyond college (as were our participants), they have experienced multiple study-test trials in various domains. That students do not seem to spontaneously engage in more effective strategies, even though they are clearly capable of doing so, would suggest that they are overconfident in the efficacy of the less effective strategies. A reasonable assumption is that learners may employ a sort of minimum-effort heuristic, in which they try to do the easiest thing that they believe will work before engaging in more effortful, but more effective, strategies. Without the experience of a test, however, they have no way to know if their current strategy is sufficient (King et al., 1980; see also deWinstanley & Bjork, 2004). Another possibility, as suggested by Szpunar et al. (2013), is that taking the initial test in our experiments reduced mind-wandering during subsequent study, regardless of whether the post-test study period involved previously studied or new material.

A related issue for future research is how long test potentiated learning of new material lasts. As previously mentioned, learners have many experiences with study-test trials before coming into the lab, but something about the act of taking a test in this context results in test potentiation for new, unrelated material. A valuable avenue for future research would be to explore how recently a test needs to have been administered for that experience to induce test-potentiated learning of new material, or even how learners could be encouraged to engage in more effective processing strategies upon first study, rather than waiting for the test. This latter

question would be particularly relevant in educational settings, where students often report surprise at how poorly they performed on a test early in the term.

Concluding Comments

The present study is, we believe, the first to demonstrate that a test over only some of the material from a multimedia lesson can potentiate restudy of the entire lesson, as well as subsequent study of a new lesson. Our findings also suggest that tests can act to decrease learner overconfidence, mitigating illusions of competence, and perhaps provide learners with the motivation to change to a more effective study strategy during opportunities for post-test study. Findings of the present research should be of considerable interest and relevance to educators. Instructors, for example, may prefer to give short quizzes (i.e., quizzes that cover only a subset of the relevant material) as they take up less class time than more comprehensive quizzes. Our research indicates that these types of quizzes are sufficient to provide a benefit to both tested and untested information upon restudy.

CHAPTER 5: GENERAL DISCUSSION

Multimedia is a potentially useful tool for instruction. Online education programs and textbook publishing companies, eager to be at the forefront of technology, now tout their extensive multimedia resources, and learners who are increasingly tech-savvy welcome, and perhaps even expect, such a method of instruction. Instructors tempted by technology must be mindful, however, that multimedia is merely a tool, and if it is to be a useful one, its use must be guided by empirical research and grounded in theory.

To that end, the research reported in the present dissertation explored several ways to promote generative processing in multimedia lessons and, in Chapter 2, how to minimize extraneous processing of irrelevant information. One of the underlying assumptions of CTML and CATLM is that learners must actively process information in order to successfully learn. Such active processing includes selecting relevant images and words, organizing them into a coherent mental model, and integrating the current to-be-learned material with relevant prior knowledge. If any stage in that process is unsuccessful, then deep learning will not occur (see Mayer, 2005).

One of the first challenges that learners experience when encountering new material, then, is selecting key information. This task can be especially difficult for novice learners if a lesson includes interesting, but irrelevant, information (i.e., seductive details). Although instructors present seductive details with the intention of enhancing learning via increased interest in the topic, previous and present findings show that the presence of seductive details impairs learning of key material (e.g., Garner et al., 1989; Harp & Mayer, 1998). According to the schema diversion hypothesis, this so-called seductive details effect occurs because seductive details divert learners' processing resources away from key information (Harp & Mayer, 1998;

Lehman et al., 2007). Prior research indicates that this diversion may occur as early as the selection stage—for example, Lehman et al. (2007) found that on a self-paced reading task, learners spent less time reading sentences containing key information when seductive details were present than when they were absent. On the other hand, previous attempts to explicitly redirect learner attention (e.g., bolded sentences) did not reduce the seductive details effect. The experiment in the present Chapter 1 thus attempted to use a lower-level manipulation that would make learners more likely to engage in deeper processing for only the key information in an animated, narrated video. Seductive details in the lesson were accompanied with on-screen text identical to the narration, a condition that has been shown to increase extraneous processing and impair learning (Kalyuga et al., 1999; Mayer et al., 2001). In contrast, key information in the video was accompanied with on-screen text that was a reworded version of the narration, a condition that has been shown to promote learning, despite being judged as subjectively more difficult (Yue et al., 2013). We found that such a presentation resulted in recall equivalent to a condition without seductive details, and significantly better than a condition in which both seductive details and key facts were accompanied by reworded on-screen text. These findings indicate that introducing a slightly more challenging learning condition for only key information may force learners to engage in deeper processing for that information; consequently, they do not succumb to the seductive details effect.

The above results describe an example of a *desirable difficulty*, or a condition that makes learning initially feel more difficult, but results in greater long-term learning (Bjork, 1994). Another desirable difficulty shown to be useful in educational settings is retrieval practice (e.g., Butler & Roediger, 2007). When learners retrieve information, they alter and actually strengthen the representation of that information in their memory (Bjork, 1975; see also Roediger & Butler,

2011). Johnson and Mayer (2009) found that taking a test after viewing a multimedia lesson benefited later long-term performance on the same type of test. In Chapter 3 of the present work, we used either the visual or the auditory modality as a cue while participants retrieved the non-presented modality. We found no benefit for retrieval practice in this context over viewing and listening to the entire lesson twice, either on immediate or 1-week delayed tests.

One possible explanation for the present results lies in the integration component of multimedia learning. When learners integrate the visual and verbal channels of information from a multimedia lesson, it is possible that they become inextricably linked as a single representation in long-term memory (Mayer, 2005; Schnotz & Bannert, 2003). When one modality is presented again, therefore, it may not require any great retrieval effort to recall the other. According to the new theory of disuse, less effort at retrieval is generally associated with a smaller boost in retrieval and storage strength (Bjork & Bjork, 1992; Pyc & Rawson, 2009). In contrast, greater effort at retrieval usually means a greater boost in retrieval and storage strength. To illustrate this concept, consider a student studying flashcards. If the student studies the same flashcard twice in a row, he needs to exert very little effort to retrieve what is on the other side the second time it appears. If that flashcard resurfaces after a delay, however, then retrieval is more difficult at that time; later retrieval, however, is more likely in the latter case than in the former. The more difficult retrieval episode increased the strength and accessibility of the memory representation.

Following this logic, the results of the experiments in Chapter 3 suggest that presenting the entire visual or aural modality of the multimedia lesson may have actually made retrieval practice too easy, thus diminishing the effectiveness of that retrieval. An interesting direction for

future research would be to examine the degree of integration of the visual and verbal representations, and how such integration might affect retrieval processes.

Beyond exploring the direct effects of testing, we also investigated the possibility that testing would potentiate future study of multimedia lessons. As demonstrated in Experiments 1a and 1b of Chapter 4, taking a cued recall test on half of the information from a multimedia lesson potentiated restudy of that lesson for the untested information. Furthermore, Experiment 2 of Chapter 4 demonstrated that taking a test potentiated subsequent study of a new lesson. Similar to previous studies (e.g., Pyc & Rawson, 2010, 2012; Soderstrom & Bjork, 2014), our data on learner confidence support the theory that a test causes learners to become metacognitively aware that their prior encoding strategies were not sufficient. Specifically, learners who experienced a test were less confident than learners who read facts. This lower confidence may have encouraged learners to engage in more effective encoding strategies upon a subsequent learning opportunity, whether it was to restudy old information or learn new information.

Theoretical Implications

The present results suggest implications for CTML. First, we presented further evidence that on-screen text that is similar, but not identical, to the narration can help promote generative processing. In addition, the results from the study in Chapter 2 suggest that the concept of desirable difficulties can be applied to multimedia learning to guide learners' initial selection and processing of relevant information from a multimedia lesson. This initial study opens the door for future research into mixing levels of redundancy within a lesson, a relatively unexplored area of research. Furthermore, it suggests that other desirable difficulties may be relevant to multimedia instructional design.

The results from the experiments in Chapter 3 have implications for retrieval practice theory as well as CTML. We found similar results using both covert and overt retrieval practice methods, lending further support that overt and covert retrieval practice have similar benefits (Smith et al., 2013). Related to CTML, our results suggest that verbal and visual information becomes well integrated relatively quickly. As a result, using one modality as a retrieval cue may not require very effortful retrieval of the non-presented modality. Future work may explore how best to cue retrieval practice of such integrated memory representations.

Finally, the experiments in Chapter 4 expand on the existing literature on test potentiation by suggesting that testing some information from a multimedia lesson can foster generative processing during subsequent study. Our results contribute to developing theories on test-potentiation and offer support for the explanation that metacognitive awareness leads to the implementation of more effective study strategies. An additional contribution of the present work is that we demonstrated that even testing only half of the information from the lesson benefitted restudy of the untested information. An interesting future direction would be to explore how the composition of the test (e.g., where in the lesson the test questions come from) influences restudy and final test performance.

Practical Implications

A primary practical goal of this dissertation was to explore methods of optimizing multimedia learning. In general, our findings suggest that certain desirable difficulties may benefit learning when applied to multimedia lessons; presenting non-identical on-screen text, for example, appears to be a useful way to engender generative processing of key information, while identical on-screen text may inhibit deeper processing of less important information. Manipulating on-screen text in this manner may be a solution for instructors concerned about

student persistence, companies concerned about learner selections, or students concerned about the interest level of a lesson. Given that students persist longer when they find lessons to be more interesting (Ainley et al., 2002), our proposed method of minimizing the seductive details effect may be a simple way to maintain student interest without sacrificing learning.

Our findings regarding single-modality retrieval practice are not quite as clear. We failed to find a benefit for single-modality retrieval practice over studying a complete multimedia lesson twice; we also found, however, no detriment of studying only one modality after the initial, integrated learning phase. Although our findings suggest caution in assuming any direct benefits from such retrieval practice, they may offer some reassurance to students who have access to only one modality (e.g., lecture slides, audio podcast) during their study. Even if students believe that studying the entire lesson twice would be more beneficial, that is not necessarily the case. In addition, we found in the first experiment of Chapter 3 that reported cognitive activity during retrieval practice was positively correlated with scores on the final recall and transfer test; thus, if students are studying with only one modality, attempting to retrieve the other may be a helpful course of action.

Our final suggestion for educational application applies to instructors as well as students. Our Chapter 4 results indicated that testing even some information from a multimedia lesson benefits restudy for both tested and untested material. Therefore, instructors may wish to administer short quizzes during their classes. These quizzes need not be comprehensive; rather than testing only for the purpose of assessment, or even for the purpose of strengthening the tested information, a few questions before restudying or continuing to learn new information may benefit encoding. Similarly, students may wish to test themselves during study—again, not

only to evaluate their progress or make sure they know the tested information, but also to make their subsequent study more effective.

Limitations and Future Directions

The studies in the present dissertation used a limited set of materials—we used only two multimedia lessons—one about the life cycle of a star and one about lightning formation—and both were shorter than five minutes. Most educational settings involve much longer lessons and retention intervals. Therefore, future work may wish to extend upon the present research and test these ideas with additional materials and over longer delays. Additionally, there remains much work to be done regarding how seductive details, specifically, affect persistence in watching educational videos.

As mentioned above, our findings in Chapter 3 were inconclusive. Thus, while we cannot recommend that single-modality retrieval practice is a useful tool, we cannot completely rule it out based on the present results. Future studies could evaluate the potential direct effects of single-modality retrieval practice using fewer (or weaker) cues or longer delays between initial study and retrieval practice.

Finally, we do not suggest conclusive evidence for the mechanisms behind the test potentiation effect observed in Chapter 4. Based on the growing evidence for test potentiation with a variety of educationally relevant materials, this issue is a valuable one to pursue. Specifically, test composition and the length of the effect (e.g., for how long after a test will the effectiveness of subsequent study improve learning) would be interesting questions to explore in more depth.

Concluding Remarks

Multimedia lessons are appealing on a variety of levels. They can be more effective for learning than purely visual or purely verbal forms of education (e.g., Mayer & Anderson, 1991, 1992). Additionally, they offer an efficient way to transmit information; one online video can reach many more students than one in-person lecture. With students and instructors increasingly opting for multimedia, the importance of basing multimedia instructional design on empirically tested methods has never been greater. The present dissertation offers several new methods and findings that can inform the efforts of those involved in multimedia learning as instructors, designers, or both. First, selective use of identical and non-identical on-screen text can be used to invoke deep processing of key information but only shallow processing of less important information—minimizing, for example, the seductive details effect. Thus, designers of multimedia materials have a way to include interest-provoking details in lessons without harming learning of the important information. Second, while single modality retrieval practice was not found to result in better learning than restudying the complete multimedia lesson, it was also not found to impair learning. Thus, instructors who want to use such educational materials can be less concerned about doing so in circumstances where it might be difficult for students to access the intact multimedia lesson for future study (rather than, say, just the visual portion or narration portion). Third, the finding that taking a test over some of the information from a multimedia lesson, as opposed to restudying that information, potentiated restudy of old information as well as subsequent study of new information provides instructors with an easy to implement, but highly effective way, to improve their students' study strategies. In summary, the present research contributes several new findings that inform our understanding of multimedia learning and provide useful suggestions for the improvement of multimedia instructional design.

Appendix A

Complete narration broken into 24 idea units and seductive details (SD). Seductive details were only present in Experiment 1.

1. Stars are born out of nebulae, which are clouds in space made up of dust and gas.
SD: The most identifiable nebula is the Horsehead Nebula, which can be found in the constellation Orion.
2. Over time, gravity causes the dust and gas to accrete and clump together to form a protostar, the very first stage in the life cycle of a star.
3. As the protostar accretes more dust and gas atoms, the density at its core increases, leading to an increase in temperature and gas pressure.
SD: Hotter stars appear white or blue when viewed from earth, while cooler stars appear red or orange.
4. The star stops accreting molecules and enters the main sequence phase when it achieves equilibrium between gas pressure pushing outward and gravity pulling inward.
SD: Earth's sun is in this phase now.
5. The star spends the majority of its life in this phase, maintaining equilibrium.
6. Since gravity is constant and no more fuel is being pulled into the star, gas pressure must be maintained by fusing hydrogen atoms into helium atoms in the star's core.
7. This nuclear fusion causes heat and energy to radiate into space, and the core of the star begins to heat up.
8. Once all the hydrogen in the core has been converted to helium, the star has entered old age and is called a red giant.
9. It continues to burn fuel by performing nuclear fusion, but now it must fuse helium atoms into carbon.
10. The star is now burning fuel more rapidly and is less stable than it was in the main sequence phase.
11. As the temperature of the star increases, the outer shell of the star expands.
SD: Earth's sun will enter this stage in approximately 5 billion years.
12. After the red giant phase, the life of a star takes a different path depending on its size. The larger a star is, the faster it must burn its fuel to maintain equilibrium and the faster it progresses through the life cycle.
SD: One of the largest stars in the Milky Way galaxy is Eta Carinae, which is sometimes the second-brightest star in the night sky.
13. In low-mass stars, thermonuclear explosions occur in the outer shell every few thousand years.
14. Because of the instability these explosions produce, the outer shell of dust and gas particles expands and eventually dissipates, leaving behind only the hot core.
15. Nuclear fusion has left the core filled with mostly carbon, which the low-mass star cannot fuse into heavier elements.
16. Without a further source of energy to create gas pressure, gravity forces the star to contract,
17. and it remains a white dwarf for billions of years.
SD: White dwarfs are among the oldest objects in the known universe.

18. Medium-mass stars also become white dwarfs,
 19. but their life continues beyond that stage as neutron stars.
 20. Their greater core mass means that gravity is strong enough to pull in the outer layers it shed as a red giant.
 21. It continues to absorb matter until it achieves enough gas pressure to produce a powerful explosion known as a supernova.
- SD: Scientists estimate that supernovas occur about twice every century in the Milky Way Galaxy.*
22. The most massive stars can fuse carbon into other elements.
 23. Once they have burned through all their fuel, these stars also turn into neutron stars and produce a supernova.
 24. After the supernova, the most massive stars contain such a strong gravitational pull that they sometimes pull in even space itself. This process results in a black hole, an area in space where not even light can escape the gravitational pull.

Appendix B

Transfer questions and possible answers. Numbers in brackets indicate the idea units required to develop a correct response to that question.

1. How does a nebula become a protostar? [1,2]
 - a. As the particles become denser, their increased gravity attracts more and more particles, leading to a protostar
2. Imagine that a star is about to exit the main sequence phase. Describe how could it be kept in the main sequence phase (i.e., how can the equilibrium between gas pressure and gravity be maintained)? [4,5,6]
 - a. Replace helium atoms w/ 2 hydrogen atoms (on a larger scale)
 - b. Add hydrogen / fuel
 - c. Keep gravity at core stronger than internal pressure / keep internal pressure lower than gravity at core
3. Suppose there is a medium-mass star that had enough initial mass to go supernova, but it did not. What could have prevented the supernova from occurring? [20,21]
 - a. Decreased mass
 - b. Another gravitational force (or something indicating they understand that the gas cloud cannot be attracted back)
 - c. Add helium
 - d. Keeping temperature low enough
 - e. Ability to fuse heavier elements
4. Imagine that two clouds of dust and gas become protostars of identical mass at the same time. What could cause these stars to enter the red giant phase at different times? [1,2,4,6,7,8]
 - a. Same mass but different proportion of hydrogen
 - b. External gravity affecting star / different gravitational forces
 - c. Different rate of particle accumulation and/or nuclear fusion
 - d. Different temperatures
5. Why is it possible for medium- and high-mass stars to go supernova? [20,21]
 - a. Greater mass means greater gravity, so they can attract more gas particles back, leading to a high enough level of gas pressure to explode

Appendix C

Complete narration from the lesson on lightning formation broken into 16 idea units, adapted from Mayer, Heiser, & Lonm (2001). Phrases in brackets indicate main point of idea unit.

1. Cool moist air moves over a warmer surface and becomes heated. [Air becomes heated]
2. Warmed moist air near the earth's surface rises rapidly. [Air rises]
3. As the air in this updraft cools, water vapor condenses into water droplets and forms a cloud. [Air forms cloud]
4. The cloud's top extends above the freezing level, so the upper portion of the cloud is composed of tiny ice crystals. [Cloud's top freezes]
5. Eventually, the water droplets and ice crystals become too large to be suspended by the updrafts. [Ice becomes large]
6. As raindrops and ice crystals fall through the cloud, they drag some of the air in the cloud downward, producing downdrafts. [Ice falls]
7. When downdrafts strike the ground, they spread out in all directions, producing the gusts of cool wind people feel just before the start of the rain. [Downdrafts strike ground]
8. Within the cloud, the rising and falling air currents cause electrical charges to build. [Charges build]
9. The charge results from the collision of the cloud's rising water droplets against heavier, falling pieces of ice. [Rising droplets collide against falling ice]
10. The negatively charged particles fall to the bottom of the cloud, and most of the positively charged particles rise to the top. [Negatives fall to bottom]
11. A stepped leader of negative charges moves downward in a series of steps. It nears the ground. [Negatives move downward]
12. A positively charged leader travels up from such objects as trees and buildings. [Positives travel up]
13. The two leaders generally meet about 165-feet above the ground. [Leaders meet]
14. Negatively charged particles then rush from the cloud to the ground along the path created by the leaders. It is not very bright. [Negatives rush to ground]
15. As the leader stroke nears the ground, it induces an opposite charge, so positively charged particles from the ground rush upward along the same path. [Positives rush upward]
16. This upward motion of the current is the return stroke. It produces the bright light that people notice as a flash of lightning. [Produces flash]

Appendix D

Transfer questions and possible answers for lightning formation test, adapted from Mayer, Heiser, & Lonn (2001).

1. What could you do to decrease the intensity of lightning?
 - a. Increase negative charges on ground / decrease positive charges from the ground
 - b. Increase positive charges in cloud / decrease negative charges in the cloud
 - c. Prevent positive ions from going up into the cloud

2. Suppose you see clouds in the sky but no lightning. Why not?
 - a. No negative charges in cloud (no separation of negative and positive charges in cloud)
 - b. No positive charges on ground (no separation of negative charges in cloud and positive charges on ground)
 - c. Top of the cloud might not be above freezing level

3. What does air temperature have to do with lightning?
 - a. Cool air hits warm surface; surface heats air
*Must indicate that the surface heats the air
 - b. Top of cloud frozen, bottom not
*Must indicate that part of the cloud is frozen
 - c. Air must be cooler than the ground

4. What causes lightning?
 - a. Positive and negative charges in cloud
 - b. The difference in electrical charges within the cloud
 - c. Positive and negative charges between cloud and ground (need to say negative charges in cloud, positive charges on ground)
 - d. Top of cloud frozen, bottom not

References

- Adesope, O. O., & Nesbit, J. C. (2012). Verbal redundancy in multimedia learning environments: A meta-analysis. *Journal of Educational Psychology, 104*, 250-263. doi: 10.1037/a0026147
- Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology, 22*, 861-876. doi: doi: 10.1002/acp.1391
- Ainley, M., Hidi, S., & Berndorff, D. (2002). Interest, learning, and the psychological processes that mediate their relationship. *Journal of Educational Psychology, 94*, 545-561. doi: 10.1037//0022-0663.94.3.545.
- Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1063-1087. doi: 10.1037/0278-7393.220.5.1063.
- Arnold, K. M., & McDermott, K. B. (2013). Test-potentiated learning: Distinguishing between direct and indirect effects of tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 940–945. doi: 10.1037/a0029199.
- Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127*, 55–68. doi: 10.1037/0096-3445.127.1.55.
- Bjork, R. A. (1975). Retrieval as a memory modifier. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123-144). Hillsdale, NJ: Lawrence Erlbaum Associates.

- Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (Vol. 1, pp. 396-401). New York, NY: Wiley.
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe and A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp.185-205). Cambridge, MA: MIT Press.
- Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 35-67). Hillsdale, NJ: Erlbaum.
- Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417–444. doi: 10.1146/annurev-psych-113011-143823.
- Butler, A. C., & Roediger, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, *19*, 514-527. doi: 10.1080/09541440701326097.
- Callender, A. A., & McDaniel, M. A. (2009). The limited benefits of rereading educational texts. *Contemporary Educational Psychology*, *34*, 30-41. doi: 10.1016/j.cedpsych.2008.07.001.
- Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative processing explanation of the testing effect. *Memory & Cognition*, *34*, 268-276. doi: 10.3758/BF03193405.
- Castel, A. D. (2008). Metacognition and learning about primacy and recency effects in free recall: The utilization of intrinsic and extrinsic cues when making judgments of learning.

- Memory & Cognition*, 36, 429-437. Doi: 10.3758/MC.36.2.429.
- Chan, J. C. K. (2009). When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language*, 61, 153–170. doi: 10.1016/j.jml.2009.04.004.
- Chan, J. C. K., McDermott, K. B., & Roediger, H. L., III. (2006). Retrieval- induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, 135, 553–571. doi: 10.1037/0096- 3445.135.4.553.
- DeWinstanley, P. A., & Bjork, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory & Cognition*, 32, 945–955. doi: 10.3758/BF03196872.
- Dixon, F. A., & Glover, J. A. (1990). Another look at number signals and preview sentences. *Bulletin of the Psychonomic Society*, 28, 287-288.
- Finn, B., & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 238–244. doi: 10.1037/0278-7393.33.1.238.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new era of cognitive- developmental inquiry. *American Psychologist*, 34, 906-911. doi: 10.1037/0003-066X.34.10.906.
- Gaddy, M. L., van den Broek, P., & Sung, Y. (2001). The influence of text cues on the allocation of attention during reading. In T. Sanders, J. Schilperoord, & W. Spooren (Eds.), *Text representation: Linguistic and psycholinguistic aspects* (pp. 89-110). Amsterdam: John Benjamins Publishing Company.

- Gardiner, J. M., Craik, F. I. M., & Bleasdale, F. A. (1973). Retrieval difficulty and subsequent recall. *Memory & Cognition*, *1*, 213-216. doi: 10.3758/BF03198098
- Garner, R., Gillingham, M. G., & White, C. S. (1989). Effects of “seductive details” on macroprocessing and microprocessing in adults and children. *Cognition and Instruction*, *6*, 41-57. doi: 10.1207/s1532690xci0601_2.
- Harp, S. F. & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*, *90*(3), 414-434. doi: 10.1037/0022-0663.90.3.414.
- Hidi, S., & Baird, W. (1986). Interestingness—A neglected variable in discourse processing. *Cognitive Science*, *10*, 179-194.
- Hidi, S., Baird, W., & Hildyard, A. (1982). That’s important, but is it interesting? Two factors in text processing. In A. Flammer & W. Kintsch (Eds.), *Discourse processing* (pp. 63-75). Amsterdam: North-Holland Publishing Company.
- Izawa, C. (1966). Reinforcement-test sequences in paired-associate learning. *Psychological Reports*, *18*, 879–919. doi: 10.2466/pr0.1966.18.3879.
- Izawa, C. (1969). Comparison of reinforcement and test trials in paired- associate learning. *Journal of Experimental Psychology*, *81*, 600–603. doi: 10.1037/h0027905.
- Izawa, C. (1970). Optimal potentiation effects and forgetting-prevention effects of tests in paired-associate learning. *Journal of Experimental Psychology*, *83*, 340–344. doi: 10.1037/h0028541.
- Izawa, C. (1971). The test trial potentiating model. *Journal of Mathematical Psychology*, *8*, 200–224. doi: 10.1016/0022- 2496(71)90012-5.

- Johnson, C. I., & Mayer, R. E. (2009). A testing effect with multimedia learning. *Journal of Educational Psychology, 101*, 621-629. doi: 10.1037/a0015183.
- Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*, 351-371. doi: 10.1002/(SICI)1099-0720(199908)13:4.
- Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General, 138*, 469-486. doi: 10.1037/a0017341.
- Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331*, 772-775. doi: 10.1126/science.1199327.
- Karpicke, J. D., & Roediger, H. L., III. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 704-719. doi: 10.1037/0278-7393.33.4.704.
- King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *The American Journal of Psychology, 93*, 329-343. doi: 10.2307/1422236.
- Kintsch, W. (1980). Learning from text, levels of comprehension, or: Why anyone would read a story anyway. *Poetics, 9*, 87-98.
- Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 187-194. doi: 10.1037/0278-7393.31.2.187.

- Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review*, *14*, 219-224. doi: 10.3758/BF03194055.
- Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 989-998. doi: 10.1037/a0015729.
- LaPorte, R., & Voss, J. F. (1974). Paired-associate acquisition as a function of a number of initial nontest trials. *Journal of Experimental Psychology*, *103*, 117-123. doi: 10.1037/h0036836.
- Lehman, S., Schraw, G., McCrudden, M. T., & Hartley, K. (2007). Processing and recall of seductive details in scientific text. *Contemporary Educational Psychology*, *32*, 569-587. doi: 10.1016/j.cedpsych.2006.07.002.
- Little, J. L., Bjork, E. L., Bjork, R. A., & Angello, G. (2012). Multiple-choice tests exonerated, at least of some charges: Fostering test-induced learning and avoiding test-induced forgetting. *Psychological Science*, *23*, 1337-1344. doi: 10.1177/0956797612443370.
- Little, J. L., Storm, B. C., & Bjork, E. L. (2011). The costs and benefits of testing text materials. *Memory*, *19*, 346-359. doi: 10.1080/09658211.2011.569725.
- Loman, N. L., & Mayer, R. E. (1983). Signaling techniques that increase the understandability of expository prose. *Journal of Educational Psychology*, *75*, 402-412. doi: 10.1037/0022-0663.75.3.402.
- Lyle, K. B., & Crawford, N. A. (2011). Retrieving essential material at the end of lectures improves performance on statistics exams. *Teaching of Psychology*, *38*, 94-97. doi: 10.1177/0098628311401587.
- Magner, U. I. E., Schwonke, R., Aleven, V., Popescu, O., & Renkl, A. (2014). Triggering

- situational interest by decorative illustrations both fosters and hinders learning in computer-based learning environments. *Learning and Instruction, 29*, 141-152. doi: 10.1016/j.learninstruc.2012.07.002.
- Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods, 44*, 1-23. doi: 10.3758/s13428-011-0124-6.
- Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31-48). New York: Cambridge University Press.
- Mayer, R. E. (2009). *Multimedia Learning*. (2nd ed.) New York: Cambridge University Press.
- Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology, 83*, 484-490. doi: 10.1037/0022-0663.83.4.484.
- Mayer, R. E., & Anderson, R. B. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology, 84*, 444-452. doi: 10.1037/0022-0663.84.4.444.
- Mayer, R. E., & Johnson, C. I. (2008). Revising the redundancy principle in multimedia learning. *Journal of Educational Psychology, 100*, 380-386. doi: 10.1037/0022-0663.100.2.380.
- Mayer, R. E., Griffith, E., Jurkowitz, I. T. N., & Rothman, D. (2008). Increased interestingness of extraneous details in a multimedia science presentation leads to decreased learning. *Journal of Experimental Psychology: Applied, 14*, 329-339. doi: 10.1037/a0013835.
- Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*, 187-198. doi: 10.1037//0022-0663.93.1.187.

- Mayer, R. E., Mathias, A., & Wetzell, K. (2002). Fostering understanding of multimedia messages through pre-training: Evidence for a two-stage theory of mental model construction. *Journal of Experimental Psychology: Applied*, *8*, 147-154. doi: 10.1037//1076-898X.8.3.147.
- McCabe, J. (2011). Metacognitive awareness of learning strategies in undergraduates. *Memory & Cognition*, *39*, 462-476. doi: 10.3758/s13421-010-0035-2.
- McDaniel, M. A., Roediger, H. L., III, & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin and Review*, *14*, 200-206. doi: 10.3758/BF03194052.
- Moreno, R. (2006). Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. *Journal of Computer Assisted Learning*, *23*, 149-158. doi: 10.1111/j.1365-2729.2006.00170.x.
- Nelson, T. O. & Narens, L. (1994). Why investigate metacognition. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1-25). Cambridge, MA: MIT Press.
- Neuman, Y., & Schwarz, B. (2000). Substituting one mystery for another: The role of self-explanations in solving algebra word-problems. *Learning and Instruction*, *10*, 203-220. doi: 10.1016/S0959-4752(99)00027-4.
- Park, B., Moreno, R., Seufert, T., & Brünken, R. (2011). Does cognitive load moderate the seductive details effect? A multimedia study. *Computers in Human Behavior*, *27*, 5-10. doi: 10.1016/j.chb.2010.05.006.
- Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction*, *12*, 61-86. doi: 10.1016/S0959-4752(01)00016-0.

- 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*, 437-447. doi: 10.1016/j.jml.2009.01.004.
- Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science, 330*, 335. doi: 10.1126/science.1191465.
- Pyc, M. A., & Rawson, K. A. (2012). Why is test-restudy practice beneficial for memory? An evaluation of the mediator shift hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 737–746. doi: 10.1037/a0026166.
- Rawson, K. A., O’Neil, R., & Dunlosky, J. (2011). Accurate monitoring leads to effective control and greater learning of patient education materials. *Journal of Experimental Psychology, 17*, 288-302. doi: 10.1037/a0024749.
- Rey, G. D. (2011). Seductive details in multimedia messages. *Journal of Educational Multimedia and Hypermedia, 20*, 283-314.
- Rhodes, M. G. & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology, 137*, 615-625. doi: 10.1037/a0013684.
- Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*, 20-27. doi: 10.1016.j.tics.2010.09.003.
- Roediger, H. L., III, & Karpicke, J. D. (2006a). Test enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249–255. doi: 10.1111/j.1467-9280.2006.01693.x.
- Roediger, H. L., III, & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181–

210. doi: 10.1111/j.1745-6916.2006.00012.x.
- Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representations. *Learning and Instruction, 13*, 141-156. doi: 10.1016/S0959-4752(02)00017-8.
- Schwan, S., & Riempp, R. (2004). The cognitive benefits of interactive videos: Learning to tie nautical knots. *Learning and Instruction, 14*, 293-305. doi: 10.1016/j.learninstruc.2004.06.005.
- Smith, M. A., Roediger, H. L., III., & Karpicke, J. D. (2013). Covert retrieval practice benefits retention as much as overt retrieval practice. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 1712-1725. doi: 10.1037/a0033569.
- Soderstrom, N. C., & Bjork, R. A. (2013). Learning versus performance. In D. S. Dunn (Ed.), *Oxford bibliographies online: Psychology*. New York, NY: Oxford University Press.
- Soderstrom, N. C., & Bjork, R. A. (2014). Testing facilitates the regulation of subsequent study time. *Journal of Memory and Language, 73*, 99-115. doi: 10.1016/j.jml.2014.03.003.
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.
- Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences USA, 110*, 6313–6317.
- Szpunar, K. K., McDermott, K. B., & Roediger, H. L., III. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1392–1399. doi: 10.1037/a0013082.
- Wade, S. E., & Adams, R. B. (1990). Effects of importance and interest on biographical text. *Journal of Reading Behavior, 22*, 331-353. doi: 10.1080/10862969009547717.

- Wade, S. E., Schraw, G., Buxton, W. M., Hayes, M. T. (1993). Seduction of the strategic reader: Effects of interest on strategies and recall. *Reading Research Quarterly*, 28, 92-114. doi: 10.2307/747885.
- Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2011). The interim test effect: Testing prior material can facilitate the learning of new material. *Psychonomic Bulletin & Review*, 18, 1140–1147. doi: 10.3758/s13423-011-0140-7.
- Yue, C. L., Bjork, E. L., & Bjork, R. A. (2013). Reducing verbal redundancy in multimedia learning: An undesired desirable difficulty? *Journal of Educational Psychology*, 105, 266-277. doi: 10.1037/a0031971.p