USING TESTING TO POTENTIATE LEARNING FROM EXPOSITORY TEXTS

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

MARK ROSE LEWIS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

SASHANK VARMA, ADVISER

May, 2013

Acknowledgments

This dissertation would not have been possible without the support of my family and friends as well as the faculty and staff of the University of Minnesota. Special thanks go to Ava Yang for her support and inspiration, Mike Mensink for toughing things out with me, and Sashank Varma for his guidance.

This research was supported by an Institute of Education Sciences Interdisciplinary Education Sciences Training Program Grant (R305C050059), through funding received from the University of Minnesota's Office of the Vice President for Research and the College of Education and Human Development, and by a Doctoral Dissertation Fellowship from the University of Minnesota.

Abstract

This dissertation proposes a novel approach for improving learning using testing. A growing body of research has demonstrated that testing can be a powerful tool not just for assessing - but also for enhancing learning. However, previous research has only demonstrated that testing can increase *past* learning (i.e., learning from materials encountered prior to testing). By contrast, this dissertation proposes that testing can also be used to potentiate *future* learning (i.e., learning from new materials encountered for the first time after testing). According to the proposed model, testing potentiates future learning by producing *generative knowledge bases*. These generative knowledge bases are composed of featurally-rich memory traces that are fluently retrieved during future text processing, supporting comprehension, learning, and even conceptual change. This proposal is tested in three experiments using naturalistic expository texts. Experiment 1 replicated previous findings of test-enhanced learning and extended them to rich expository texts on topics in psychological science: being tested on a text increased the amount of information participants learned from that text relative to an equivalent amount of restudy, as indicated by delayed tests administered a week later. Experiment 2 demonstrated that testing potentiates future learning and conceptual change: being tested on a text increased the amount of information that participants learned from a new, but related, text that negated, qualified and elaborated the original text. Experiment 3 demonstrated that testing can potentiate future learning as measured by educationally relevant tasks: being tested on a recently read text improved students' critical essays and short answer questions about a new text encountered after testing. Experiment 3 also demonstrated that strongly held beliefs can interfere with the benefits of testing, preventing test-potentiated conceptual change. These results contribute to and extend the emerging view of testing as a valuable tool for promoting learning and represent a new tool for improving learning from expository texts.

Table of Contents

Acknowledgments i

Abstract ii

List of Tables v

List of Figures vi

Overview 1

Chapter 1: The Difficulty of Learning from Expository Texts 7

Prior Knowledge and Text Comprehension 8

Improving Prior Knowledge to Improve Comprehension 12

Misconceptions, Beliefs, and Learning 14

Chapter 2: Enhancing Past Learning with Testing 17

The Testing Effect 17

Proposed Mechanisms for the Testing Effect 20

Chapter 3: Using Testing to Potentiate Future Learning 24

Previous Studies of Test-Potentiation 24

Mechanisms for Test-Potentiated Learning 26

Potentiating Learning from Expository Texts 28

Chapter 4: Experiment 1 31

Method 31

Results 35

Discussion 38

Chapter 5: Experiment 2 40

Method 41

Results 45

Discussion 53

Chapter 6: Experiment 3 55

Method 57

Results 62

Discussion 71

Chapter 7: General Discussion 73

Mechanisms of Test Potentiation 74

Similarity and Transfer 76

Refutational Texts and Conceptual Change 78

Educational Implications 79

Chapter 8: Future Directions 82

Articulating the Mechanisms of Test-Potentiation 82

Establishing and Maximizing Educational Utility 86

References 90

Appendix A 107

Appendix B 111

List of Tables

Table 4.1. *Likert-Type Items Used to Measure Belief, Interest, and Prior Knowledge.* ... 34

Table 4.2. *Mean (SE) Belief, Interest, and Prior Knowledge Ratings for Tested and Restudied Topics.* ... 37

Table 4.3. *Mean (SE) Belief, Interest, and Prior Knowledge Ratings Disaggregated by Text.* ... 37

Table 5.1. *Example D1D8 Idea Units (Idea Units Present in Both Texts) and D8 Idea Units (Idea Units Present Only in the Day 8 Text).* ... 44

Table 5.2. *Mean (SE) Proportional Recall Disaggregated by Text* ... 48

Table 5.3. *Mean (SE) Belief, Interest, and Prior Knowledge Ratings for Tested and Restudied Topics by Experiment.* ... 49

Table 5.4. *Mean (SE) Belief, Interest, and Prior Knowledge Ratings Disaggregated by Topic for Experiments 1 and 2.* ... 51

Table 5.5. *Mean (SE) Belief, Interest, and Prior Knowledge Ratings Disaggregated by Text for Experiment 2.* ... 52

Table 6.1. *Mean (SE) Likert-Type Item Responses by Day 1 Learning Condition.* ... 66

List of Figures

Figure 3.1. *Test-potentiated learning from expository texts*. 29

Figure 4.1. *Proportion of idea units recalled on Day 8 as a function of Day 1 learning condition*. 36

Figure 5.1. *Proportion of idea units recalled on Day 8 as a function of Day 1 learning condition for the related text* 46

Figure 6.1. *Proportion of idea units from the Day 1 text (D1), both texts (D1D8), and the Day 8 text (D8) included in essay responses as a function of Day 1 learning condition*. 63

Figure 6.2. *Proportion of ideas units from the Day 1 text (D1), both texts (D1D8), and the Day 8 text (D8) included in essay and question responses as a function of Day 1 learning condition*. 65

Figure 6.3. *Interaction between beliefs and learning condition for the essay writing task*. 68

Figure 6.4. *Interaction between beliefs and learning condition for the short answer task*. 69

Figure 6.5. I*nteraction between beliefs and learning condition for the combined results (critical essay and short answer questions)* 70

Overview

Expository texts are central to learning at all stages of life. But unfortunately, many children and adults struggle to comprehend the types of expository prose that fill textbooks, scholarly journals, and popular press articles (see Chapter 1 for a review of key reasons that readers fail to comprehend and learn from expository texts). This dissertation presents a novel way of improving learning from expository tests. Building upon recent research that has shown that testing can enhance *past* learning (i.e., learning of information encountered prior to testing) (for a review, see Roediger & Karpicke, 2006a), we propose that testing can also be used to build generative knowledge bases that potentiate *future* learning (i.e. learning of new information encountered in future learning sessions). These generative knowledge bases are composed of featurally-rich memory traces that are fluently retrieved during future text processing, supporting comprehension, learning, and even conceptual change. This proposal is investigated in three experiments using ecologically valid expository texts.

Testing is often thought of as a way of assessing learning (e.g. classroom tests), ensuring political accountability (e.g., testing administered under policies like No Child Left Behind), and selecting individuals for limited spaces in educational institutions (e.g., tests like the SAT or ACT). However, a growing body of research from the cognitive and educational sciences demonstrates that testing can also be a tool for enhancing learning (e.g., Glover, 1989; Roediger & Karpicke, 2006b). When participants are tested on materials they have already been exposed to, they typically learn more than if they spend an equivalent amount of time restudying the materials. For example, if participants read an expository text and then take a recall test, they typically learn more than if they spend the same amount of time rereading the text. This enhanced learning is not usually apparent immediately, but is revealed by assessments administered days, weeks, and even months after testing (Butler & Roediger, 2007; Thompson, Wenger, & Bartling, 1978). The utility of testing in enhancing learning from previously encountered materials is discussed in Chapter 2.

The novel proposal presented in this dissertation is that – in addition to enhancing past learning – testing can potentiate future learning. According to this proposal, testing builds a *generative knowledge base* that enables participants to learn more from new, but related, materials encountered after testing. For example, if a participant reads an introductory text about a topic and is then tested on it, she will learn more from subsequent texts that build upon the ideas in the introductory text than if she had simply restudied the introductory passage. Testing potentiates learning by building generative knowledge bases that promote the fluent retrieval of prior knowledge via automatic, similarity-driven processes. This fluent retrieval increases the integration of tested knowledge with new textual information, improving learning and comprehension. In the context of learning from a refutational expository text, fluent retrieval increases the coactivation of conflicting information, promoting conceptual change. The proposed model of test-potentiated learning is introduced in Chapter 3.

This proposed model was investigated in three experiments that employed naturalistic expository texts. Experiment I investigated whether testing can be used to increase the knowledge learned from materials encountered prior to testing. In this experiment, participants read two naturalistic expository texts that presented arguments in favor of using discovery learning and learning styles to improve student learning. Participants were tested on one text and restudied the other for an equivalent amount of time. A week later, participants were tested on both texts to determine whether they had learned more from the tested or the restudied text. Consistent with previous research that has demonstrated that testing can enhance past learning, participants recalled a higher proportion of ideas from the text they were tested on than from the text they had restudied. Experiment 1 is presented in Chapter 4.

Experiment 2 investigated whether the increased knowledge built by testing transfers to new learning experiences, potentiating the learning of new information from new materials. In this experiment, participants again read the two texts that presented arguments in favor of using discovery learning and learning styles. After reading these texts, participants were tested on one and spent an equivalent amount of time restudying the other. A week later, participants read two new texts that presented arguments that refuted, qualified, and elaborated the claims presented in the original texts. After reading these texts, participants were tested to determine whether their learning had been potentiated by earlier testing on a related text. In contrast to prior experiments that have explored whether testing can potentiate the learning of previously encountered ideas to which a participant is re-exposed after tested, Experiment 2 explored whether testing can potentiate the learning of new ideas encountered for the first time after testing. Corroborating the proposed model of test-potentiated learning, having been previously tested on a related text a week before increased the amount of information that participants learned from a new text that refuted and elaborated the tested text. Critically, testing increased the subsequent learning of ideas that negated and qualified originally tested ideas, indicating that testing potentiated conceptual change. Experiment 2 is presented in Chapter 5.

Experiment 3 investigated whether testing can potentiate critical comprehension and learning in the context of two educationally relevant tasks. Although the results of the first two experiments demonstrated that testing can both enhance the knowledge that readers build and potentiate their future learning of new ideas, these learning gains were assessed using free-recall tasks. Free recall is a convenient tools for psychological research, but it is not directly relevant to educational practice. As a first step in demonstrating the potential educational utility of test-potentiated learning, Experiment 3 investigated whether prior testing could improve participants' performance on critical essays and short answer questions by enhancing their past learning and potentiating their future learning and critical comprehension.

Experiment 3 also investigated whether strong prior beliefs could prevent participants from benefiting from testing on ideas that contradicted these beliefs. Many people hold strong beliefs that everyone learns in different ways and that education is most effective when teaching is targeted to these "styles". Strongly held misconceptions like these can prevent readers from learning and/or using scientifically accurate ideas that conflict with these beliefs.

In Experiment 3, participants first read a text that introduced key criticisms of the validity and utility of learning styles approaches to education. After reading, one half of the participants were tested on the text and the other half restudied it for an equivalent amount of time. One week later, participants read a new text that presented an argument in favor of learning styles approaches to education. Participants were then asked to write an essay that critically evaluated this text and to answer a set of short answer questions. Consistent with previous research into test-enhanced learning, participants who were tested learned more ideas from the Day 1 text than participants who had restudied it. Corroborating the proposed model of test-potentiated learning, testing on Day 1 also enabled participants to learn more information from the new text they read on Day 8. This enhanced and potentiated learning, allowed participants to generate richer responses to the essay stem and short answer questions. However, participants' performance was also characterized by an interaction between their belief in learning styles and whether they had been tested on Day 1. While participants who were skeptical of learning styles approaches to education integrated more ideas from the critical Day 1 text into their responses when they had been tested on it, participants who strongly believed in the efficacy of learning styles failed to integrate more critical ideas from the Day 1 text regardless of whether they had been tested. Experiment 3 is presented in Chapter 6.

The results of these three experiments represent a significant expansion of our understanding of the role of testing in learning. They build upon the growing recognition of testing as more than a tool for assessing learning and demonstrate that testing can be used to both enhance past learning and potentiate future learning. Critically for potential educational applications, these results demonstrate that testing can amplify conceptual change and enhance/potentiate learning from the types of rich materials that students are likely to encounter in the real world. A full discussion of the results of these experiments is presented in Chapter 7, and suggestions for future research are presented in Chapter 8.

Chapter 1: The Difficulty of Learning from Expository Texts

Expository texts are central to learning at all stages of life. But unfortunately, many children and adults struggle to comprehend the types of expository prose that fill textbooks, scholarly journals, and popular press articles (RAND Reading Study Group, 2002). American students in particular demonstrate shortcomings in the comprehension of expository texts that have implications for both civic engagement and national competitiveness (RAND Reading Study Group, 2002). For example, the results of the Progress in International Reading Literacy Study 2006 (PIRLS 2006) indicated that American 4^{th} graders were significantly worse at comprehending text for “informational purposes” (i.e., comprehending expository texts) than they were at comprehending texts for “literary purposes” (i.e., comprehending narrative texts) (Mullis, Martin, Kennedy, Foy, 2007). The results of PIRLS 2006 also indicated that American students were significantly worse than students from 12 countries in the 40 country sample in understanding texts for informational purposes. Although the Program for International Student Assessment (PISA) does not report separate results for expository and narrative texts, PISA assessments place a heavy emphasis on students’ ability to comprehend and learn from expository texts. The results of PISA 2009 indicated that American 15 year olds’ reading ability was only “average” and was significantly worse than the performance of students in Asian countries and regions like Shanghai-China, Korea, Hong-Kong, and Singapore; European countries like Finland, the Netherlands, Belgium, and Norway; and Commonwealth countries like Canada, New Zealand, and Australia against whom we compete in the global marketplace (OECD, 2010).

Why do students have so much difficulty comprehending and learning from expository text? A key reason that readers struggle to comprehend expository texts is because they fail to recruit relevant prior knowledge (Chiesi, Spilich, & Voss, 1979; McNamara, E. Kintsch, Songer, Kintsch, 1996). When readers do not possess relevant prior knowledge, they may generate incomplete and/or erroneous representations of a text (Schneider & Korkel, 1989; Steffensen, Joag-Dev, & Anderson, 1979). Even when readers do have relevant prior knowledge, the constraints of the human cognitive system may prevent them from fluently integrating it with textual information, preventing them from generating accurate, fully elaborated text representations (Kinstch, 1988; McKoon & Ratcliff, 1992).

Prior Knowledge and Text Comprehension

Successful text comprehension requires that readers integrate relevant textual information and prior knowledge at appropriate times during reading (Kintsch, 1988; Kintsch & van Dijk, 1978; Thibadeau et al., 1982). This integration is constrained by limits in how much information a reader can maintain and process in working memory (WM) (Fletcher, 1981; Fletcher & Bloom, 1988; Just & Carpenter, 1992; McKoon & Ratcliff, 1992), by restrictions in how quickly and selectively readers can retrieve information from long term memory (LTM) (Anderson, Budui, & Reder, 2001; Cook, Halleran, & O'Brien, 1998), and by the nature of readers' prior knowledge (Bartlett, 1932; Chiesi et al., 1979). Readers have a limited amount of WM resources that they must allocate to both the processing and short-term storage of textual information and prior knowledge (Fletcher, 1981, 1986; Just & Carpenter, 1992; Kintsch & van Dijk, 1978). Although readers can elaborate the information that is temporarily held in WM with information retrieved from LTM, retrieval from LTM can be alternatively slow and selective or fast but imprecise (Anderson et al., 2001; Cook et al., 1998).

Readers must rely upon their prior knowledge to fill gaps in elliptical texts (Bower, Black & Turner, 1979), explain how textual events are causally related (Trabasso, Secco, & van den Broek, 1984), and construct mental models that transcend linguistically encoded relations (Bransford, Barclay, & Franks, 1972; Johnson-Laird, 1983; van Dijk & Kintsch, 1983). To understand the critical role that recruiting relevant background knowledge plays in text comprehension, consider the following example:

> With hocked gems financing him, our hero bravely defied all scornful laugher that tried to prevent his scheme. "Your eyes deceive," he had said, "an egg, not a table, correctly typifies this unexplored planet". Now three sturdy sisters sought proof, forging along, sometimes through calm vastness, yet more often over turbulent peaks and valleys. Days became weeks as many doubters spread fearful rumors about the edge. At last, from nowhere, welcome winged creatures appeared, signifying momentous success. (Dooling & Lachman, 1971)

If you have never seen this text before, it was probably difficult to comprehend. However, if the appropriate title had been included (see footnote on next page), you would have most likely recruited relevant background knowledge and had little difficulty understanding the vague references that pervade this passage.

Although this example is a bit extreme, it demonstrates a fundamental principle of text comprehension: successful comprehension requires that readers recruit relevant background knowledge during reading (Adams & Bruce, 1980; Bartlett, 1932; Bransford & Johnson, 1972; Walker & Yekovich, 1987). When readers do not have relevant prior knowledge, they are entirely dependent on the information encoded in the words of a text. As a result, they may generate inaccurate or incomplete representations of the underlying meaning of a text. When readers do have access to relevant prior knowledge, they learn more from a text (e.g., Chiesi, et al., 1979; Recht & Leslie, 1988), they can understand ambiguous language (Wiley & Rayner, 2000), resolve subtle references (Walker & Yekovich, 1987), and make sense of even minimally coherent texts (McNamara et al., 1996). When this prior knowledge is extensive, readers may even be able to circumvent limitations in their own general cognitive abilities (Schneider & Korkel, 1989; Schneider, Korkel, & Weinert, 1989; Walker, 1987).[1]

One way that prior knowledge facilitates text comprehension is by providing readers with information that can help them to elaborate or fill gaps in texts that are not fully coherent (McNamara, 2001; McNamara & Kintsch, 1996; McNamara et al., 1996). For example, McNamara et al. (1996) had students who varied in prior knowledge read biology texts that varied in their level of coherence, ranging from a minimally coherent text that was locally but not globally coherent to a fully coherent text that included signals of the underlying structure of the text as well as language that explicitly explained the relations between text segments. Students with high prior knowledge were able to employ their knowledge to construct a coherent representation of the minimally coherent text and actually learned more from this apparently impoverished text than from the fully

---

[1] *Christopher Columbus Discovering America*

coherent version. Students with low prior knowledge learned little from the incoherent text because they were not able to recruit relevant knowledge during reading.

Another way that prior knowledge facilitates text comprehension is by enabling readers to circumvent limitations in their general cognitive abilities (Schneider & Korkel, 1989; Schneider et al., 1989; Walker, 1987). For example, Schneider et al. (1989) had students who varied in general verbal ability and prior knowledge about soccer read texts that described soccer matches. Students with low verbal ability who were soccer experts actually learned more from the texts than did students with high verbal ability who were soccer novices. This result demonstrates that high domain knowledge can compensate for general cognitive deficits. Ericsson and Kintsch (1995) have proposed that expert knowledge can allow readers to dramatically expand the information that they have quick, precise access to by exploiting over-learned LTM structures. According to their theory of Long Term Working Memory (LT-WM), expertise allows readers to expand their WM in much the same way as chess masters use their expert knowledge to dramatically expand their WM for chess positions (Chase & Ericsson, 1982; Chase & Simon, 1973).

Deficits in prior knowledge may be especially culpable in the case of failed expository text comprehension. Although life experiences may prepare readers to readily understand narrative texts, the knowledge structures that are necessary for understanding expository texts are often unique to a discipline and must be learned (Britton, Graesser, Glynn, Hamilton, & Penland, 1983; Kintsch & Young, 1984)[2]. The large amount of relevant knowledge that readers bring to narrative texts, such as knowledge about typical human activity and extensive experience with narrative story structures, may allow them to easily construct representations of narrative texts that capture large thematic relations (Britton, et al., 1983; Kintsch & Young, 1984). By contrast, the relatively impoverished knowledge base that readers bring to most expository texts may drive them to focus on the processing of isolated information that is explicitly encoded in a text (McDaniel, Einstein, Dunay, & Cobb, 1986; McDaniel & Kerwin, 1987).

Improving Prior Knowledge to Improve Comprehension

Given the centrality of prior knowledge to successful text comprehension, it makes sense to try to improve readers' comprehension by helping them to build requisite prior knowledge or helping them to recruit the relevant knowledge they already possess. Several methods have been used to help readers build the prior knowledge necessary for comprehending a text. These include "advance organizers" (Ausubel, 1960) and short lessons that aim to introduce relevant content.

Attempts to provide readers with requisite background knowledge through short targeted lessons have yielded inconsistent results. For example, McNamara and Kintsch (1996) found that giving readers a short lesson on the Vietnam War prior to reading a difficult text on the conflict in Indochina only helped readers who already had high prior knowledge about the war. In fact, these short lesson actually hurt readers who had low prior knowledge. However, other studies have found that exposing readers to relevant

[2] While the boundaries between different text types are definitely fuzzy, cognitive and educational psychology have historically differentiated between two super-classes, expository and narrative texts. The purpose of expository texts being to inform, and the purpose of narrative texts to entertain.

knowledge can increase the amount they subsequently recall from a passage (Mayer, 1983; Rawson & Kintsch, 2002, 2004). For example, Rawson and Kintsch (2002, Experiment 1) had participants read a series of short texts about governmental interventions in the labor market. Before reading, half of the participants read information about the issues covered in the texts. After reading, all participants completed a free- and a cued- recall test. Participants who read background information prior to reading produced more textual information on the free recalls than did participants who had not.

Advance organizers are intended to introduce and/or activate conceptual structures that can help a reader subsume or anchor new textual material (Ausubel, 1960; Mayer, 1979). The potential efficacy of advance organizers was investigated extensively in the 1960s, 70s, and 80s. The results of this research can be best summarized as "confused" (Hartley & Davies, 1976, p. 256). While many reviewers of this research have concluded that advance organizers do little to improve learning from texts (e.g., Barnes & Clawson, 1975; Faw & Waller, 1976), others have concluded that advance organizers can improve learning if properly constructed (Corkill, 1992; Mayer, 1979).

The sporadic success of these two methods suggests that it can be difficult to help readers build the knowledge they need to understand a text. Looking forward, the proposed model of test-potentiated comprehension aims to overcome this difficulty by using testing to not only help readers build relevant knowledge but to increase the fluency with which this knowledge is retrieved during future reading, supporting comprehension and learning.

Misconceptions, Beliefs, and Learning

It is important to recognize that not all prior knowledge facilitates learning. In fact, readers often possess knowledge or beliefs that can actually interfere with learning (Alvermann, Smith, & Readence, 1985; Diakidoy & Kendeou, 2001; Kardash & Scholes, 1996; Kendeou & van den Broek, 2007). For example, a student who believes that seasonal change is the result of the earth being closer to the sun during the summer than during the winter can read a text that describes the actual mechanism of seasonal change (the Earth's tilt and elliptical orbit) without revising their misconception (Broughton, Sinatra, & Reynolds., 2010). In general, misconceptions are highly resistant to revision. In the case of strongly held beliefs (e.g., a belief in creationism), readers' strong commitment to a misconception can further reduce the likelihood that they will learn a new concept from a text (Dole & Sinatra, 1998; Kardash & Scholes, 1996).

Changing misconceptions and erroneous beliefs. According to models of conceptual change that have emerged from modern philosophy of science and science education, there are several pre-requisites for changing misconceptions. These prerequisites include dissatisfaction with an existing concept, the presence of an intelligible new concept, and a demonstration of the plausibility and fruitfulness of a new concept (Posner, Strike, Hewson, & Gertzog, 1982).[3] In addition to these philosophically-inspired prerequisites, educational, cognitive, and social psychology have identified several additional factors that are key to fostering conceptual changes (Dole & Sinatra, 1998). For example, conceptual change is most likely to occur when learners are highly

---

[3] In the context of scientific theory building, conceptual change is also driven by "the possibility of a fruitful research program" associated with the new concept (p. 214 Posner et al., 1982)

engaged, when a topic is personally relevant or interesting, and when learners have a high "need for cognition" (Cacioppo & Petty, 1982; Dole & Sinatra, 1998; Gregoire, 2003; Pintrich, Marx, & Boyle, 1993).

In the context of learning from texts, conceptual change is promoted when new concepts are presented in a refutational structure (for a review, see Guzzetti, Snyder, Glass, & Gamas, 1993). In a refutational text, a misconception is explicitly highlighted and then the correct concept is presented. For example, consider the following excerpt from a refutational text about seasonal change:

> Many people believe that the changing seasons are the result of the Earth being closer to the Sun during the summer months and farther away from the Sun during the winter months. Perhaps you hold similar beliefs. However, seasons do not change because the distance between the Earth and the Sun changes. In fact, Earth is closer to the Sun in winter and farther away from the Sun in summer. Seasonal change is the result of two features of the Earth: its tilted axis and its elliptical orbit around the Sun. (p. 421 Broughton et al., 2010)

Notice that this text explicitly highlights potential misconceptions in the opening sentence "Many people believe that the changing seasons are the result of the Earth being closer to the Sun during the summer months and farther away from the Sun during the winter months." It then continues to explain the actual mechanism for seasonal change.

When participants with misconceptions read a refutational text, they are more likely to revise their misconception than if they read a non-refutational text (Broughton et al., 2010; Diakidoy & Kendeou, 2001; Kardash & Scholes, 1996; Kendeou & van den

Broek, 2008; Guzzetti, et al., 1993). One proposed mechanisms for this change is the cognitive conflict caused by the co-activation of misconceptions and new concepts that is evoked by a refutational structure (van den Broek & Kendeou, 2008). Another mechanism for conceptual change may be that refutational texts are easier to process than their non-refutational counterparts (Broughton et al., 2010).

Looking ahead to the experiments, an important dimension of Experiments 2 and 3 is investigating whether testing can help participants learn ideas that contradict prior knowledge. These experiments feature texts that refute commonly held beliefs about learning, such as the belief that students learn best when taught in a matched “learning style”. Experiment 2 investigates whether testing can set the stage for students to learn more from refutational texts they encounter after testing. Experiment 3 investigates whether prior testing can improve participants’ responses to critical essays and short-answer questions that require integrating conflicting ideas. Experiment 3 also investigates whether participants’ prior beliefs interact with the potential benefits of testing by reducing their learning/use of ideas that conflict with their beliefs.

Chapter 2: Enhancing Past Learning with Testing

We usually think of tests as a way of assessing learning (e.g., classroom tests or accountability tests administered in accordance with policies like No Child Left Behind), diagnosing individuals, incentivizing study, or selecting individuals for limited spaces in educational institutions (e.g., tests like the ACT and SAT). However, a growing body of research indicates that testing can also be a powerful learning experience (for a review, see Roediger & Karpicke, 2006a). In a typical investigation of what is widely known as the *testing effect,* researchers compare the effects of testing to other types of study. For example, Roediger and Karpicke (2006b, Experiment 1) had students read two prose passages on unfamiliar biology topics. After reading each passage, students either produced a free recall of the text (a type of test that emphasizes memory retrieval) or spent an equivalent amount of time restudying the passage. On free recall tests two days or one week later, students produced more extensive free recalls of the text that they had been previously tested on than of the text that they had restudied, demonstrating enhanced learning of tested material.

The Testing Effect

A long history of research has demonstrated that testing can be used to enhance learning (e.g., Abbot, 1909; Gates, 1917; Glover, 1989; Spitzer, 1939). In a typical investigation of the testing effect, participants study verbal materials and are either tested one or more times on these materials or spend an equivalent amount of time restudying them. After a delay of one or more days, all participants are tested. On this delayed test, participants who were initially tested on a set of materials typically outperform participants who restudied the materials. The testing effect (defined as better learning of tested than restudied material) has been reliably demonstrated for word lists (Congleton & Rajaram, 2011; Hogan & Kintsch, 1971) and paired associates (Allen, Mahler, Estes, 1969; Bartlett & Tulving, 1974; Estes, 1960), and – in a growing number of studies – for educationally-relevant materials (Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008; Butler & Roediger, 2007; Chan, McDermott, & Roediger, 2006; Gates, 1917; Glover, 1989; McDaniel & Fisher 1991; Nungester & Duchastel, 1982; Spitzer, 1939). Because of this overwhelming evidence, testing is one of only two instructional practices that the Institute of Education Sciences has determined to have a "strong" level of support (Pashler et al., 2007).

The testing effect literature is characterized by three key findings. First, the benefits of testing usually appear only after a delay (Runquist, 1983; Thompson et al., 1978; Wenger, Thompson, & Bartling, 1980). When final tests are administered immediately, restudying often produces better performance than initial testing. This pattern changes after a delay (usually of a day or more) and often increases with longer and longer delays (Runquist, 1983). For example, Thompson et al. (1978) had participants study word lists under repeated study or repeated test conditions and then take a free recall test after either 20 minutes or 48 hours. On the test administered after only a 20 minute delay, participants who had repeatedly studied the text recalled more words. On the test administered after a 48 hour delay, this pattern reversed, and participants who had been repeatedly tested recalled more. This reversal has been described by some authors as lower levels of forgetting following testing than following restudy (Izawa, 1966; Runquist, 1983)

Testing can also improve the organization of knowledge (Bregman & Weiner, 1970; Congleton & Rajaram, 2011; Gates, 1917; Rosner, 1970; Zaromb & Roediger, 2010). For example, Zaromb and Roediger (2010) found that testing could increase both the subjective and objective organization of participants' delayed recalls. In two experiments, participants learned word lists in repeated study or study followed by testing conditions. Words in each list were presented in a random order, but belonged to one of five taxonomic categories. In the first experiment, repeated testing did not increase the degree to which participants clustered their recall with respect to these underlying categories (category clustering was near ceiling in all conditions), but it did increase the degree to which participants clustered their recalls with respect to their own idiosyncratic ordering[4]. In a second experiment in which the materials were revised to avoid ceiling effects, testing increased objective organization as demonstrated by the degree to which participants recalled items from the same category together[5].

A third finding that is not as well established as the first two is that testing also increases the speed with which tested information can be retrieved in the future (Allen et al., 1969; Eimas & Zeaman, 1963; Izawa, 1966). For example, Allen et al. (1969) had participants learn paired associates (A-B pairs) under a variety of repeated study or testing conditions. After a twenty four hour delay, participants were given a series of cued recall tests (A-). Participants were more accurate and faster in producing responses

---

[4] Subjective organization was measured using Sternberg & Tulving's (1977) paired frequency measure.
[5] Objective organization was measured using ARC scores (Roenker, Thompson, & Brown, 1971).

(B’s) when they had been previously tested. These faster responses reflect an increase in the subsequent accessibility of tested information. This increase in accessibility may be especially important for promoting text comprehension, because successful comprehension is contingent on the fluent retrieval of relevant information.

Proposed Mechanisms for the Testing Effect

Most contemporary accounts of the testing effect attribute the benefits of testing to retrieval from LTM (Allen et al., 1969; Carrier & Pashler, 1992; Glover, 1989). These accounts are built on the assumption that retrieval directly alters the state of an item in memory (Bjork, 1975) and can be contrasted with alternative models that attribute the benefits of testing to the amount of processing or number of presentations that are produced by testing (cf, Dempster 1996; Glover, 1989).

Amount of processing/rehearsal/study accounts. Prior to the widespread adoption of retrieval-based accounts, several authors proposed that testing enhances learning in much the same way as study or presentation trials. According to these accounts, testing enhances memory because it increases the amount of time participants dedicate to processing stimuli or provides participants with self-generated re-presentations of stimuli (Bregman & Wiener, 1970; Thompson et al., 1978). While these accounts can explain increased learning in experiments in which testing is contrasted with a condition that involves no re-presentation or processing of stimuli (e.g., Bartlett, 1977; Bartlett & Tulving, 1974; Darley & Murdock, 1971), they cannot explain the results of experiments in which testing is contrasted with a study condition in which stimuli is presented for an equivalent or even greater amount of time (Allen et al., 1969; Carrier &

Pashler, 1992; Hogan & Kintsch, 1970; Kuo & Hirshman, 1996). For example, Carrier and Pashler (1992) compared participants' learning of paired-associates in a pure study condition that involved continuous presentation of A-B pairs for 10 seconds to learning in a hybrid testing condition that involved five seconds of testing (A- presentation) followed by five seconds of full A-B pair presentation. Despite the advantage of the pure study condition in terms of the amount of presentation time, on a delayed cued recall task (A-) participants produced significantly more associates (B's) in the test condition than the study condition. In the case of studies in which an equivalent amount of restudy time is contrasted with a free recall testing condition, accounts based on time and/or the number of presentations have difficulty explaining test-enhanced learning. Not only is time equated in these studies, but since participants rarely reproduce 100 percent of stimuli during free recall testing, they are usually exposed to far fewer re-presentations of stimuli during testing than during restudy.

Retrieval-based accounts. In addition to the results of studies that have carefully equated the amount of time or number of presentations in study and test conditions, there are several additional reasons to believe that retrieval from LTM is the operant mechanism in producing the testing effect. First, the testing effect is much weaker when testing only requires producing information from WM rather than retrieving information from LTM (Bjork, 1975; Bjork & Whitten, 1974; Gotz & Jacoby, 1974; Landauer & Eldridge, 1967). For example, Gotz and Jacoby (1974) had participants learn short word lists in testing conditions where testing was either immediate or delayed by a number subtraction task. Consistent with retrieval-based accounts of the testing effect, immediate testing (which presumably elicited information held in WM) was far less effective than delayed testing (which presumably elicited the retrieval of information from LTM). In educational contexts, the assumption that LTM retrieval plays a key role in producing the testing effect is supported by several decades of adjunct-question research. In a meta-analysis, Hamaker (1986) found that post-reading questions (i.e. questions that require retrieval from LTM) were more effective than pre-reading questions which may produce online attentional effects (Lewis & Mensink, 2012; Rothkopf & Billington, 1975) but may not elicit retrieval from LTM.

A second reason to believe that LTM retrieval is the operant mechanism in the testing effect is that the effect tends to be more robust with recall than recognition tests (Carpenter & Delosh, 2006; Glover, 1989). For example, Glover (1989, Experiment 4) had students read a prose passage and then take one of three tests: a free-recall test, a cued-recall test, or a recognition test that involved deciding whether a statement had occurred in the passage. After a two-day delay, participants took a final free-recall, cued-recall, or recognition test. On all final tests, participants who were initially tested with free-recall performed better than participants who were initially tested with cued-recall, who performed better than participants who were initially tested with a recognition test. The greater efficacy of recall tests can be attributed to the fact that they are more likely to elicit retrieval than are recognitions tests. In educational contexts, the differential effects of recall and recognition are indicated by studies that have found that short-answer tests tend to produce more robust testing effects than do multiple-choice tests (Butler & Roediger, 2007; Kang, McDermott, & Roediger, 2007).

A third reason to believe that LTM retrieval is the operant mechanism in the testing effect is that the size of the testing effect is highly sensitive to manipulations that impact retrieval processes. Manipulations that impact the type, difficulty, or spacing of retrieval have all been found to influence the magnitude of the testing effect (Bartlett, 1977; Bartlett & Tulving, 1974; Gardiner, Craik, & Bleasdale, 1973).

Although most contemporary accounts posit that retrieval from LTM plays a critical role in the testing effect, there are some potentially important differences in how different authors frame the role of retrieval. While some authors have posited that retrieval enhances learning by establishing multiple retrieval routes or a mnemonic mediator (Bjork, 1975; Pyc & Rawson, 2010), others have appealed to the difficulty or depth of retrieval involved in initial practice (Bartlett, 1977; Bjork, 1975; Gardiner et al., 1973; Jacoby, 1978) to account for the impact of initial retrieval practice on subsequent test performance. In the next chapter, we propose an alternate set of mechanisms that generates new predictions for how testing can potentiate future comprehension and learning.

Chapter 3: Using Testing to Potentiate Future Learning

Prior research has convincingly demonstrated that testing enhances past learning. The novel claim of this dissertation is that testing can also potentiate future learning, facilitating the acquisition of new information. This proposal is inspired by studies that have demonstrated the key role of prior knowledge in text comprehension (outlined in Chapter 1) as well as previous studies of testing that have documented different kinds of potentiation, but never of future learning.

Previous Studies of Test-Potentiation

Several previous studies have demonstrated that testing can potentiate learning and performance after the initial testing event. For example, taking a short answer test can potentiate future performance on multiple-choice questions that target the same tested information (Kang et al., 2007; McDaniel, Anderson, Derbish, & Morrisette, 2007). The most direct antecedents for the current research are the handful of studies that have demonstrated that testing can potentiate the learning of previously encountered information that has been tested, but not successfully produced (Congleton & Rajaram, 2011; Izawa, 1966, 1967).

Potentiating learning of previously encountered information. A small number of studies have demonstrated that testing can potentiate the learning of previously encountered ideas upon re-exposure (Congleton & Rajaram, 2011; Izawa 1966, 1967). Izawa (1967) found that the occurrence of test trials increased the effectiveness of subsequent presentation (study) trials of previously encountered information. In this study, participants learned paired associates under four conditions that included the same number of study trials (A-B) but different numbers of test trials (A-). One of these conditions involved twice as many test trials as the others. Although this condition initially showed the slowest learning, after a few trials, this condition produced faster learning than the other three conditions and eventually produced the lowest number of errors. The increased testing in this condition had potentiated participants learning from re-presentations of previously encountered pairs.

A more recent study that utilized a more naturalistic paradigm, collaborative recall, found that initial testing increased the amount of information that participants subsequently learned from other group members (Congleton & Rajaram, 2011). During the first phase of this study, participants learned word lists on their own. Next, participants either repeatedly restudied the word lists or were repeatedly tested with free recall tests. Later, participants collaborated in triads to produce a group recall. Participants who had been repeatedly tested were able to learn more from re-exposure to words that their group-members produced but they themselves had not remembered.

Potentiating transfer across tests. A growing number of studies have demonstrated that testing can facilitate the transfer of information learned in the past from one kind of test to another (Carpenter & Delosh, 2006; Glover, 1989; Kang et al., 2007; McDaniel et al., 2007; Rohrer, Taylor, & Sholar, 2010). For example, initial short-answer testing can improve subsequent performance on related multiple-choice questions (Kang et al., 2007; McDaniel et al., 2007) and initial testing with factual questions can improve subsequent performance on inferential questions (Butler, 2010; Johnson & Mayer, 2009). This is a powerful finding given that transfer is typically reduced when encoding tasks and retrieval tasks differ (Morris, Bransford, & Franks, 1977; Roediger & Blaxton, 1987).

A related finding is that testing promotes transfer from old problem contexts to new problem contexts. Butler (2010) showed that participants who were initially tested with questions like “A bat has a very different wing structure from a bird. What is the wing structure of a bat like relative to that of a bird?” were able to successfully transfer the knowledge built by answering these test questions to new questions like “The U.S. Military is looking at bat wings for inspiration in developing a new type of aircraft. How would this new type of aircraft differ from traditional aircraft like fighter jets?” Results like these demonstrate that the benefits of initial testing are not limited to enhancing subsequent performance on the same types of tests that were used to promote learning; they show that testing can in fact promote the *future application of old information* to new problems. The model of test-potentiated learning proposed here claims that testing can also promote the *future learning of new information*.

Mechanisms for Test-Potentiated Learning

We propose that testing builds a *generative knowledge base* that enables people to learn more from new, related materials encountered after testing. For example, if a person reads an introductory text about a topic and is then tested on it, she will learn more from subsequent texts that build upon and complement the ideas in the introductory text than if she had simply restudied the introductory text. Generative knowledge bases are characterized by two defining features: (1) They support the *fluent* retrieval of information learned in the past in the service of future learning, and (2) this retrieval is driven by *similarity*, resulting in automatic and effortless access.

Fluency. We propose that testing potentiates future learning by establishing knowledge which can be fluently retrieved in the future. This proposal is based on results that have demonstrated that testing can increase the speed with which tested information is retrieved in the future (Allen et al., 1969; Eimas & Zeaman, 1963; Izawa, 1966). These faster responses reflect an increase in the subsequent accessibility of tested information. This accessibility is particularly important when learning from texts, because text comprehension places high demands on processing resources.

Similarity. Testing increases the fluency with which prior knowledge is retrieved from LTM by establishing memory traces that are evoked automatically and effortlessly on the basis of similarity. This proposal is based on empirical results from the study of discourse comprehension that have shown that new texts automatically evoke featurally similar memory traces from LTM (Albrecht & O'Brien, 1993; Cook et al., 1998; Myers & O'Brien, 1998; O'Brien, Rizzella, Albrecht, & Halleran, 1998). During online comprehension, the propositions of an incoming text are encoded in WM, and these propositions continuously elicit retrieval of potentially relevant traces from LTM (McKoon, Gerrig, & Greene, 1996; Myers & O'Brien, 1998). Traces are retrieved on the basis of their semantic, lexical, or phonological similarity to the contents of working memory (e.g., Cook, Halleran, & O'Brien, 1998; Lea, Rapp, Elfenbein, Mitchel, & Romine, 2008). Once retrieved, this information can be integrated into the reader's emergent representation, facilitating comprehension. We propose that testing establishes memory traces that are richer in their featural content than the memory traces established by restudy. This increases their potential similarity (i.e., featural overlap) with new textual propositions in WM, amplifying the probability and speed of their retrieval and facilitating their integration into emergent representations.

Potentiating Learning from Expository Texts

The generative knowledge bases built by testing promote future comprehension and learning from expository texts in two key ways. First, they enable the fluent integration of tested information into new text representations by establishing featurally-rich memory traces that are subsequently evoked via automatic similarity-based retrieval processes. Second, they promote conceptual changes through the interaction of this highly accessible information with new ideas from refutational texts that negate and/or qualify originally tested ideas.

Promoting the fluent integration of relevant knowledge. As readers proceed through a text, they construct an emergent representation of its meaning by integrating information from LTM with incoming information from the text (Kintsch, 1988, 1998). Testing improves comprehension and learning by increasing the fluency with which relevant prior knowledge can be subsequently accessed, thereby increasing readers' integration of tested information into their emergent representations (see Figure 3.1). Testing increases the fluency of this access by establishing featurally-rich memory traces that are subsequently evoked by similar texts that expand upon, qualify, and even refute tested knowledge. Because similarity-based retrieval processes are automatic and place no demands on WM resources, tested ideas are rendered highly accessible during comprehension without taxing readers' limited processing resources. High accessibility is important because prior knowledge that is not readily available during online comprehension is unlikely to be integrated into a reader's emergent representation (McKoon & Ratcliff, 1992).

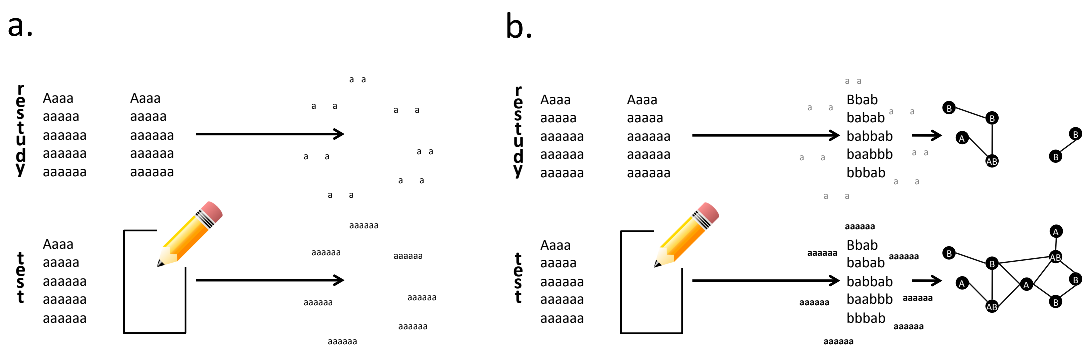

Figure 3.1. Test-potentiated learning from expository texts. (a) Testing builds generative knowledge bases composed of featurally-rich memory traces (b) When tested readers read a new text, they fluently retrieve information from these generative knowledge bases, allowing them to integrate this information into their emergent text representations, improving comprehension and learning.

Preparing readers to revise misconceptions. Another way that testing potentiates learning from expository texts is by preparing readers to revise inaccurate knowledge and erroneous beliefs. Testing does this by maximizing the coactivation of conflicting tested and new information, promoting cognitive conflict and conceptual change. In some cases this sets up an apparent paradox: Testing can both strengthen an inaccurate idea and enable its revision. For example, initial testing may first strengthen a misconception by establishing a memory trace that a reader is even more likely to access in the future. In this case, testing has actually strengthened the misconception. However, if a tested reader subsequently reads a text that explicitly refutes and qualifies this misconception, its greater accessibility increases the probability that the misconception will be accessed, evaluated, and rejected. This prediction is particularly important given the need to forestall scientific misconceptions and to improve scientific literacy more generally.

Chapter 4: Experiment 1

Experiment 1 replicated and extended the testing effect with rich, naturalistic materials – expository texts on topics from psychological science. Because the benefits of testing are not usually apparent when final assessments of learning are administered on the same day as testing (Thompson et al., 1978), this experiment took place over two sessions separated by one week. On Day 1, participants read two expository texts. One text introduced commonly made claims about the validity and utility of learning styles approaches to education, and the other introduced commonly made claims about the efficacy of using discovery learning in math and science education. Participants were tested on one text and restudied the other for an equivalent amount of time. On Day 8, participants were tested on both texts. We predicted that participants would recall a higher proportion of ideas from the tested text than the restudied text. Corroboration of this prediction is necessary to set the stage for the investigations of test-potentiated learning in Experiments 2 and 3.

Method

Participants. Thirty-two participants (*M* age = 19.05 years, 21 female) at a large American university participated for course extra credit or a $15 gift card. All participants were native speakers of English. Participants were tested in groups or individually.

Design. A within-participants design was employed in which the sole factor was the Day 1 Learning Condition (Test, Restudy). In the Test condition, participants read a text for four minutes, completed math problems for two minutes, and then completed a free recall of the text for four minutes. The purpose of the math problems was to reduce readers' access to the surface form of the text and to increase participants' reliance on LTM retrieval during the recall task. A recall test was chosen because it emphasizes retrieval from LTM, which is thought to be the primary cognitive mechanism that drives test-enhanced learning (Carrier & Pashler, 1992), and because recall tests produce larger testing effects than other types of tests (Glover, 1989). In the Restudy condition, participants read a different text for four minutes, completed math problems for two minutes and then were asked to reread the text for four additional minutes.

One week later (Day 8), participants were tested on both texts to assess whether initial testing or restudy had resulted in greater learning. The primary dependent variable was the proportion of idea units that participants recalled from the Day 1 texts. We also collected responses to Likert-type items measuring familiarity with and belief in learning styles and discovery learning as well as their interest in the experimental texts.

Materials. Two texts were constructed to approximate the passages that participants might encounter in the popular psychology literature (see Appendix A for texts). One argued in favor of using "learning styles" to improve education, and included ideas like "learning is best when instruction matches a student's style." This *Pro Learning Styles* text was 393 words long and contained 39 idea units. The other text argued in favor of using "discovery learning" to help students in math and science education, and conveyed ideas like "when students are allowed to discover scientific principles, they are conducting experiments just like a real scientist." The *Pro Discovery*

Learning text was 407 words long and contained 39 idea units. These texts were constructed from popular trade books, online sources, and news articles.

Procedure. On Day 1, participants first completed an informed consent process. They were then given a packet and informed that they would read two texts, complete math problems, restudy one of the texts, and recall as much as they could of the other text. Participants were also informed that they would be tested on both texts the following week. In greater detail, the experimental packets were divided into seven sections. Each section had a fixed amount of time allotted to it, and participants were informed when to start and stop working on each section. In the first section, participants read one of the two texts for four minutes. If they reached the end of the text before time elapsed, then they were instructed to reread the text until it was time to move on; this served to equate time on task in the Test and Restudy conditions. In the second section, participants completed simple math problems for two minutes. In the third section, participants either restudied the text they had read in the first section or recalled as much of the text as they could. Participants were encouraged to recall as much as they could even if they could only recall the gist of an idea, and were told not to worry about spelling or punctuation. The third section lasted four minutes. The fourth section comprised two more minutes of math problems. Over the fifth, sixth, and seventh sections, participants read the other text for four minutes, completed math problems for two minutes, and either recalled the second text (if they had restudied the first text) or restudied the second text (if they had recalled the first text) for four minutes. The order of text presentation and recall versus restudy were counterbalanced across participants. Finally, participants were reminded to return one week later and were excused.

One week later (Day 8), participants returned to complete the experiment. They were informed that they would be recalling the texts they read the week before and completing more math problems. They were then given a new packet divided into five sections. In the first section, participants recalled as much as they could from either the "text you read about Discovery Learning" or the "text you read about Learning Styles" for five minutes. In the second section, participants completed short math problems for two minutes. In the third section, participants recalled as much as they could from the text they did not recall in the first section, again for five minutes. In the fourth section, participants completed two additional minutes of math problems. The fifth section

Table 4.1. *Likert-Type Items Used to Measure Belief, Interest, and Prior Knowledge.*

Domain	Learning Styles	Discovery Learning
Belief that practice is effective	*I believe that instruction is most effective when it is provided in a mode that matches a learner's style.*	*I believe that people learn most when they get to discover concepts.*
Interest in text	*I found the text on Learning Styles interesting.*	*I found the text on Discovery Learning interesting.*
Prior-knowledge about topic	*I already knew a lot about Learning Styles.*	*I already knew a lot about Discovery Learning.*

Note: Response options ranged from "1-Strongly Disagree " to "7-Strongly Agree ".

contained Likert-type items measuring their familiarity with and belief in learning styles and discovery learning as well as their interest in the texts (see Table 4.1). The order of text recall was counterbalanced across participants.

Results

The primary research question for Experiment 1 was whether, on Day 8, participants would recall a higher proportion of ideas from the text they had been tested on a week before than from the text they had restudied. To assess this, each participant's recalls were parsed into idea units and coded by the first author and a research assistant who was blind to the research question. Throughout coding, experimental condition was masked. Reliability was high (*Cohen's kappa* = .95 on 528 in common idea units) and disagreements were resolved through discussion.

Free recalls. Figure 4.1 displays the mean proportion of idea units recalled on Day 8 as a function of Day 1 learning condition. Participants recalled a higher proportion (*paired-t*(31) = 2.88, $p = 0.007$, *Cohen's d* = .62) of the idea units from the tested text ($M = .18$, $SE = .02$) than from the restudied text ($M = .13$, $SE = .01$). Post-hoc comparisons showed that these learning increases held for both texts. Participants who were tested on the learning styles text recalled marginally more ideas ($M = .21$, $SE = .02$) from the text than participants who restudied it ($M = .16$, $SE = .02$) ($t(30) = 1.85$, $p = 0.07$), and participants who were tested on the discovery learning text also recalled marginally more ideas ($M = .14$, $SE = .02$) from that text than participants who restudied it ($M = .11$, $SE = .01$) ($t(30) = 1.78$, $p = 0.08$).

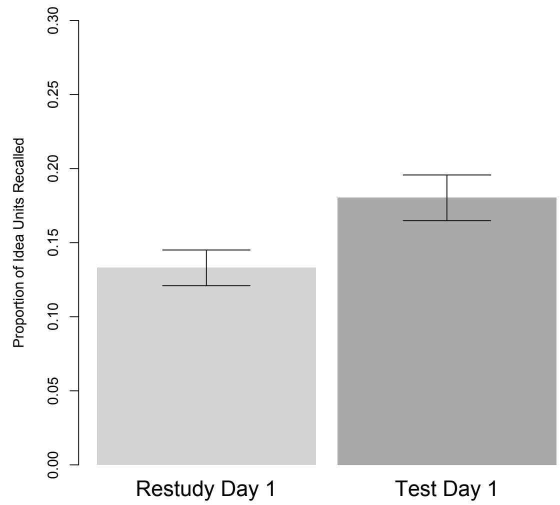

Figure 4.1. Proportion of idea units recalled on Day 8 as a function of Day 1 learning condition; error bars represent standard errors. Results reveal a testing effect, with D1 ideas better recalled if information was initially tested versus restudied.

Belief, interest, and prior knowledge ratings. To investigate whether Day 1 testing influenced the strength of self-reported beliefs, we compared the strength of participants' belief in the practice they were tested on to the strength of their belief in the practice they had restudied (see Table 4.2). To account for differences in the average strength of belief in the efficacy of discovery learning and learning styles, we first normalized responses by mean-centering responses to each statement[6]. Participants did not report stronger beliefs in either Tested or Restudied topics (*paired-t*(30) = .007, *p* = .99) Note that one participant did not complete the Likert-type items. Parallel analyses revealed that participants did not report significantly different levels of interest (*paired-t*(30) = .16, *p* = .87) or prior knowledge (*paired-t*(30) = .25, *p* = .80) in Tested or Restudied topics.

[6] Responses were mean-centered using data from both Experiment 1 and Experiment 2.

Table 4.2. *Mean (SE) Belief, Interest, and Prior Knowledge Ratings for Tested and Restudied Topics.*

Domain	Tested	Restudied
Belief that practice is effective	.50 (.27)	.49 (.26)
Interested in text	-.09 (.25)	-.18 (.25)
Prior-knowledge about topic	-.24 (.22)	-.18 (.23)

Note: Responses were mean-centered for each statement to account for differences in agreement rates between statements. Mean-centering was performed on data from both Experiments 1 and 2.

Table 4.3. *Mean (SE) Belief, Interest, and Prior Knowledge Ratings Disaggregated by Text.*

Domain	Statement	Tested on Discovery Learning	Tested on Learning Styles	$t(29)$	p
Belief that practice is effective	I believe that people learn most when they get to discover concepts.	4.75 (.48)	4.60 (.39)	.24	.81
	I believe that instruction is most effective when it is provided in a mode that matches a learner's style.	5.75 (.36)	5.60 (.25)	.34	.75
Interest in text	I found the text on Discovery Learning interesting.	5.13 (.27)	4.80 (.39)	.69	.50
	I found the text on Learning Styles interesting.	5.19 (.25)	4.73 (.34)	1.08	.29
Prior-knowledge about topic	I already knew a lot about Discovery Learning.	4.25 (.37)	4.20 (.35)	.10	.92
	I already knew a lot about Learning Styles.	4.69 (.35)	4.80 (.35)	.23	.82

Note: Response options ranged from "1-Strongly Disagree " to "7-Strongly Agree ".

Additional post-hoc comparisons (see Table 4.3) confirmed that testing did not influence self-reports of belief in a practice's efficacy, interest in the texts, or prior knowledge for either topic.

The disaggregated belief ratings presented in Table 4.3 illustrate two important points. First, participants believe strongly that these educational practices are effective. Second, participants believed more strongly in the efficacy of learning styles ($M = 5.68$, $SE = .22$) than discovery learning ($M = 4.68$, $SE = .31$) (*paired-t*(30) = 2.78, $p < .01$, *Cohen's d* = .68)

Discussion

Experiment 1 demonstrated a testing effect for naturalistic expository texts on topics in psychological science. Participants recalled a significantly higher proportion of idea units from the text they had been tested on one week earlier than from the text they had restudied. This complements a large number of demonstrations that testing can enhance the learning of previously encountered stimuli like word lists and paired associates (e.g., Carrier & Pashler, 1992; Hogan & Kintsch, 1971) as well as a growing body of research showing that testing can enhance the learning of previously encountered information from expository texts (e.g., Glover, 1989; Roediger & Karpicke, 2006a). Experiment 1 also demonstrated that participants believed strongly in the efficacy of popular educational theories, especially learning styles, and that testing had no effect on these beliefs. These results set the stage for Experiment 2, which investigates whether testing produces a generative knowledge base that potentiates learning from new materials encountered in the future, even when these ideas conflict with strongly held beliefs.

Chapter 5: Experiment 2

Experiment 2 investigated whether the benefits of testing transfer to new learning experiences: in this case, comprehending and learning from a new text which qualified, negated, and elaborated initially tested ideas. Based on the proposal that testing potentiates future learning by establishing a generative knowledge base that increases the fluency and automaticity with which information is subsequently retrieved from LTM, we predicted that testing would facilitate participants' comprehension and learning from these new texts. Because highly accessible ideas come into conflict with new textual ideas that explicitly negated and qualified them, we also predicted that testing would facilitate conceptual change.

The procedure for Experiment 2 was the same as the procedure for Experiment 1 on Day 1: Participants read the *Pro Learning Styles* and *Pro Discovery Learning* texts and were tested on one and restudied the other. The procedure differed on Day 8: Participants read two new refutational texts that *elaborated*, *qualified*, and *negated* the information in the Day 1 texts. Participants were then tested on these new texts to assess whether being tested on versus restudying the Day 1 texts potentiated the learning of new information from the Day 8 texts. The major prediction was that participants would recall a higher proportion of ideas that occurred in the new text that was related to the Day 1 text they had been tested on than ideas that occurred in the new text that was related to the Day 1 text they had restudied. Critically, this advantage should even be present for new ideas that negate or qualify ideas in the old texts, signaling conceptual change. We also predicted that testing might attenuate participants' beliefs in the efficacy of learning styles and discovery learning.

Method

Participants. Twenty-four participants (*M* age = 20.23 years, 20 female) at a large American university participated for course extra-credit or a $15 gift card. All were native speakers of English.

Design. As in Experiment 1, a within-participants design was employed in which the sole factor was the Day 1 Learning Condition (Test, Restudy). The key difference was in the dependent measures. On Day 8, participants read two new texts, one related to the text on which they had been tested on Day 1 and the other to the text they had restudied on Day 1. The new texts contained different types of ideas: D1D8 idea units occurred in both the Day 1 and Day 8 texts, whereas D8 idea units were only present in the D8 text. D8 idea units were further divided into two subtypes. Refutational D8 idea units built upon previously encountered D1D8 idea units by explicitly negating or qualifying them. For example, the D8 idea unit "Discovery learning is not a new idea" explicitly negates the D1D8 idea unit "Discovery learning is a new idea". By contrast, elaborative D8 idea units introduced new content unrelated to D1D8 idea units. For example, the D8 idea unit "the brain learns through multiple senses working together" does not explicitly build upon any of the D1D8 idea units in the Day 1 *Pro Learning Styles* text. After reading each new text, participants completed a recall measure on that text. The major dependent variables were the proportion of D1D8 and D8 idea units that participants recalled from the new texts. We also measured their familiarity with and belief in learning styles and discovery learning after reading the refutational texts on Day 8 using the Likert-type items used in Experiment 1.

Materials. On Day 1, participants read the same *Pro Learning Styles* and *Pro Discovery Learning* texts used in Experiment 1. On Day 8, participants read two new texts that refuted the Day 1 texts. The *Anti Learning Styles* text introduced scientific criticisms of the validity and utility of learning styles approaches to education. The *Anti Discovery Learning* text presented scientific criticisms of the effectiveness of pure discovery learning in math and science education. The content of these refutational texts was inspired by influential review articles from the psychological literature (Mayer, 2004; Pashler et al., 2008).

To understand the relations between the original Day 1 texts and the new Day 8 texts, consider the following excerpt from the *Pro Learning Styles* text read on Day 1:

> It is clear that people think and learn in different ways. Like most people, you have probably had a teacher who just did not work for you. This was probably due to the fact that the teacher's instructional style did not match your learning style. More and more educators are recognizing that some people learn visually, others learn verbally, and still others learn kinesthetically. Participants learn best when instruction matches their particular style. For example, a visual learner learns best with visual instruction and a verbal learner learns best with verbal instruction.

This text makes a number of widely held, but empirically unsubstantiated claims. Next, consider an excerpt from the *Anti Learning Styles* text read on Day 8, which questions many of these claims:

> The idea that people learn in different ways has grown in popularity in recent years. Most learning styles theories claim that some people learn verbally and others learn visually. The appeal of these theories is often based on anecdotal evidence like most people's experience that some teacher just did not work for them. Unfortunately, there is no strong evidence that learning is best when instructional format matches a student's putative style.

This text contains a combination of previously encountered D1D8 ideas and new D8 ideas. For example, "learning is best when instruction matches a student's style" is a D1D8 idea, whereas "there is no strong evidence that learning is best when instruction matches a student's style" is a refutational D8 idea. (See Table 5.1 for additional examples.) The *Anti Learning Style* text was 522 words long and contained 22 previously encountered D1D8 idea units and 43 new D8 idea units. The *Anti Discovery Learning* text was 531 words long and contained 19 D1D8 idea units and 47 D8 idea units.

Procedure. Experiment 2 took place over two sessions separated by one week. The Day 1 procedure was the same as in Experiment 1. When participants returned on Day 8, they were informed that they would read two new texts on educational issues, complete more math problems, and recall as much as they could from the new texts.

Table 5.1. *Example D1D8 Idea Units (Idea Units Present in Both Texts) and D8 Idea Units (Idea Units Present Only in the Day 8 Text).*

Text Pair	D1D8 idea unit	D8 idea unit	Day 1 Text Source	Day 8 Text Source
Learning Styles	Students are happier when instruction matches style	Although students are happier when instruction matches preferred style, they do not learn more. **Refutational**	*Additionally, when students receive instruction that matches their style, they are much happier and more engaged in the lesson.*	*Although students are definitely happier when they receive instruction that matches their preferred style, they do not appear to learn more.*
Learning Styles	NA	Massa and Mayer conducted a study in 2006. **Elaborative**	NA	*For example, consider a study by Massa and Mayer (2006).*
Discovery Learning	Students who learn through discovery are better able to apply concepts to new problems in the future.	There is no evidence that students who learn through discovery are better able to apply concepts to new problems in the future. **Refutational**	*Because they have already encountered significant hurdles during learning, students who learn through discovery are better able to apply previously learned concepts to new problems in the future.*	*In addition to not promoting initial learning, there is no strong evidence that discovery learning improves students' ability to apply learned concepts to new problems.*
Discovery Learning	NA	Background knowledge can reduce the demands that discovery places on cognitive resources. **Elaborative**	NA	*Background knowledge and prior experience can reduce the demands that discovery learning places on cognitive resources.*

They were then given an experimental packet divided into eight sections. In the first section, participants read one of the two new refutational texts for six minutes. In the second section, participants completed simple math problems for two minutes. In the third section, participants recalled as much of the text they read in the first section as they could for seven minutes. The fourth section comprised two more minutes of math problems. During the fifth section, participants read the other new refutational text for six minutes, and during the sixth section they solved math problems for two minutes. In the seventh section, participants recalled as much as they could from the text they read in the fifth section for seven minutes. The final section contained Likert-type items measuring familiarity with and belief in learning styles and discovery learning as well as interest in the texts. The order of text presentation on Day 8 was counterbalanced across participants.

Results

The major research question for Experiment 2 was whether participants would learn a higher proportion of ideas from a new text that refuted an old text if they had been tested on the old text a week earlier as opposed to having restudied it. To assess this, participants' recalls were parsed into D1D8 idea units and D8 idea units (both refutational and elaborative) and coded by the first author and a research assistant who was blind to the research questions. Reliability was high (*Cohen's kappa* = .92 on 744 in common idea units) and disagreements were resolved through discussion.

Free recalls. Figure 5.1a displays the mean proportion of D1D8 and D8 idea units recalled from the new texts read on Day 8 as a function of the learning condition for the related Day 1 texts. Participants recalled a higher proportion (*paired-t*(23) = 2.73, *p* = .01, *Cohen's d* = .55) of the D1D8 ideas common to the old and new texts if they had been tested on the old text (*M* = .25, *SE* = .03) than if they had restudied it (*M* = .19 *SE* = .02). This replicates the testing effect documented in Experiment 1.

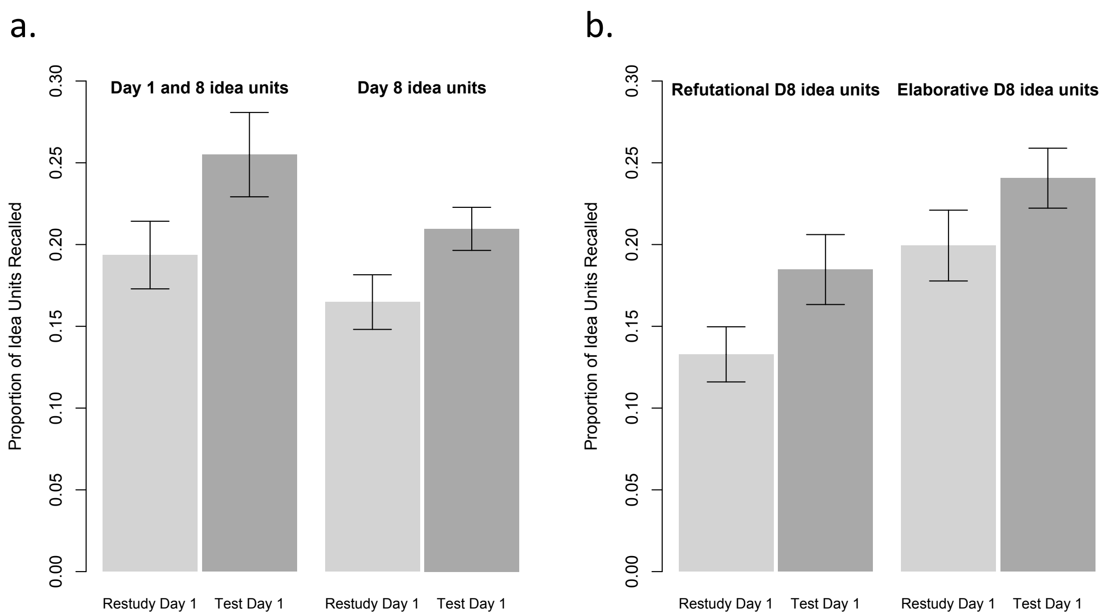

Figure 5.1. Proportion of idea units recalled on Day 8 as a function of Day 1 learning condition for the related text; error bars represent standard errors. a) Experiment 2 replicated the testing effect observed in Experiment 1 for D1D8 ideas; with D1D8 ideas better recalled if information was initially tested versus restudied. Critically, Experiment 2 also revealed a new potentiated learning effect, with new D8 ideas better recalled if Day 1 texts were tested versus restudied, presumably due to the generative knowledge base established by testing on Day 1. b) Test potentiated learning was reflected by better recall of new D8 ideas that refuted/qualified D1D8 ideas and marginally better recall of new ideas that elaborated previous ideas without explicitly building on them.

Critically, participants recalled a higher proportion (*paired-t*(23) = 3.30, *p* = .003, *Cohen's d* = .81) of the new D8 ideas if they had been tested on the related Day 1 text (*M* = .21, *SE* = .01) than if they had restudied it (*M* = .16, *SE* = .02). This demonstrates that initial testing potentiated the learning of new information. Next, we separately analyzed the two classes of D8 idea units (see Figure 5.1b). Participants recalled a higher proportion (*paired-t*(23) = 2.52, *p* = .02, *Cohen's d* = .54) of refutational D8 idea units, which qualified and negated D1D8 idea units, if they had been tested on the old text (*M* = .18, *SE* = .02) than if they had restudied it (*M* = .13 *SE* = .02). This was also true for elaborative D8 idea units, which introduced new content, but the effect was only marginal: Participants recalled a marginally higher proportion (*paired-t*(23) = 1.90 , *p* = .07) of elaborative D8 idea units if they had tested on the old text (*M* = .24, *SE* = .02) than if they had restudied it (*M* = .20 *SE* = .02).

Post-hoc comparisons showed that these learning increases were significant only for the discovery learning texts (see Table 5.2). Participants who were tested on the Day 1 *Pro Discovery Learning* text recalled more D1D8 ideas, as well as more D8 ideas units (including refutational but not elaborative ideas), from the Day 8 *Anti Discovery Learning* text than participants who had restudied the first text. Participants who were tested on the Day 1 *Pro Learning Styles* text did not recall significantly more of any of these idea types from the new *Anti Learning Styles* text than did participants who had restudied the first text. One reason for this discrepancy may be the greater strength of participants' belief in the efficacy of Learning Styles practices identified in Experiment 1.

Table 5.2. *Mean (SE) Proportional Recall Disaggregated by Text*

	Discovery Learning				Learning Styles			
	Restudy	Test	*t*(22)	*p*	Restudied	Tested	*t*(22)	*p*
D1D8	.13 (.02)	.29 (.04)	3.63,	<.01	.26 (.02)	.22 (.03)	.98	.34
D8	.15 (.02)	.22 (.02)	2.13	.04	.18 (.03)	.20 (.02)	.72	.48
D8 Refutational	.12 (.02)	.22 (.04)	2.56	.02	.15 (.03)	.16 (.02)	.15	.88
D8 Elaborative	.19 (.03)	.22 (.02)	1.02	.32	.22 (.03)	.26 (.03)	.90	.37

Belief, interest, and prior knowledge ratings. We next evaluated whether testing, reading a refutational text, or the interaction of these two experiences affected beliefs about learning styles and discovery learning. To do this, we analyzed participants' beliefs about these practices gathered at the end of both Experiments 1 and 2. Recall that participants in Experiment 1 only read texts that reinforced commonly held beliefs about learning styles and discovery learning. By contrast, participants in Experiment 2 also read texts refuting these commonly held beliefs. Table 5.3 presents participants' mean levels of belief in the efficacy of the tested and restudied topics in Experiments 1 and 2. To account for differences in mean rates of agreement for different statements, responses were mean-centered by statement.

To analyze this belief data, we performed a mixed model Analysis of Variance (ANOVA) with Day 1 Learning Condition (Test, Restudy) as a within subjects factor and

Experiment (E1, E2) as a between subjects factor. Note that one participant in Experiment 1 did not respond to these items. For the belief that an approach is effective, the main effect of Experiment was significant (F (1,53) = 15.61, MSE = 34.90, p < .001), but the main effect of Learning Condition was not (F (1,53) = .11, MSE = .21 p = .74). The Experiment x Learning Condition interaction was also not significant (F (1,53) = .15, MSE = .30, p = .69). The main effect of Experiment shows that reading refutational texts attenuated beliefs in the efficacy of learning styles and discovery learning. The absence of an interaction between Learning Condition and Experiment fails to support the prediction that testing can amplify the belief change caused by reading a refutational text.

Table 5.3. *Mean (SE) Belief, Interest, and Prior Knowledge Ratings for Tested and Restudied Topics by Experiment.*

	Experiment 1		Experiment 2	
	Tested	Restudied	Tested	Restudied
Belief that approach is effective	.50 (.27)	.49 (.26)	-.74 (.27)	-.54 (.31)
Interested in text(s)*	-.09 (.25)	-.18 (.25)	.09 (.27)	.26 (.26)
Prior-knowledge about topic	-.24 (.22)	-.18 (.23)	.11 (.27)	.44 (.22)

Note: Responses were mean-centered for each statement to account for differences in agreement rates between statements.

* Statements read “text” in Experiment 1 and “texts” in Experiment 2.

We also conducted ANOVAs on responses that measured interest in the text(s) as well as prior knowledge about these two practices. For self-reported interest, there was a marginal main effect of Experiment, with participants who had read the additional refutational texts reporting higher levels of interest (F (1,53) = 2.98, MSE = 6.51, p = .09). Neither the main effect of Learning Condition (F (1,53) = 1.06, MSE = .88, p = .31), nor the interaction (F (1,53) = .61, MSE = .51, p = .44) were significant. For self-reported prior knowledge, neither main effect was significant, nor was the interaction (all Fs < 1.05).

Additional post-hoc comparisons (see Tables 5.4 and 5.5) showed that participants in Experiment 2 reported lower levels of agreement with both the statement “I believe that instruction is most effective when it is provided in a mode that matches a learner’s style” and the statement “I believe that people learn most when they get to discovery concepts” than did participants in Experiment 1. This suggests that reading a refutational text drove changes in beliefs about both of these educational practices. Post-hoc comparisons also showed that the higher levels of interest associated with reading a refutational text were significant only for the statement that asked readers about the learning styles text(s).

The only evidence that is congruent with the prediction that testing can amplify the belief change caused by refutational texts comes from the marginally lower rates of endorsement for the statement “I believe that people learn most when they get to discovery concepts” provided by participants who were tested on the *Pro Discovery Learning* text in Experiment 2 (see Table 5.5).

Table 5.4. *Mean (SE) Belief, Interest, and Prior Knowledge Ratings Disaggregated by Topic for Experiments 1 and 2.*

		Exp 1	Exp 2	*t*(53)	*p*
Belief that practice is effective	I believe that people learn most when they get to discover concepts.	4.68 (.31)	3.44 (0.27)	2.94	<.01
	I believe that instruction is most effective when it is provided in a mode that matches a learner's style.	5.68 (.22)	4.65 (0.31)	2.80	<.01
Interested in text(s)	I found the text(s) on Discovery Learning interesting.*	4.97 (.23)	5.29 (0.28)	.89	.37
	I found the text(s) on Learning Styles interesting.*	4.97 (.21)	5.62 (0.22)	2.16	.04
Prior-knowledge about topic	I already knew a lot about Discovery Learning.	4.23 (.25)	4.21 (0.32)	.04	.97
	I already knew a lot about Learning Styles.	4.74 (.25)	5.38 (0.21)	1.90	.06

Note: Response options ranged from "1-Strongly Disagree " to "7-Strongly Agree ".

* Statements read "text" in Experiment 1 and "texts" in Experiment 2.

Table 5.5. *Mean (SE) Belief, Interest, and Prior Knowledge Ratings Disaggregated by Text for Experiment 2.*

		Tested on Discovery Learning	Tested on Learning Styles	*t*(22)	*p*
Belief that practice is effective	I believe that people learn most when they get to discover concepts.	3.00 (.43)	3.89 (.30)	1.67	.11
	I believe that instruction is most effective when it is provided in a mode that matches a learner's style.	4.42 (.54)	4.89 (.30)	.74	.47
Interested in text(s)	I found the text(s) on Discovery Learning interesting.*	5.17 (.49)	5.42 (.29)	.44	.66
	I found the text(s) on Learning Styles interesting.*	5.83 (.34)	5.42 (.26)	.97	.34
Prior-knowledge about topic	I already knew a lot about Discovery Learning.	3.75 (.46)	4.67 (.41)	1.48	.15
	I already knew a lot about Learning Styles.	5.08 (.33)	5.67 (.22)	1.44	.16

Note: Response options ranged from "1-Strongly Disagree " to "7-Strongly Agree ".

* Statements read "text" in Experiment 1 and "texts" in Experiment 2.

Discussion

Experiment 2 replicated the testing effect documented in Experiment 1. Participants recalled a significantly higher proportion of previously encountered (D1D8) ideas from the old text they had been tested on one week earlier than from the text they had restudied.

The novel finding of Experiment 2 was that testing potentiated the future learning of new information. When participants were tested on an old text and, one week later, read a new text that refuted and elaborated some of the old text's claims, they recalled a significantly higher proportion of new (D8) ideas than when the old text was simply restudied. This finding extends previous demonstrations of the benefits of testing. It demonstrates that testing can potentiate the learning of new ideas encountered after initial testing, in addition to potentiating the learning of previously encountered ideas that are re-presented after initial testing (Congleton & Rajaram, 2011; Izawa, 1966, 1967) and the transfer of previously learned ideas to new tests (Butler, 2010; Carpenter & Delosh, 2006; Glover, 1989; Kang et al., 2007; McDaniel et al., 2007).

These results support the proposal that testing promotes future learning by producing a generative knowledge base that can be fluently integrated with new textual information. Importantly, successful comprehension and learning from the texts used in this experiment required more than simple integration. The Day 8 texts presented ideas that refuted and qualified the Day 1 texts, requiring a fundamental reorganization of Day 1 ideas. Testing may contribute to this conceptual change by promoting the coactivation of previously held ideas and new refutational text via low-level mechanisms like fluency and similarity.

These results also suggest that beliefs may play a key role in determining whether testing is effective in promoting conceptual change. Post-hoc comparisons showed that although participants who were tested on the *Pro Discovery Learning* text subsequently learned more from the *Anti Discovery Learning* text than those who restudied the original text, testing on the *Pro Learning Styles* texts did not promote subsequent learning from the *Anti Learning Styles* text. This discrepancy may have been due to strong beliefs about the efficacy of learning styles that participants brought to the experiment. These strong beliefs may have prevented participants from revising misconceptions about learning styles. We investigate this question more directly in Experiment 3, by comparing the effects of testing for participants who strongly believe in learning styles to those who are more skeptical of the practice's efficacy.

Chapter 6: Experiment 3

The primary goals of Experiment 3 were to investigate whether testing can potentiate comprehension and learning in the context of two educationally relevant tasks and to investigate whether prior beliefs in a topic modulate test-potentiated learning. Although the results of the first two experiments demonstrated that testing can both enhance the knowledge that readers build and potentiate their future learning of new ideas, these learning gains were demonstrated using free-recall tasks. Free recall is a convenient tool for psychological research, but it is not directly relevant to educational practice. As a first step in demonstrating the potential educational utility of test-potentiated learning, Experiment 3 investigated whether prior testing could improve participants' performance on critical essays and short answer questions by enhancing their past learning and potentiating their future learning and critical comprehension.

Experiment 3 further investigated how prior beliefs in a topic may modulate test-potentiated learning. The results of Experiment 2 suggested that test-potentiated learning may be most effective when participants do not hold strong, pre-existing beliefs that contradict educational materials. In Experiment 3, we test this assumption more directly, by contrasting the effects of testing on materials critical of learning styles practices for participants who strongly believe in these practices to those who are more skeptical of their efficacy.

To investigate these questions, we modified the procedure and materials used in Experiment 2. On Day 1, participants read a text that introduced key criticisms of the validity and utility of "learning styles". Half of the participants were tested on this *Anti*

Learning Styles text, and the other half restudied it for an equivalent amount of time. After a week's delay, all participants read a new *Pro Learning Styles* text that advocated the use of learning styles to improve educational outcomes. They were then asked to write a critical essay about the *Pro Learning Styles* text, to answer a series of short answer questions, and to respond to a series of Likert-type items that were designed to capture their beliefs about the utility and validity of learning styles approaches to education.

There were three central predictions for Experiment 3. Based on previous research into test-enhanced learning (including Experiments 1 and 2), we predicted that initial testing on the *Anti Learning Styles* text would increase that amount of information that students retained from this text and that this would be reflected in their increased inclusion of ideas from the Day 1 text in critical essays and short answer responses on Day 8. Based on the proposed model of test potentiated learning and its preliminary validation in Experiment 2, we also predicted that initial testing on the *Anti Learning Styles* text would increase the amount of information that participants learned from the *Pro Learning Styles* text they read a week later and that this increased learning of new ideas would be reflected by the increased inclusion of ideas from the new text in critical essays. Finally, based on previous research into the powerful role of misconceptions in learning (e.g., Alvermann et al., 1985; Kardash & Scholes, 1996), we predicted that essay performance would be characterized by an interaction between participants' beliefs and learning condition. Specifically, we predicted that participants who strongly endorsed statements in favor of learning styles and who weakly endorsed statements against learning styles would not demonstrate test-enhanced learning of ideas that were critical of learning styles in the context of writing a critical essay. On the other hand, participants who were more skeptical of learning styles were predicted to demonstrate test-enhanced learning of ideas that are critical of the construct and were predicted to include more of these critical ideas in their essay responses.

Method

Participants. Forty-two participants (*M* age = 20.05 years, 22 female) at a large American university participated for course extra-credit or a $15 gift card. The data from one participant was excluded from the subsequent analyses due to a failure to follow instructions. All participants were native speakers of English.

Materials. Two texts were constructed for Experiment 3: one for the Day 1 session and another for the Day 8 session. The Day 1 text introduced key scientific and economic criticisms of the utility and validity of "learning styles" approaches to education. The Day 8 text presented arguments, both anecdotal and evidence-based, in favor of using learning styles to improve educational outcomes. To understand the relation between the Day 8 text and the Day 1 text, consider the following excerpts (see Appendix B for the full texts). First, consider an excerpt from the Day 1 text that presented an argument which was critical of learning styles:

> Another reason for the popularity of learning styles is the widely held belief that all students should be treated by educators as unique individuals. However, rejecting learning styles approaches to education does not mean rejecting individuality. In fact, it may allow educators to focus on differences between

> students that are more important to educational outcomes such as differences in aptitude or personality.

Now, consider an excerpt from the Day 8 text that supported using learning styles to improve education

> Each student is an individual and deserves an education that recognizes it. By tailoring instruction to each student's learning style, we can maximize learning and ensure that each student receives the individualized education they deserve. We need to start assessing each student's learning style and partitioning our classrooms to ensure that each student is taught in the most appropriate way.

These texts were constructed to contain three types of ideas: D1 ideas that were present only in the text read on Day 1, D1D8 ideas that were present in both texts, and D8 ideas that were present only in the text read on Day 8. An example of a D1 idea in the first excerpt is "rejecting learning styles does not mean rejecting individuality". In the context of writing a critical essay or answering the short answer questions posed in this experiment, ideas like this allow a participant to qualify, elaborate, or refute the argument presented in the new Day 8 text. D1D8 ideas were contained in both texts. An example of a D1D8 idea unit from the excerpts above is "all students deserve to be treated as individuals". D8 ideas were only presented by the new text. An example of a D8 idea unit in the second excerpt above is "Learning styles based education is the only way to respect students' individuality". In the context of the essay and short answer tasks, D8 ideas reflect the ideas that participants were expected to refute and qualify.

The Day 1 *Anti Discovery Learning Styles* text was 602 words long and contained 42 D1 idea units that were not repeated in the Day 8 *Pro Learning Styles* text as well as 13 D1D8 idea units that were also presented in the Day 8 text. The Day 8 *Pro Learning Styles* text was 432 words long and contained the 13 D1D8 idea as well as 30 D8 idea units that were not presented in the Day 1 text.

Six short answer questions were constructed that were designed to assess the quality of the participants' critical representations of the Day 8 text. An example question is "5- What would be the economic consequences of the author's plan? Be specific." The short answer questions were designed to primarily elicit information from the Day 1, not the Day 8 text. See Appendix B for the full set of questions.

Six Likert-type items were also constructed to provide a gauge of each participant's background knowledge, beliefs, and interest in the topic. Participants were asked to rate their agreement with statements like "I already knew a lot about the topics covered in the text" on a seven point scale (1-"Strongly Disagree" to 7-"Strongly Agree") (see Table 6.1 on p. 66 for all statements). Four items (the first four in Table 6.1) were designed to reveal the potential impact of learning condition on beliefs about learning styles. For example, lower levels of agreement with statements like "I believe that instruction is most effective when it is provided in a mode that matches a learner's style." by Test participants than by Restudy participants would suggest that testing influenced beliefs about learning styles approaches to education. These four items were also designed to support an exploration of the impact of beliefs on essay and short answer performance by allowing participants to be divided into two groups: LS Believers and LS Skeptics.

Procedure. As in Experiments 1 and 2, Experiment 3 took place over two sessions separated by one week. On Day 1, participants first completed an informed consent procedure. After providing informed consent, participants were given an experimental packet that contained the *Anti Discovery Learning* text, several math problems, and blank pages for the participants who would be asked to produce a recall of the text. Participants were informed that they would be asked to read the text, to complete some math problems, and to either restudy the text or to recall as much as they could from it. Participants were also informed that the text they read in the fist session would be important to the second experimental session and were asked to read and study it carefully.

The Day 1 procedure and the accompanying experimental packet were divided into three sections. Each section had a specific amount of time allotted to it, and participants were informed by the experimenter when to start and when to finish each section. In the first section, participants read the *Anti Learning Styles* text for six minutes. In the second section, participants completed simple math problems for two minutes. In the third section, one half of the participants were asked to restudy the text for six minutes. The other half were asked to recall as much as they could from the text for an equivalent amount of time. As in the first two experiments, tested participants were encouraged to recall as much as they could even if they could only recall the gist of an idea and were also told not to worry about spelling or punctuation. At the end of the Day 1 session, participants were reminded to return one week later and were excused.

One week later (Day 8), participants returned to complete the experiment. Upon returning, participants were informed that they would be asked to read a new text on educational issues, to write a critical essay about the new text, and to answer a set of short answer questions. Participants were then given a new packet. This packet and the procedure for Day 8 were divided into four sections. Unlike the Day 1 session, each section was self-paced and participants were free to move on to the next section as soon as they completed the current one. In the first section, participants read the new *Pro Learning Styles* text. In the second section, participants were asked to write an essay that critiqued the text they had just read. The instructions for this section asked participants to point out “as many potential criticisms or important points” as they could “no matter what” their “personal beliefs … about Learning Styles” were. In the third section, participants answered the six questions that were designed to elicit a critical evaluation of the arguments presented by the text they had just read. In the fourth and final section, participants rated their agreement with the six Likert-type items.

Coding. Each participant’s essay and short answer responses were parsed into idea units and coded by the first author and a research assistant who was blind to the research questions. Reliability was high (*Cohen’s kappa* = .88 on 1488 in common idea units) and disagreements were resolved through discussion.

Results

To preview the results, participants' responses to the essay stem and the short answer questions confirmed the prediction that testing can be used to enhance past learning and to potentiate future learning from expository texts, thereby enabling participants to produce richer responses to educationally relevant learning assessment. Overall, the responses produced by tested participants contained more ideas from both the tested Day 1 text and the new text read on Day 8. The results of Experiment 3 also demonstrated the important role of beliefs in test-potentiated learning. Participants' inclusion of critical ideas from the Day 1 text was driven not by learning condition alone (i.e., testing versus restudy), but by an interaction between their beliefs and learning condition. While tested participants' who were skeptical of learning styles produced more of the critical D1 ideas than did skeptical participants who had simply restudied the Day 1 text, tested participants' who believed strongly in the efficacy and validity of learning styles did not produce more of these critical ideas than participants who had restudied them.

Essay responses. Writing a successful critical essay about the new text required participants to state the claims made by the author of the new text and to present contradictory or qualifying evidence. In the context of this text set, the claims made by the author of the second text included both new D8 ideas and repeated D1D8 ideas, and the critical/ qualifying evidence was represented by the old D1 ideas from the first text. When essay responses were considered by themselves (see Figure 6.1), participants who were tested on Day 1 failed to integrate a higher proportion of D1 ideas (M = .06, SE =

.02) into their critical essays than participants who had restudied the Day 1 text (*M* = .05, *SE* = .01) ($t(39) = .83$, $p > .05$). However, participants who were tested on Day 1 did integrate more D1D8 ideas (*M* = .15, *SE* = .02) into their critical essays than did participants who had restudied the Day 1 text (*M* = .08, *SE* = .02) ($t(39) = 2.6$, $p = .02$,

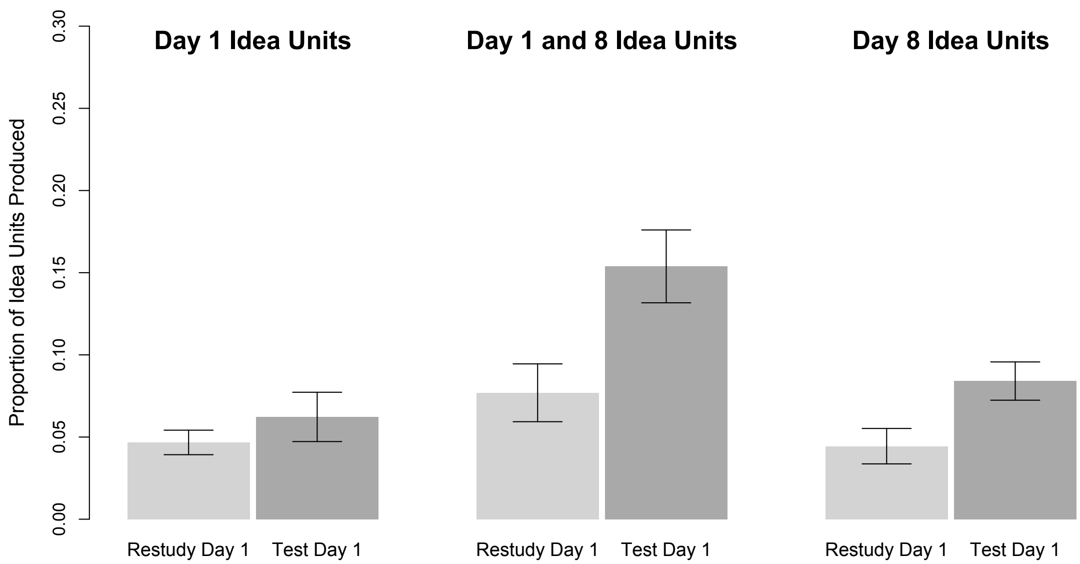

Figure 6.1. Proportion of idea units from the Day 1 text (D1), both texts (D1D8), and the Day 8 text (D8) included in essay responses as a function of Day 1 learning condition. Error bars represent standard errors.

Cohen's d = .83). Participants who were tested on Day 1 also included a higher proportion of the new D8 ideas in their essay responses (*M* = .08, *SE* = .01) than did participants who had restudied the Day 1 text (*M* = .04, *SE* = .01) ($t(39) = 2.44$, $p = .02$, *Cohen's d* = .78).

Question responses. Some participants (e.g., participants who believed strongly in the efficacy of learning styles) may have failed to include critical ideas from the Day 1 text in their essay responses even though they had successfully learned these ideas. To determine whether testing had enhanced the learning of these ideas even though tested participants failed to produce more of them on essay responses, we administered a set of six questions that were designed to as an additional assessment of the participants' learning of critical ideas from the Day 1 text. As predicted, tested participants integrated more critical D1 ideas (M = .10, SE = .01) into their question responses than did students who had restudied the Day 1 text (M = .07, SE = .01) (t(39) = 2.47, p = .02, *Cohen's d* = .79).

Combined essay and question responses. When responses to both the essay stem and the short-answer questions were combined (see Figure 6.2), it was revealed that tested participants produced more of all three types of ideas. Participants who were tested on Day 1 integrated a higher proportion of D1 ideas (M = .13, SE = .02) into their responses than did participants who had restudied the Day 1 text (M = .09, SE = .01) (t(39) = 2.06, p = .046, *Cohen's d* = .66). Participants who were tested on Day 1 also integrated more D1D8 ideas (M = .26, SE = .02) into their critical essays than did participants who had restudied the Day 1 text (M = .17, SE = .02) (t(39) = 2.71, p = .01, *Cohen's d* = .87). The increased inclusion of D1 and D1D8 ideas reflects the enhanced learning of previously encountered information that was generated by testing. The increased inclusion of D1 ideas also reflects a positive impact on critical comprehension, because it reflects an increased use of elaborative and qualifying information to critically evaluate a new text. Participants who were tested on Day 1 also included a higher proportion of the new D8 ideas in their responses (M = .14, SE = .02) than did participants who restudied the Day 1 text (*M* = .09, *SE* = .02) ($t(39) = 2.26$, p = .03, *Cohen's d* = .72). This reflects the potentiation of learning from a new text encountered a week after initial testing.

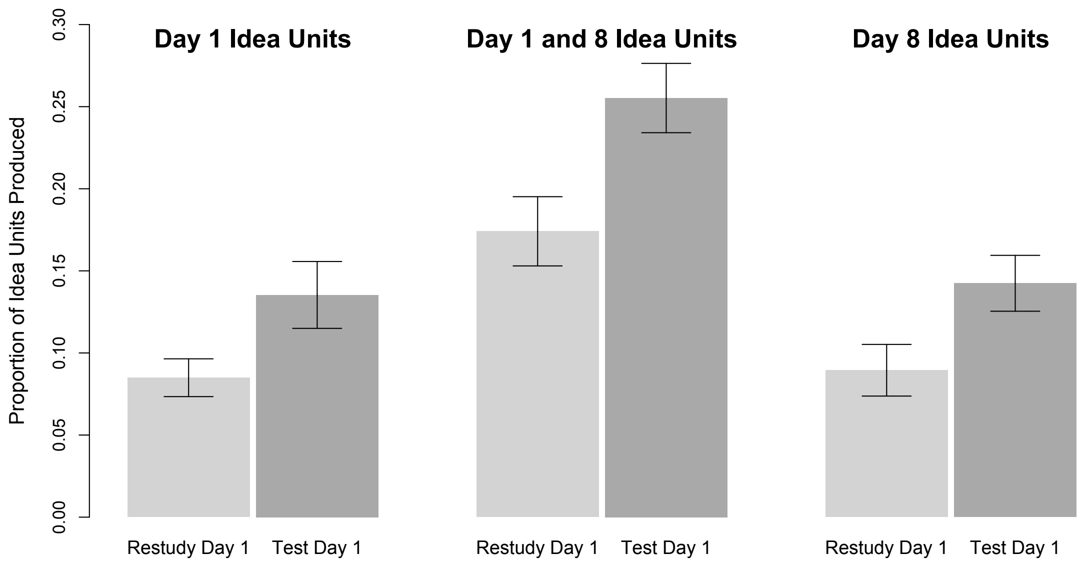

Figure 6.2. Proportion of ideas units from the Day 1 text (D1), both texts (D1D8), and the Day 8 text (D8) included in essay and question responses as a function of Day 1 learning condition. Error bars represent standard errors.

Beliefs about learning styles. For the most part, initial testing did not significantly influence participants' levels of agreement with the Likert-type items (see Table 6.1). However, there were two marginal effects worth noting. Tested participants reported marginally lower levels of agreement with the statement "I believe there is strong evidence for learning styles" than did restudy participants. Tested participants also reported marginally higher levels of agreement with the statement "I believe that other individual differences may be more important than learning styles" than did restudy participants. These responses are congruent with a reduction in belief in learning styles that was driven by testing on the Day 1 text that contained ideas that were critical of the construct.

Table 6.1. *Mean (SE) Likert-Type Item Responses by Day 1 Learning Condition.*

	Restudy Day 1	Testing Day 1	*t*(39)	*p*
I believe that different people learn information in different ways.	5.56 (0.23)	5.4 (0.33)	.39	.70
I believe that instruction is most effective when it is provided in a mode that matches a learner's style.	4.56 (0.24)	4.4 (0.31)	.40	.69
I believe that there is strong evidence for learning styles.	4.75 (0.32)	3.98 (0.3)	1.77	.08
I believe that other individual differences may be more important than learning styles.	5.12 (0.21)	5.65 (0.22)	1.70	.09
I found the text interesting.	4.56 (0.36)	5.25 (0.25)	1.62	.11
I already knew a lot about the topics covered in the text.	4.31 (0.3)	4.62 (0.28)	.75	.45

Note: Response options ranged from "1-Strongly Disagree " to "7-Strongly Agree ".

Beliefs and essay responses. The participants' responses to the Likert-type items as well as their critical essay responses allowed us to examine the potentially important interaction between testing and beliefs. In order to test whether participants' rates of inclusion of critical (D1) ideas were influenced by their beliefs about learning styles, we used the Likert data collected on Day 8 to identify participants who strongly believed in learning styles (LS Believers) as well as participants who were more skeptical (LS Skeptics). Specifically, we separated students into the two groups on the basis of their responses to the four statements that were designed to capture their beliefs in learning styles (the first four statements in Table 6.1). The internal consistency of these four items was acceptable (*Cronbach's alpha* = .79 with the coding of the fourth statement reversed). To divide participants into belief groups, we summed their responses to these four items (reversing the coding of the fourth statement) and performed a median split. Participants who had higher than median levels of agreement with these statements were classified as *LS Believers* ($N = 23$). Participants who had lower than median levels of agreement with these statements were classified as *LS Skeptics* ($N = 18$).

To evaluate whether belief in learning styles influenced participants' inclusion of critical ideas in their essay responses, we performed a 2 x 2 ANOVA on the proportion of D1 ideas included with Day 1 Learning Condition (*Test, Restudy*) and LS Belief (*LS Believers, LS Skeptics*) as between subjects factors. The main effect of Belief (F (1,37) = 9.87, MSe = .03, p = .003) was significant, but the main effect of Learning Condition (F (1,37) = .96 , MSe = .002, p = .33) was not. These main effects were qualified by a significant Belief x Learning Condition interaction (F (1,37) = 6.87, MSe = .02, p = .01) (see Figure 6.3). Tested participants who were skeptical of learning styles included significantly more D1 ideas in their essays than did restudy participants who were skeptical ($t(21) = 2.35$, p = .03, *Cohen's d* = 1.03). Tested participants who believed in learning styles did not include significantly fewer or more D1 ideas in their essays than did restudy participants who also believed in learning styles ($t(18) = 1.31$, $p > .05$).

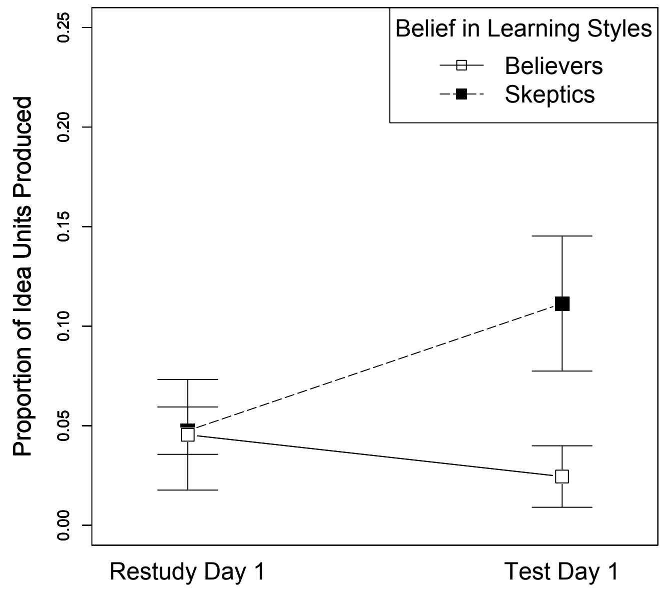

Figure 6.3. Interaction between beliefs and learning condition for the essay writing task. Error bars represent standard errors.

Beliefs and question responses. To evaluate whether belief in learning styles influenced the inclusion of critical information in question responses, we performed the same analysis. The main effect of Belief ($F(1,37) = 2.49$, $MSe = 10.20$, $p = .12$) was not significant, but the main effect of Learning Condition ($F(1,37) = 6.61$, $MSe = 27.00$, $p = .01$) was. These main effects were qualified by a marginal Belief x Learning Condition interaction ($F(1,37) = 2.82$, $MSe = 25.13$, $p = .10$) (see Figure 6.4). Tested participants who were skeptical of learning styles included significantly more D1 ideas in their essays than did restudy participants who were skeptical ($t(21) = 3.44$, $p = .003$, *Cohen's d* =

1.50). Tested participants who believed in learning styles did not include significantly more D1 ideas in their essays than did restudy participants who also believed in learning styles ($t(18) = .18$, $p > .05$).

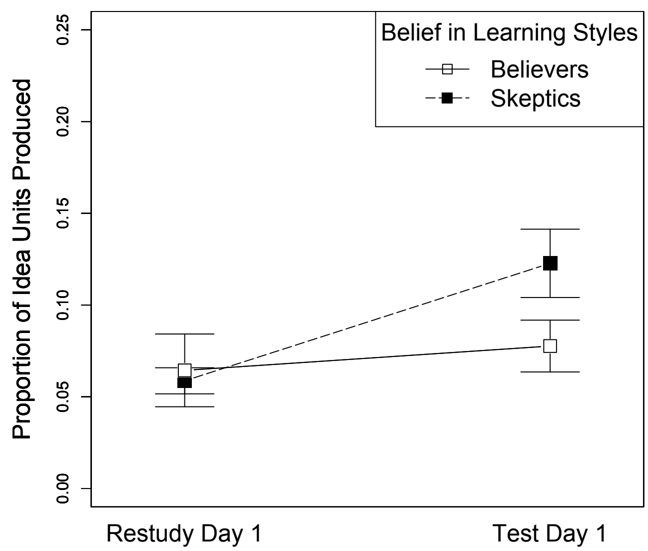

Figure 6.4. Interaction between beliefs and learning condition for the short answer task. Error bars represent standard errors.

Beliefs and the overall inclusion of critical ideas. To evaluate whether belief in Learning Styles influenced the inclusion of critical information overall (in the combined essay and question responses), we performed the same analysis. The main effects of Belief ($F(1,37) = 6.32$, $MSe = 59.63$, $p = .02$) and Learning Condition ($F(1,37) = 5.37$, $MSe = 50.14$, $p = .03$) were both significant. These main effects were qualified by a significant Belief x Learning Condition interaction ($F(1,37) = 5.31$, $MSe = 50.09$, $p =$

.03) (see Figure 6.5). Tested participants who were skeptical of learning styles included significantly more D1 ideas in their essays than did restudy participants who were skeptical ($t(21) = 3.29$, $p = .003$, *Cohen's d* = 1.43). Tested participants who believed in learning styles did not include significantly more D1 ideas in their essays than did restudy participants who also believed strongly in this educational philosophy ($t(18) = .35$, $p > .05$).

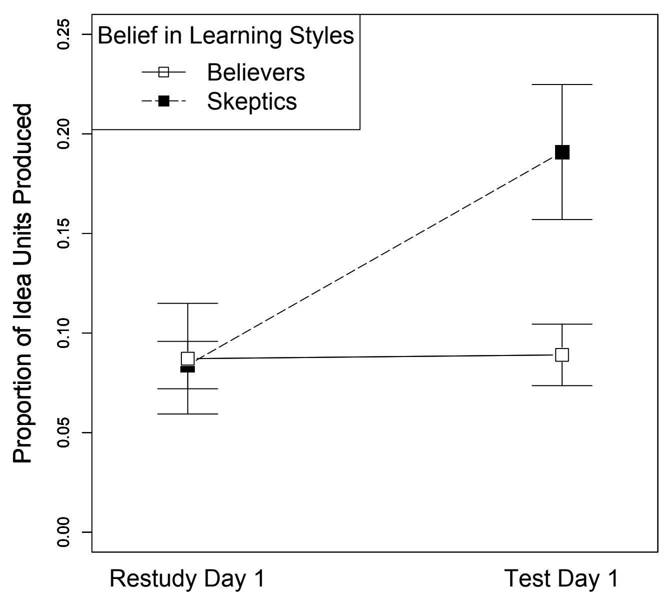

Figure 6.5. Interaction between beliefs and learning condition for the combined results (critical essay and short answer questions). Error bars represent standard errors

Discussion

The results of Experiment 3 demonstrate that testing can improve performance on two educationally relevant tasks, writing a critical essay and answering a set of short answer questions, by enhancing past learning and potentiating future learning from naturalistic expository texts. Corroborating the prediction that testing enhances past learning, participants who were tested after reading an expository text that introduced key criticisms of learning styles learned more ideas from this text than did participants who simply restudied the text. Corroborating the prediction that testing can potentiate future learning from related materials, participants who were tested on the critical text subsequently learned more ideas from a new text they read a week later that presented a contradictory argument. This enhancement of past learning and potentiation of future learning enabled participants who were initially tested to produce richer essays and short answer question responses a week after testing.

These results also demonstrate the powerful role that misconceptions play in test-potentiated learning. Participants' performance was characterized by an interaction between their belief in learning styles and whether they had been tested on Day 1. While participants who were skeptical of learning styles approaches to education integrated more ideas from the Day 1 text into their responses when they had been tested on it, participants who strongly believed in learning styles did not integrate more critical ideas from Day 1 text regardless of whether they had been tested. Since beliefs were only assessed at the end of the experiment, it is not clear whether participants brought these beliefs with them to the experiment, or if these beliefs were influenced by exposure to the experimental materials. It is also unclear if this interaction reflects a failure for believers in learning styles to benefit from testing on a text that challenged their beliefs, or if this interaction reflects a failure to use ideas that were learned but that were not congruent with prior beliefs.

These results represent a first step in translating the novel test-potentiated learning paradigm from the laboratory to the real world. They demonstrate that testing can potentiate future performance on educationally relevant tasks. However, they also demonstrate the complexity of translating a promising laboratory paradigm to real world problems, as evidenced by the interaction between participants' beliefs and testing.

Chapter 7: General Discussion

This dissertation proposed that testing can be used to produce generative knowledge bases that support future comprehension, learning, and conceptual change. This proposal was supported by the results of three experiments using naturalistic expository texts on topics in psychological science. Experiments 1, 2, and 3 replicated the classic testing effect, showing superior delayed recall of tested versus restudied information. Experiment 2 demonstrated for the first time that testing can be used to potentiate the learning of new information encountered after testing. Experiment 3 demonstrated the potential educational utility of this paradigm, showing that testing can also potentiate performance on educationally meaningful tasks like writing a critical essay.

These findings are important for empirical, theoretical, and pragmatic reasons. Empirically, they extend previous demonstrations that testing on *old* information can potentiate transfer to *new* domains (Butler, 2010) and *new* tests (Carpenter & Delosh, 1989; Glover, 1989; Kang et al. 2007, McDaniel et al., 2007). They also extend demonstrations that testing can potentiate the learning of *old* information that has been previously tested, but not successfully produced (Congleton & Rajaram, 2011; Izawa, 1966, 1967). Specifically, these results show that testing can also potentiate the learning of *new* information encountered for the first time after testing. Theoretically, these experiments provide evidence for the proposal that testing establishes generative knowledge bases, which in turn potentiate future learning. These knowledge bases support successful comprehension because they are efficiently accessed. Although we have proposed that generativity derives from simple mechanisms such as fluency and similarity, generative knowledge bases are not limited to supporting learning from conceptually similar materials. In fact, generative knowledge bases effectively support learning from materials that refute prior knowledge. Pragmatically, these experiments demonstrate that testing can potentiate learning from rich, naturalistic texts, and not just the paired-associates and word lists used in most previous studies. They also demonstrate that testing can potentiate learning and performance in the context of educationally relevant tasks. In doing so, they establish the viability of testing for improving learning in educational contexts, including notoriously difficult forms of learning such as transfer and conceptual change.

Mechanisms of Test Potentiation

We have proposed that testing potentiates future learning by building a knowledge base that is generative, facilitating comprehension and learning from future texts. One mechanism by which testing potentiates future learning is by increasing the fluency with which tested information is retrieved from LTM and integrated with new information. There is some evidence from research with paired associates (e.g., Allen et al., 1969) that testing can increase the speed with which tested content is subsequently recalled. However, it is unclear whether this finding generalizes to the retrieval and integration of textual ideas during online comprehension. The results of Experiments 2 and 3 are congruent with this proposal, but they do not provide direct evidence for fluency as a necessary mechanism for generativity. Future research should employ online measures to more directly measure the contribution of fluent retrieval to future learning. Additionally, there are multiple ways in which testing may increase fluency as well as multiple ways in which fluent retrieval may potentiate learning.

In addition to supporting the fluent integration of tested information, testing may also support future learning in an indirect fashion, by freeing up WM memory resources during the processing of new texts (Ericsson & Kintsch 1995). This possibility is supported by previous research that has demonstrated that possessing relevant prior knowledge can reduce the information processing demands posed by a new text. For example, Fincher-Kiefer, Post, Greene, and Voss (1988) compared resource availability in readers who had high versus low prior knowledge about baseball as they read sentences that either described actions from a baseball game or described neutral topics (e.g., a typical workplace scenario) about which all participants had equivalent prior knowledge. Consistent with the prediction that prior knowledge frees up processing resources during comprehension, high-knowledge participants demonstrated higher working memory capacity when reading sentences about baseball than sentences about neutral topics, whereas low-knowledge participants had lower working memory capacity when reading sentences about baseball than sentences about everyday topics. In the context of test-potentiated learning, prior testing on a related text may reduce the processing demands posed by a new text, thereby enabling participants to learn more from it.

We have argued that the fluency of retrieval emerges from the similarity of featurally rich memory traces established through testing and new textual content, but fluency may also emerge from other mechanisms. One candidate mechanism is the superior organization of tested knowledge. A considerable body of research with word lists has demonstrated that testing produces more organized knowledge relative to restudy (Bregman & Weiner, 1970; Congleton & Rajaram, 2011; Rosner, 1970; Zaromb & Roediger, 2010). If this generalizes to learning from texts, then the improved organization of knowledge built by testing may increase the fluency of retrieval through mechanisms like those spelled out in Ericsson and Kintsch's (1995) Long Term Working Memory Theory (LT-WM). According to this theory, readers can dramatically expand the information that they have ready access to by keeping pointers to well-organized LTM structures in short term working memory. The potential role of organization in promoting the fluent access necessary for test-potentiated learning warrants further investigation.

Similarity and Transfer

The potentiation of future learning demonstrated in Experiments 2 and 3 represent cases of successful transfer of learning. This expands on prior research that has demonstrated that testing can increase participants' transfer from one initially tested domain to a new one (Butler, 2010) and from one initially tested format to a new one (e.g. Kang et al., 2007; McDaniel et al., 2007). Experiments 2 and 3 demonstrate that testing can facilitate a third form of transfer, namely the learning of new information. As in all cases of transfer, it will be important to determine how "far" these benefits of testing transfer (Barnett & Ceci, 2002).

The featural similarity of tested and new content is likely to be an important factor in determining how far testing transfers. In Experiment 2, new (Day 8) texts were carefully constructed to build upon ideas from the old (Day 1) texts. This resulted in the use of similar wording across texts. For example, the Day 1 text on discovery learning contained the sentence “When participants are allowed to discover scientific principles, they are conducting experiments just like a real scientist.” and the complementary Day 8 text contained the sentence “However, most students are not prepared to conduct experiments like a real scientist”. These sentences are similar in both their surface wording (lexical similarity) and their underlying ideas (semantic similarity). Because incoming text automatically evokes the retrieval of semantically and lexically similar memory traces through passive retrieval processes (Albrecht & O’Brien, 1993; Cook et al., 1998; Myers & O’Brien, 1998; O’Brien et al., 1998), the featural similarity of the Day 8 texts likely played a key role in promoting the reactivation of memory traces built through testing on the related Day 1 text. This automatic reactivation may have promoted the fluent integration of ideas into the participants’ emergent text representations.

Because automatic memory processes do not tax a reader’s limited processing resources, reactivation driven by featural similarity may have also helped to minimize the processing needed for previously encountered ideas, thereby increasing the amount of new information that participants were able to learn from the new text. If this is true, then the distance over which testing potentiates learning may be dependent upon the similarity of new materials to old, tested materials. An interesting direction for future research will be to determine how similar new materials must be to originally tested materials for testing to potentiate learning.

Refutational Texts and Conceptual Change

Although new ideas sometimes build directly upon prior knowledge, they often require a fundamental reconceptualization of old knowledge or beliefs (Carey, 1985; Posner et al., 1982; Rumelhart & Norman, 1978; Vosniadou, 1994). Unfortunately, people often have difficulty learning new ideas that counter previously held beliefs. One established way to increase the learning of contradictory information is to explicitly refute previously held ideas (Guzzetti et al., 1993). The results of Experiment 2 suggest that testing can help to amplify the effects of explicit refutation, at least with respect to the acquisition of knowledge. Testing on a text that presented erroneous ideas about discovery learning increased the amount of information that participants subsequently learned from new texts that refuted, qualified, and elaborated these ideas. However, testing on a text that presented erroneous ideas about learning styles did not produce parallel gains. This discrepancy may have been due to the stronger beliefs about the efficacy of learning styles that participants brought to the experiment and that prevented them from revising misconceptions. This interpretation is supported by the results of Experiment 3 that showed that although participants who were skeptical of learning styles benefited from testing on ideas that were critical of these theories, participants who strongly believed in learning styles did not benefit from testing. Although not conclusive, testing also potentiated marginal changes in beliefs about discovery learning (Experiment 2) and learning styles (Experiment 3). This suggests that testing may be able to amplify attitudinal change as well as the reorganization of erroneous knowledge structures.

One explanation for why testing can precipitate conceptual change is that testing increases the fluency with which old information is retrieved, increasing the accessibility of conflicting information during the processing of a refutational text. Several researchers have posited that refutational text structures are particularly effective in promoting conceptual change because explicit refutation highlights the conflict between new information and previously held beliefs. This results in the coactivation of conflicting information (van den Broek & Kendeou, 2008) which promotes cognitive conflict and conceptual change (Guzzetti et al., 1993). Because testing increases the accessibility of previously tested ideas, it further increases the likelihood that new refutational texts will evoke the coactivation of novel ideas and conflicting information.

Educational Implications

The results of these three experiments represent a significant expansion of our understanding of the role of testing in learning. They build upon upon the growing recognition of testing as more than a tool for assessing learning, and demonstrate that testing can be used to both enhance past learning and potentiate future learning. Critically for potential educational applications, these results demonstrate that testing can amplify conceptual change and enhance/potentiate learning from the types of rich materials that students are likely to encounter in the real world.

The finding that testing enhances and potentiates learning from expository texts is particularly important for future educational applications given the centrality of expository texts to education and the difficulty many students experience comprehending and learning from expository sources. Two key reasons that readers struggle to comprehend expository texts are because they lack requisite prior knowledge (Chiesi et al., 1979; McNamara et al., 1996) and because the architecture of the human mind severely limits how effectively readers can integrate the knowledge they do have with textual information (Fletcher & Bloom, 1988; Kintsch & van Dijk, 1978; Thibadeau et al., 1982). Testing helps to address both of these shortcomings by helping readers to build a generative knowledge base that can be fluently integrated into their emerging representations of new texts read in the future. Critically, testing does not simply promote the integration of tested knowledge "as is" but can even potentiate the revision of misconceived knowledge. This result is particularly important given the need to combat non-scientific misconceptions and to improve scientific literacy.

Test-potentiated learning is likely to promote learning form a broad array of materials. Although the most direct application of this research is to improve learning from expository texts, test-potentiated learning is not limited to improving learning from expository or even textual materials. As currently theorized, the mechanisms of test-potentiated learning should be able to promote learning from any content that is featurally similar to tested content. Clearly, this suggests that test-potentiated learning should extend to other forms of text and discourse, such as narrative texts or classroom lectures. It also suggests that test-potentiation could help with learning any materials characterized by featural similarity. For example, testing could potentiate word learning in morphophonemic orthographies like Chinese by establishing memory traces which are subsequently evoked by new, featurally similar characters containing the same phonetic or semantic radicals. Future work should investigate whether test-potentiation generalizes to other expository text genres, to narrative texts, and to non-textual stimuli such as scientific visualizations (Hegarty, 2004; Johnson & Mayer, 2009). This work should focus not only on whether testing can potentiate learning from these different types of materials, but also on finding ways to maximize this potentiation.

Although these results significantly expand the emergent view of testing as a powerful tool for promoting learning, considerable work needs to be done before these results can be translated into educational practice. There are significant differences between the type of learning that occurs in the context of psychological experimentation and the type of learning that occurs in real-world learning environments. Additionally, psychological experimentation allows a level of control that is not always possible in educational settings. It will be important to replicate these findings in the real-world before proposing any changes to educational practice.

Chapter 8: Future Directions

There are two key directions for this research in the future. The first direction is to further articulate the mechanisms that are at the heart of test-potentiated learning. The central assumption of the proposed model is that testing improves future learning by increasing the fluency with which tested knowledge is subsequently retrieved, but there is more than one way in which testing can increase subsequent retrieval fluency. Determining how testing increases fluency (e.g. by altering the featural content and/or organization of memory traces) has profound implications not only for future inquiry into test-potentiation but also for our understanding of human memory and language comprehension. The second direction for this research is to begin to leverage these findings to improve real world educational outcomes. Ideally, this research will not only demonstrate the utility of test-potentiation in improving real world learning, but will also help to further articulate the mechanisms of test-potentiation and begin to have a measurable impact on the learning of actual students, especially the cultivation of scientific literacy.

Articulating the Mechanisms of Test-Potentiation

According to the proposed model, testing potentiates future comprehension by increasing the fluency with which tested content is retrieved in the future. This fluent retrieval potentiates comprehension directly by increasing the likelihood that tested information will be integrated with new textual content and/or indirectly by freeing up WM memory resources during the processing of new texts. We have proposed two mechanisms that may increase subsequent retrieval fluency. According to the first proposed mechanism, testing establishes featurally-rich memory traces that are subsequently evoked by related texts through automatic similarity-based LTM retrieval processes. According to the second proposed mechanism, testing increases the organization of knowledge, increasing fluency by expanding the amount of information that readers can maintain in WM. These mechanisms are not necessarily mutually exclusive, and future research should investigate both.

Featural similarity. A core assumption of the proposed model of test-potentiated learning as well as many modern models of text comprehension (e.g., McKoon et al., 1996; Myers & O'Brien, 1998) and human memory (e.g. Hintzman, 1988; Murdock, 1983; Ratcliff, 1978) is that passive retrieval processes evoke the activation of featurally similar memory traces. Unfortunately, "featural similarity" has rarely been explicitly defined or operationalized with respect to text comprehension (for an exception, see Lewis & Mensink, 2012). Additionally, the extant empirical evidence remains equivocal with respect to which features evoke memory traces. For example, in an influential series of studies, target sentences like *Mary ordered a cheeseburger and fries* were found to evoke the automatic retrieval of information from source sentences like *she refused to eat anything fried or cooked in grease* (Albrecht & O'Brien, 1993; Myers, O'Brien, Albrecht, & Mason, 1994; O'Brien et al., 1998). The featural similarity of these sentences could be interpreted as shared referential (Mary), thematic (the agent), and/or semantic features (*cheeseburger* and *eat/fried/grease*). In addition, other studies have implicated both phonological similarity (Lea et al., 2008) and lexical similarity (Lewis &

Mensink, 2012) as additional featural dimensions that may be relevant to online text comprehension and learning.

This current theoretical and empirical equivocation means that it will be necessary to not only demonstrate that testing produces featurally-rich memory traces, but to actually articulate which features are relevant to fluent text comprehension. One way to do this is to systematically manipulate the surface and semantic features of the texts that participants encounter after testing and to directly measure the impact this featural manipulation has on the retrieval of tested information. During an initial learning session, participants would read a text that contains multiple source sentences. These source sentences would contain at least one key source idea that would not be repeated in future texts. After a delay, participants would read a new text that contains target sentences which systematically vary in the their semantic, lexical, phonological, referential, and thematic similarity to source sentences, but do not repeat the key source ideas. Immediately after reading these sentences, the retrieval of a source idea would be assessed by having participants name or make a lexical decision on a word related to the source idea. The featural richness of memory traces would be reflected by faster naming and lexical decision responses following a wide array of target sentences. The relevance of different types of featural similarity would be implicated by variation in the efficacy of different target sentences in reducing response latencies. For example, semantically similar cues may be more effective than lexically similar cues in evoking the retrieval of memory traces.

Investigating the question of featural similarity will also help to answer key questions with respect to the application of test-potentiation to real world learning problems. Chief among these is the question of how "far" the benefits of testing transfer (cf., Barnett & Ceci, 2002). Determining how similar tested and new materials must be is a first step in deciding when testing should be inserted into real-world curricula in order to facilitate the transfer of currently studied content to content that will be presented in the future. For example, this research may suggest that it is beneficial to break textbook chapters into small sections with tests after each one. That way, testing can *enhance* the learning of content from early in a chapter and *potentiate* the learning of content later in a chapter.

Organization. Future research should also investigate testing's effect on the organization of knowledge built from expository texts as well as the ways in which this organization can facilitate subsequent comprehension. Previous research has demonstrated that testing can increase the subjective and/or taxonomical organization of participants' subsequent retrieval of word lists (Bregman & Weiner, 1970; Congleton & Rajaram, 2011; Rosner, 1970; Zaromb & Roediger, 2010), but there is no relevant evidence regarding the organization of textual information. Because organization may be play an important role in promoting the fluency of readers' access to tested information, it will also be important to determine whether any increases in organization are accompanied by increases in accessibility. The traditional approach to measuring testing's effect on organization has been to calculate measures of clustering in participants' recalls, such as ARC scores (Roenker et al., 1971). Using clustering metrics to analyze recalls from multi-topic texts (as opposed to the single topic texts used in this experiment) will be a good first step in investigating testing's effect on the organization of text representations. However, clustering metrics will not help to uncover the role that this organization may play in fluent retrieval.

To investigate the role of organization in fluent retrieval, it will be important to use paradigms that can probe both structure and retrieval latency. One such approach is the priming paradigm created by Ratcliff and McKoon (1978) (see also Lewis & Varma, 2011; Zwaan, 1996). In this paradigm, participants make judgments on lists of statements derived from earlier texts while their reaction times are recorded. Within these lists, target statements are preceded by different prime statements. If judging a prime statement speeds participants' target judgments, the information represented by the prime and target statements are assumed to be connected in LTM. In the context of test-potentiated learning, priming paradigms will help to determine whether testing increases the organization of text representations as well as whether this organization speeds the retrieval of connected information.

Establishing and Maximizing Educational Utility

There is a growing recognition that testing is more than a tool for assessing learning. The consensus that testing can solidify students' memory for previously encountered information has lead the Institute of Education Sciences to identify testing as one of only two instructional practices with a "strong" level of empirical support (Pashler et al., 2007). This dissertation suggests an even more radical reconceptualization of the role of testing in education, suggesting that testing can be integral not only to solidifying memory for previously encountered information, but also to building generative knowledge that will support future learning. Demonstrating that this model generalizes to more naturalistic learning environments will require pervasive real world experimentation. If conducted properly, this experimentation will not only establish the educational utility of test-potentiated learning, but will also help to tease apart the underlying mechanisms of test-potentiation by subjecting the model to testing with rich naturalistic materials and students who vary with respect to important individual differences.

One approach that has been used to demonstrate the educational utility of testing has been to investigate testing's impact on learning in "simulated classrooms" (e.g., Butler & Roediger, 2007). In these simulated classrooms, students learn short lessons, take tests designed to enhance learning, and complete summative assessments of their knowledge. The primary benefit of simulated classrooms is that they allow a level of control that is not possible in the real world. The drawbacks of simulated classrooms are that they test a limited range of materials and students, have no direct impact on real-world learning outcomes, and sacrifice many of the online measures (like reading times and probe response latencies) that are critical to measuring real-time cognition. Therefore, we argue for a more pervasive, more technologically sophisticated approach to studying test potentiation in educational contexts.

One particularly intriguing platform for investigating test-potentiation in the real world is digital textbooks. By their very nature, digital texts are conducive to the systematic manipulation of text characteristics (e.g. lexical, semantic, and phonological features) and test features (e.g. test format and spacing) that is necessary for teasing apart the mechanisms of test-potentiation and for identifying best practices. Additionally, digital texts can be easily distributed to a wide array of students at multiple sites increasing the ecological validity of results and enabling an investigation of important individual characteristics such as prior knowledge, beliefs, and reading ability. As results are produced, they can be automatically forwarded to researchers, enabling the rapid analysis of data and a "design-build" approach to text/test construction. As the latencies for digital reading devices such as tablet computers improve, researchers will be able to collect online data which will enable a more direct assessment of students' cognition than test results alone. For example, researchers will be able to investigate the fluency with which readers integrate tested and new information by analyzing reading times for new content. As the cameras on these devices improve, it will even be possible to collect usable eye-movement data that can help to further elucidate the impact of previous testing on online comprehension.

The superordinate goal for this research should be to investigate how testing can be employed to promote the type of nuanced scientific literacy that few students achieve. Scientific literacy requires a continuous revision of old knowledge and beliefs. The proposed model of test potentiated learning provides a new perspective on how we can help students learn new ideas that expand, qualify, and refute old knowledge. However, there is considerable work that needs to be done in order to maximize the utility of testing in promoting conceptual change and scientific literacy. For example, it will be important to further explore the interaction of beliefs and testing, and to identify methods for maximizing the conceptual change of learners who hold strong, erroneous beliefs.

References

Abbot, E. E. (1909). On the analysis of the memory consciousness in orthography. *The Psychological Review: Monograph Supplements*, *11*(1), 127-158.

Adams, M., & Bruce, B. (1980). Background knowledge and reading comprehension. *Reading Education Report No. 13.* University of Illinois.

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*(7), 861-876.

Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(5), 1061–1070.

Allen, G. A., Mahler, W. A., & Estes, W. K. (1969). Effects of recall tests on long-term retention of paired associates. *Journal of Verbal Learning and Verbal Behavior*, *8*(4), 463-470.

Alvermann, D. E., Smith, L. C., & Readence, J. E. (1985). Prior knowledge activation and the comprehension of compatible and incompatible text. *Reading Research Quarterly*, *20*(4), 420–436.

Anderson, R. C., & Biddle, W. B. (1975). On asking people questions about what they are reading. *Psychology of Learning and Motivation*, *Volume 9*, 89-132.

Anderson, J. R., Budiu, R., & Reder, L. M. (2001). A theory of sentence memory as part of a general theory of memory. *Journal of Memory and Language*, *45*(3), 337-367.

Ausubel, D. P. (1960). The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of Educational Psychology*, *51*(5), 267–272.

Barnes, B. R., & Clawson, E. U. (1975). Do advance organizers facilitate learning? Recommendations for further research based on an analysis of 32 studies. *Review of Educational Research*, *45*(4), 637–659.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612–637.

Bartlett, F. C. (1932). *Remembering: a Study in Experimental and Social Psychology.* Oxford, England: Macmillan.

Bartlett, J. C. (1977). Effects of immediate testing on delayed retrieval: Search and recovery operations with four types of cue. *Journal of Experimental Psychology: Human Learning and Memory*, *3*(6), 719-732.

Bartlett, J. C., & Tulving, E. (1974). Effects of temporal and semantic encoding in immediate recall upon subsequent retrieval. *Journal of Verbal Learning and Verbal Behavior*, *13*(3), 297-309.

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. *Information processing and cognition: The Loyola symposium* (pp. 123-144). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology*, *6*(2), 173-189.

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, *11*(2), 177-220.

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, *3*(2), 193-209.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 717-726.

Bregman, A. S., & Wiener, J. R. (1970). Effects of test trials in paired-associate and free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, *9*(6), 689-698.

Britton, B. K., Graesser, A. C., Glynn, S. M., Hamilton, T., & Penland, M. (1983). Use of cognitive capacity in reading: Effects of some content features of text. *Discourse Processes*, *6*(1), 39-57.

Broughton, S. H., Sinatra, G. M., & Reynolds, R. E. (2010). The nature of the refutation text effect: An investigation of attention allocation. *The Journal of Educational Research*, *103*(6), 407–423.

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1118-1133.

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

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116–131.

Carpenter, S., & Delosh, E. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, *34*(2), 268-276.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, *20*(6), 633-642.

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

Chase, W. G, & Ericsson, K. A. (1982). Skill and working memory. *The Psychology of Learning and Motivation: Advances in Research and Theory*, *16*, 1-58.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55-81.

Chiesi, H. L., Spilich, G. J., & Voss, J. F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, *18*(3), 257-273.

Congleton, A. R., & Rajaram, S. (2011). The influence of learning methods on collaboration: Prior repeated retrieval enhances retrieval organization, abolishes collaborative inhibition, and promotes post-collaborative memory. *Journal of Experimental Psychology: General*, *140*(4), 535-551.

Cook, A. E., Halleran, J. G., & O'Brien, E. J. (1998). What is readily available during reading? A memory-based view of text processing. *Discourse Processes*, *26*(2), 109 - 129.

Corkill, A. J. (1992). Advance organizers: Facilitators of recall. *Educational Psychology Review*, *4*(1), 33–67.

Darley, C. F., & Murdock, B. B. (1971). Effects of prior free recall testing on final recall and recognition. *Journal of Experimental Psychology*, *91*(1), 66-73.

Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E.L. Bjork & R.A. Bjork (Eds,), *Memory* (pp. 317-344). San Diego, CA, US: Academic Press.

Diakidoy, I.-A. N., & Kendeou, P. (2001). Facilitating conceptual change in astronomy: a comparison of the effectiveness of two instructional approaches. *Learning and Instruction*, *11*(1), 1–20.

Dole, J. A., & Sinatra, G. M. (1998). Reconceptualizing change in the cognitive construction of knowledge. *Educational Psychologist*, *33*(2-3), 109–128.

Dooling, D. J., & Lachman, R. (1971). Effects of comprehension on retention of prose. *Journal of Experimental Psychology*, *88*(2), 216-222.

Eimas, P. D., & Zeaman, D. (1963). Response speed changes in an Estes' paired-associate "miniature" experiment. *Journal of Verbal Learning and Verbal Behavior*, *1*(5), 384–388.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211-245.

Estes, W. K. (1960). Learning theory and the new "mental chemistry." *Psychological Review, 67*(4), 207-223.

Faw, H. W., & Waller, T. G. (1976). Mathemagenic behaviours and efficiency in learning from prose materials: Review, critique and recommendations. *Review of Educational Research*, *46*(4), 691–720.

Fincher-Kiefer, R., Post, T. A., Greene, T. R., & Voss, J. F. (1988). On the role of prior knowledge and task demands in the processing of text. *Journal of Memory and Language*, *27*(4), 416–428.

Fletcher, C. R. (1981). Short-term memory processes in text comprehension. *Journal of Verbal Learning and Verbal Behavior*, *20*(5), 564-574.

Fletcher, C. R. (1986). Strategies for the allocation of short-term memory during comprehension. *Journal of Memory and Language*, *25*(1), 43-58.

Fletcher, C. R., & Bloom, C. P. (1988). Causal reasoning in the comprehension of simple narrative texts. *Journal of Memory and Language*, *27*(3), 235-244.

Gardiner, F., Craik, F., & Bleasdale, F. (1973). Retrieval difficulty and subsequent recall. *Memory & Cognition*, *1*(3), 213-216.

Gates, A. I. (1917). Recitation as a factor in memorizing. New York: Science Press.

Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, *81*(3), 392-399.

Gotz, A., & Jacoby, L. L. (1974). Encoding and retrieval processes in long-term retention. *Journal of Experimental Psycholog, 102*(2), 291-297.

Gregoire, M. (2003). Is it a challenge or a threat? A dual-process model of teachers' cognition and appraisal processes during conceptual change. *Educational Psychology Review*, *15*(2), 147–179.

Guzzetti, B. J., Snyder, T. E., Glass, G. V., & Gamas, W .S. (1993). Promoting conceptual change in science: A comparative meta-analysis of instructional interventions from reading education and science education. *Reading Research Quarterly* , 28(2), 116-159.

Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research*, *56*(2), 212-242.

Hartley, J., & Davies, I. K., Preinstructional strategies: The role of pretests, behavioral objectives, overviews and advance organizers. *Review of Educational Research*, *46*(2), 239-265.

Hegarty, M. (2004). Dynamic visualizations and learning: getting to the difficult questions. *Dynamic Visualisations and Learning*, *14*(3), 343–351.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*(4), 528-551.

Hogan, R. M., & Kintsch, Walter. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, *10*(5), 562-567.

Izawa, C. (1966). Reinforcement-test sequences in paired-associate learning. *Psychological Reports*, *18*(3), 879-919.

Izawa, C. (1967). Function of test trials in paired-associate learning. *Journal of Experimental Psychology, 75*(2), 194-209.

Izawa, C. (1970). Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. *Journal of Experimental Psychology*, *83*(2, Pt.1), 340-344.

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, *17*(6), 649–667.

Johnson, C. I., & Mayer, R. E. (2009). A testing effect with multimedia learning. *Journal of Educational Psychology*, *101*(3), 621–629.

Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, inference and consciousness*. Cambridge: Cambridge University Press.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122-149.

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

Kardash, C. M., & Scholes, R. J. (1996). Effects of preexisiting beliefs, epistemological beliefs, and need for cognition on interpretation of controversial issues. *Journal of Educational Psychology*, *88*(2), 260–271.

Kendeou, P., & Van den Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & Cognition*, *35*(7), 1567–1577.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*(2), 163-182.

Kintsch, W. (1998). *Comprehension: A Paradigm for Cognition.* New York, NY, US: Cambridge University Press.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363-394.

Kintsch, W., & Young, S. (1984). Selective recall of decision-relevant information from texts. *Memory & Cognition*, *12*(2), 112-117.

Kuo, T.-M., & Hirshman, E. (1996). Investigations of the testing effect. *The American Journal of Psychology*, *109*(3), 451-464.

Landauer, T. K., & Eldridge, L. (1967). Effects of tests without feedback and presentation-test interval in paired-associate learning. *Journal of Experimental Psychology*, *75*(3), 290-298.

Lea, R. B., Rapp, D. N., Elfenbein, A., Mitchel, A. D., & Romine, R. S. (2008). Sweet silent thought. *Psychological Science*, *19*(7), 709–716.

Lewis, M. R., & Mensink, M. C. (2012). Prereading questions and online text comprehension. *Discourse Processes*, *49*(5), 367–390.

Lewis, M.R., & Varma, S. (2011). *Episodic structure, episodic memory, and text comprehension.* Poster presented at the annual meeting of the Society for Text and Discourse, Poitiers, France.

Mayer, R. E. (1979). Twenty years of research on advance organizers: Assimilation theory is still the best predictor of results. *Instructional Science*, *8*(2), 133–167.

Mayer, R. E. (1983). Can you repeat that? Qualitative effects of repetition and advance organizers on learning from science prose. *Journal of Educational Psychology*, *75*(1), 40-49.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist*, *59*(1), 14–19.

McDaniel, M. A., & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology*, *16*(2), 192-201.

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

McDaniel, M. A., Einstein, G. O., Dunay, P. K., & Cobb, R. E. (1986). Encoding difficulty and memory: Toward a unifying theory. *Journal of Memory and Language*, *25*(6), 645-656.

McDaniel, M. A., & Kerwin, M. L. E. (1987). Long-term prose retention: Is an organizational schema sufficient? *Discourse Processes*, *10*(3), 237-252.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, *99*(3), 440-466.

McKoon, G., Gerrig, R. J., & Greene, S. B. (1996). Pronoun resolution without pronouns: Some consequences of memory-based text processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(4), 919–932.

McNamara, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *55*(1), 51-62.

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

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*(3), 247-288.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 519–533.

Mullis, V. S., Martin, M. O., Kennedy, A. M., & Foy, P. (2007). *PIRLS 2006 International Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Murdock, B. B. (1983). A distributed memory model for serial-order information. *Psychological Review*, *90*(4), 316-338.

Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, *26*(2), 131 – 157.

Myers, J. L., O'Brien, E. J., Albrecht, J. E., & Mason, R. A. (1994). Maintaining global coherence during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 876–886.

Nungester, R. J., & Duchastel, P. C. (1982). Testing versus review: Effects on retention. *Journal of Educational Psychology*, *74*(1), 18-22.

O'Brien, E. J., Rizzella, M. L., Albrecht, J. E., & Halleran, J. G. (1998). Updating a situation model: A memory-based text processing view. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(5), 1200–1210.

OECD. (2010). *PISA 2009 Results: What Students Know and Can Do: Student Performance in Reading, Mathematics and Science (Volume I)*.

Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007) *Organizing Instruction and Study to Improve Student Learning (NCER 2007-2004).* Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles. *Psychological Science in the Public Interest*, *9*(3), 105-119.

Pintrich, P. R., Marx, R. W., & Boyle, R. A. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research*, *63*(2), 167–199.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, *66*(2), 211–227.

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: mediator effectiveness hypothesis. *Science*, *330*(6002), 335.

RAND Reading Study Group. (2002). *Reading for Understanding: Toward an R&D Program in Reading Comprehension*. Santa Monica, CA: Rand Corp.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59-108.

Ratcliff, R., & McKoon, G. (1978). Priming in item recognition: Evidence for the propositional structure of sentences. *Journal of Verbal Learning and Verbal Behavior*, *17*(4), 403–417.

Rawson, K., & Kintsch, W. (2002). How does background information improve memory for text content? *Memory & Cognition*, *30*(5), 768-778.

Rawson, K. A., & Kintsch, W. (2004). Exploring encoding and retrieval effects of background information on text memory. *Discourse Processes*, *38*(3), 323-344.

Recht, D. R., & Leslie, L. (1988). Effect of prior knowledge on good and poor readers' memory of text. *Journal of Educational Psychology*, *80*(1), 16-20.

Robert S., L. (1975). The facilitation of recognition by recall. *Journal of Verbal Learning and Verbal Behavior*, *14*(3), 253-258.

Roediger III, H. L., & Blaxton, T. A. (1987). Retrieval modes produce dissociations in memory for surface information. In D. S. Gorfein R. R. Hoffman (Ed.), *Memory and learning: The Ebbinghaus Centennial Conference* (pp. 349–379). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Roediger, H. L., & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*(3), 181-210.

Roediger, H. L., & Karpicke, J. D. (2006b). Test-enhanced learning. *Psychological Science*, *17*(3), 249-255.

Roenker, D. L., Thompson, C. P., & Brown, S. C. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin*, *76*(1), 45–48.

Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 233–239.

Rosner, S.R. (1970). The effects of presentation and recall trials on organization in multitrial free recall. *Journal of Verbal Learning and Verbal Behavior*, *9*(1), 69–74.

Rothkopf, E. Z., & Billington, M. J. (1975). Relevance and similarity of text elements to descriptions of learning goals. *Journal of Educational Psychology*, *67*(6), 745-750.

Rumelhart, D. & Norman, D. (1978). Accretion, tuning and restructuring: Three modes of learning. In. J.W. Cotton & R. Klatzky (eds.), *Semantic Factors in Cognition*. Hillsdale, NJ: Erlbaum.

Runquist, W. (1983). Some effects of remembering on forgetting. *Memory & Cognition*, *11*(6), 641-650.

Schneider, W., & Körkel, J. (1989). The knowledge base and text recall: Evidence from a short-term longitudinal study. *Contemporary Educational Psychology*, *14*(4), 382-393.

Schneider, W., Körkel, J., & Weinert, F. E. (1989). Domain-specific knowledge and memory performance: A comparison of high- and low-aptitude children. *Journal of Educational Psychology*, *81*(3), 306-312.

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, *30*(9), 641-656.

Steffensen, M. S., Joag-Dev, C., & Anderson, R. C. (1979). A cross-cultural perspective on reading comprehension. *Reading Research Quarterly*, *15*(1), 10–29.

Sternberg, R. J., & Tulving, E. (1977). The measurement of subjective organization in free recall. *Psychological Bulletin*, *84*(3), 539-556.

Sue R., R. (1970). The effects of presentation and recall trials on organization in multitrial free recall. *Journal of Verbal Learning and Verbal Behavior*, *9*(1), 69-74.

Szpunar, K. K., McDermott, K. B., & Roediger III, H. L. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1392-1399.

Thibadeau, R., Just, Marcel Adam, & Carpenter, P. A. (1982). A model of the time course and content of reading. *Cognitive Science*, *6*(2), 157-203.

Thompson, C. P., Wenger, S. K., & Bartling, C. A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(3), 210-221.

Trabasso, T., Secco, T., & van den Broek, P. (1984). Causal cohesion and story coherence. In H. Mandl, N. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 83-111). Lawrence Erlbaum Associates.

van den Broek, P., & Kendeou, P. (2008). Cognitive processes in comprehension of science texts: the role of co-activation in confronting misconceptions. *Applied Cognitive Psychology*, *22*(3), 335–351.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. New York: Academic Press.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction*, *4*(1), 45–69.

Walker, C. H. (1987). Relative importance of domain knowledge and overall aptitude on acquisition of domain-related information. *Cognition and Instruction*, *4*(1), 25-42.

Walker, C. H., & Yekovich, F. R. (1987). Activation and use of script-based antecedents in anaphoric reference. *Journal of Memory and Language*, *26*(6), 673-691.

Wenger, S. K., Thompson, C. P., & Bartling, C. A. (1980). Recall facilitates subsequent recognition. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 135-144.

Wiley, J., & Rayner, K. (2000). Effects of titles on the processing of text and lexically ambiguous words: Evidence from eye movements. *Memory & Cognition*, *28*(6), 1011-1021.

Zaromb, F., & Roediger, H. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition*, *38*(8), 995-1008.

Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1196–1207.

Appendix A

Day 1 Pro Learning Styles Text (Experiments 1 and 2)

It is clear that people think and learn in different ways. Like most people, you have probably had a teacher who just did not work for you. This was probably due to the fact that the teacher's instructional style did not match your learning style. More and more educators are recognizing that some people learn visually, others learn verbally, and still others learn kinesthetically. Students learn best when instruction matches their particular style. For example, a visual learner learns best with visual instruction and a verbal learner learns best with verbal instruction.

One way to address individuals' learning styles is to teach a topic in multiple ways. There is good evidence that this type of broad instruction is more effective than the narrower type of instruction that teachers have traditionally used. Broad instruction helps everyone and works because teaching a topic in multiple ways simultaneously targets the different learning styles of different students.

Although good teaching is definitely characterized by teaching the same concept in multiple ways; if we want to truly maximize student learning, we need to start tailoring instruction directly to the individual styles of students. There is strong evidence that matching instructional format to a student's style maximizes learning. Additionally, when students receive instruction that matches their style, they are much happier and more engaged in the lesson.

Every student is an individual and deserves to be treated like one. Only by tailoring teaching to every student's learning style, can we respect this individuality. To do this, we need to start assessing the learning style of every student and giving them appropriate instruction. Luckily, there are easy and reliable ways to identify students' learning styles. There are many excellent learning styles assessments and books for sale. There are even some free assessments available on the internet. These assessments accurately measure learning styles by asking people to rate their agreement with simple statements.

Once we have determined every student's learning style, we need to group students by style and give them customized instruction. This is well within the means of most school districts. In fact, many model schools across the country have already implemented educational interventions based on learning styles theories. The wide spread recognition of distinct learning styles has been one of the most important revolutions in education. If we pay attention to learning styles theories, we can dramatically improve learning.

Day 1 Pro Discovery Learning Text (Experiments 1 and 2)

Discovery learning is an exciting new idea. In discovery learning, students "learn by doing". In one particularly effective physics lesson, young students learn about basic physics principles by freely manipulating an adjustable ramp. They learn about friction by changing the surface of the ramp from rough sandpaper to smooth linoleum. They learn about acceleration by adjusting the ramp's slope and rolling a ball down it.

A student can learn more by actively conducting experiments than by listening to a lecture. A major strength of discovery learning is that students engage in the same activities as experts. When students are allowed to discover scientific principles, they are conducting experiments just like a real scientist. Like a scientist, they generate and test hypotheses. Also like a scientist, students encounter roadblocks that they must find a way around. As a result, they are learning real-world skills that could never be taught through direct instruction.

As the name "discovery learning" implies, students literally "discover" the concepts and rules of a new domain. For example, when students are asked to solve novel algebra problems with no guidance, they must explore the myriad possible routes to a solution that exist for every math problem. This might involve a considerable amount of trial and error. As a result, students develop critical problem-solving skills.

Students learn best when they construct knowledge themselves. The active learning fostered by discovery learning is the only way to guarantee that students construct their own knowledge. Additionally, discovery learning benefits all types of students. The amount of previous experience that students have with a domain does not matter.

One of the major goals of education is to promote *transfer*. Transfer is the ability to apply previously learned concepts to new problems. Discovery learning is more effective than direct instruction in promoting transfer. Because they have already encountered significant hurdles during learning, students who learn through discovery are better able to apply previously learned concepts to new problems in the future.

In successful teaching, a teacher's role should be radically altered. Teachers should not lecture or offer students explicit guidance. Instead a teacher should establish learning goals and allow students to discover relevant features of a problem on their own. To maximize learning, teachers should not intervene when students struggle. This does not reduce a teacher's involvement in education, it transforms it. Instead of preparing lectures that students will passively consume, teachers should design active learning experiences.

Day 8 Anti Learning Styles Text (Experiment 2)

The idea that people learn in different ways has grown in popularity in recent years. Most learning styles theories claim that some people learn verbally and others learn visually. The appeal of these theories is often based on anecdotal evidence like most people's experience that some teacher just did not work for them. Unfortunately, there is no strong evidence that learning is best when instructional format matches a student's putative style.

Although there are many ideas about how learning styles can improve education, the most popular idea is the meshing hypothesis. The meshing hypothesis is the idea that learning is best when the mode of instruction matches an individual's style. For example, a common claim is that a visual learner will learn best when information is presented visually. Although students are definitely happier when they receive instruction that matches their preferred style, they do not appear to learn more.

For example, consider a study by Massa and Mayer (2006). The researchers first gave students a series of learning styles assessments. Next, they randomly assigned them to an electronics lesson that emphasized either the visual or verbal presentation of ideas. Finally, they gave students a test to assess their learning. Contrary to the predictions of the meshing hypothesis, “verbal students” did not learn better with verbal instruction and “visual students” did not learn better with visual instruction.

A weaker version of the learning styles idea is the claim that good instruction is characterized by teaching concepts in multiple ways. According to these claims, broad instruction works because teachers are able to target multiple learning styles at the same time. There is some evidence that broad instruction is better than narrow instruction. However, this evidence does not mean that learning is improved because broad instruction targets multiple learning styles. Instead, broader instruction may work because the human brain has evolved to learn from multiple senses working together. Therefore, broad instruction may help everyone in a similar way.

Tailoring instruction to the learning styles of every student would be incredibly expensive. This would be expensive because students would need to be assessed, grouped by style, and then given customized instruction. Teaching to learning styles would require the creation of new teaching methods. Additionally, teachers would need to be trained in these methods, and more teachers would need to be hired to teach to every style.

Many advocates of learning styles approaches to education claim that there are easy and reliable ways to identify students’ styles. Unfortunately, most learning styles assessments do not actually measure styles. Instead, they measure students’ preferences. For example, there are many online assessments that ask people to rate their agreement with statements like “I prefer to read instructions about how to do something rather than have someone show me”. Learning preferences are very real, but they do not reflect how people actually learn best.

We all agree that students should be treated as individuals, but learning styles based education is neither the only way nor the best way to respect students’ individuality. It may be more effective to target instruction to other differences such as differences in general aptitude or personality.

Day 8 Anti Discovery Learning Text (Experiment 2)

Discovery learning is not a new idea. Similar approaches to learning have gone by many names, including inquiry learning, problem-based learning, and experiential learning. Proponents of discovery learning often claim that students "learn by doing". However, it is not always true that students learn more by conducting an experiment than by listening to a lecture. Despite the hopes of many teachers, students are unlikely to learn physics principles by simply manipulating a ramp or a set of springs with little guidance.

Many science teachers believe that a major strength of discovery learning is that students engage in the same types of activities as experts. However, most students are not prepared to conduct experiments like a real scientist. Scientists have considerable background knowledge that students do not, so the experiences of scientists and students are very different. Without guidance or instruction, students may not be able to generate and test hypotheses, at least not meaningful ones.

In many cases, students do not "discover" the rules of a domain. Discovery learning often involves a lot of trial and error, and therefore students may never successfully discover appropriate concepts. This is especially true in mathematics. Because math problems have multiple routes to a solution, students may never select the right one on their own.

Additionally, discovery learning may have emotional consequences. For example when students are asked to solve math problems with no guidance, they often get frustrated. Not every type of student benefits from discovery learning. In fact, discovery learning may only benefit students who already have fairly high background knowledge about a domain. Background knowledge and prior experience can reduce the demands that discovery learning places on cognitive resources.

In addition to not promoting initial learning, there is no strong evidence that discovery learning improves students' ability to apply learned concepts to new problems. It is often claimed that discovery learning is more effective than direct instruction in promoting transfer. However, this claim has not received empirical support. For example, Klahr and Nigam (2004) conducted an experiment in which they compared students' ability to apply knowledge to new situations. Students learned about basic scientific principles either through discovery learning or direct instruction. Students who learned through discovery did not learn more and were no better at applying basic scientific principles to new problems.

Calls for discovery learning often appeal to the educational philosophy of Constructivism. According to Constructivism, students learn best when they construct knowledge. However, the visible activity fostered by discovery learning is not the only way to guarantee that learners construct knowledge. Proponents of discovery learning often confuse visible activity with mental activity. Students rarely passively consume lectures and can learn as much or more from a lecture than from discovery.

Many proponents of discovery learning claim that teachers should not intervene when students struggle, but the research suggests that teachers can help students learn by intervening. According to new models of "guided" discovery, a teacher's role should not be radically altered. Unlike "pure" discovery, in "guided" discovery teachers help students by pointing out salient features of a problem and by providing help at key moments. Unlike pure discovery, guided discovery may be a valuable addition to every teacher's toolkit.

Appendix B

Day 1 Anti Learning Styles Text (Experiment 3)

The educational literature is filled with claims that individuals differ with respect to the way they learn. For example, many popular theories differentiate between “verbal” and “visual” learning styles. Proponents of learning styles argue that optimal learning requires assessing an individual’s learning style and tailoring instruction accordingly. However, there is no strong scientific evidence to support this claim.

One of the most popular ideas amongst proponents of learning styles is the “meshing hypothesis”. According to the meshing hypothesis, instruction is most effective when it is provided in a format that matches the learning style of an individual. For example, a visual learner is assumed to learn best when concepts are illustrated with visual-aids.

Proponents of the meshing hypothesis often appeal to anecdotal evidence to support their claims. The learning styles literature is filled with personal stories about educational successes and failures. Perhaps learning styles advocates resort to this anecdotal evidence because the results of high-quality research studies rarely support strong claims like the meshing hypothesis. Consider a study by Massa and Mayer (2006). They had students take a series of learning styles assessments and then randomly assigned them to an electronics lesson that emphasized either the visual or verbal presentation of ideas. Next, they gave students a test to assess their learning. Contrary to the predictions of learning styles theories, “verbal students” did not learn better with verbal instruction and “visual students” did not learn better with visual instruction.

A weaker version of the learning styles position contends that broad instruction is better than narrow instruction. According to these weaker claims, good instruction is characterized by teaching concepts in multiple ways. By teaching in multiple ways, a teacher is assumed to target the learning styles of many students during the same lesson. There is some limited evidence that broader instruction is associated with improved learning. However, this does not support stronger claims like the meshing hypothesis. Students may learn better when instruction is delivered in multiple ways not because it targets their individual learning styles, but because the human brain has evolved to learn through multiple senses working in unison.

An important distinction that is rarely made is the distinction between “styles” and “preferences”. Most learning style assessments do not measure how a person actually learns. Instead, they measure how a person prefers to learn or to study. Study preferences are very real, but they are not the same as learning styles. When students are allowed to choose an instructional format that matches their study preferences, they tend to report higher levels of satisfaction. However, they do not appear to learn more.

Teaching to putative learning styles is likely to be expensive. Students must first be assessed and grouped by learning style and then given some sort of customized instruction. Delivering customized instruction will require the creation and validation of instructional activities for each learning style. Partitioning children within a given classroom and teaching each subset differently might require increasing the number of teachers.

If there is no strong evidence that tailoring instruction to students' learning styles increases learning, why is the learning styles idea so popular? One reason for the popularity of learning styles are the commercial interests involved. Publishers aggressively market a broad range of books and assessments. Another reason for the popularity of learning styles is the widely held belief that all students should be treated by educators as unique individuals. However, rejecting learning styles approaches to education does not mean rejecting individuality. In fact, it may allow educators to focus on differences between students that are more important to educational outcomes such as differences in aptitude or personality.

Day 8 Pro Learning Styles Text (Experiment 3)

People learn and think in different ways. You have probably taken a course where the teacher's mode of instruction just did not work for you. Although I am now a successful research psychologist, I struggled to earn a C in my first statistics class. I put in hours of work, but I could never grasp the fundamental concepts. Everything changed when my next statistics teacher showed me a new way to look at statistics. Instead of describing the key concepts verbally, he presented them geometrically. My previous teacher had been describing everything with words and I had struggled. My new teacher presented ideas visually, and I thrived.

Luckily, a lot has changed since I was student. One of the most important revolutions has been the wide spread recognition of distinct learning styles. More and more educators are recognizing that some students learn visually, other verbally, and still other kinesthetically. By matching the presentation of information to an individual's style, an individual's learning can be maximized.

It is important to assess each student's learning style. There are reliable and easy ways to identify a student's learning style. Some of these are commercially available. Others are available for free on the internet. For example, one online assessment asks students to rate their agreement with statements like "I prefer to read instructions about how to do something rather than have someone show me". Students are much happier when they receive instruction that matches their responses on assessments like this.

There is strong evidence that aligning instructional style with a student's learning style improves learning. For example, Sternberg et al. (1998) compared the effects of instruction that targeted multiple styles of learning to more traditional instruction that only targeted one style. The middle school students they studied learned better when they received the broader type of instruction than when they received the narrower, more traditional instruction.

Assessments can help teachers to identify their current teaching style as well as to identify their students' individual learning styles. I strongly suggest that teachers consider purchasing one of the excellent books available at the web site of the International Learning Styles Network (www.learningstyles.net). These books can help teachers to tailor their teaching to the individual needs of their students.

Each student is an individual and deserves an education that recognizes it. By tailoring instruction to each student's learning style, we can maximize learning and ensure that each student receives the individualized education they deserve. We need to start assessing each student's learning style and partitioning our classrooms to ensure that each student is taught in the most appropriate way.

Day 8 Questions (Experiment 3)

1- What version or versions of the learning styles idea does this week's author support?

2-What type or types of evidence does the author appeal to? Why do you think the author appeals to this type of evidence?

3-What types of conclusions does the study the author cited support? Are there any alternative interpretations?

4- What does the online assessment the author described probably measure? Why is this important?

5- What would be the economic consequences of the author's plan? Be specific.

6-Why might this argument appeal to a lot of people?