INSTRUCTIONAL EFFECTS OF REFUTATION TEXT ON DIFFERENT TYPES OF KNOWLEDGE

A Dissertation in
Learning, Design, and Technology

by

General Mfanasibili Ntshalintshali

© 2014 General Mfanasibili Ntshalintshali

Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

May 2014
The dissertation of General Mfanisibili Ntshalintshali was reviewed and approved* by the following:

Roy B. Clariana  
Professor of Education (Learning, Design, and Technology)  
Dissertation Advisor  
Chair of Committee

Kyle L. Peck  
Professor of Education (Learning, Design, and Technology)  
Director of Graduate Studies

Edgar Yoder  
Professor of Extension Education

Irene Petrick  
Senior Lecturer of Information Sciences and Technology

Simon R. Hooper  
Professor of Education (Learning, Design, and Technology)

*Signatures are on file in the Graduate School
ABSTRACT

This quantitative experimental study examined the instructional impact of conceptual change-oriented refutation text on the development of different types of knowledge of college students. First and second year college students (N = 66) engaged with refutation text (RT) at different levels and their declarative knowledge, conceptual knowledge, and procedural knowledge were tested. Some students were asked to read the refutation text, while others were asked to read and paraphrase their understanding. The students were enrolled in two sections of a 200-level class, with different instructors teaching the same syllabus. Pretest and posttest data were collected. The three tests that assessed different types of knowledge were used to infer misconceptions at the declarative, conceptual, and procedural knowledge level.

Using MANOVA (with follow-up ANOVAs) and mixed within-between subjects statistical analysis, results suggested that RT simultaneously had a profound negative influence on declarative knowledge performance and a positive influence on conceptual knowledge performance. For declarative knowledge performance, results were therefore counter-intuitive. Previous research has shown that conceptual change-oriented and generative learning strategies can impede lower forms of knowledge such as verbatim (declarative) knowledge, and simultaneously improve higher forms of knowledge such as conceptual, structural, or procedural knowledge.

Moreover, a median split that created low-score and high-score groups showed pretest-posttest comparisons suggesting that RT had a profound negative impact on some
students’ declarative knowledge performance. Low-score groups who received RT group one treatment as well as those who received RT group two treatment improved from pretest to posttest, while high-score groups either decreased (RT group one) in conceptual knowledge performance or increased (RT group two). There were statistically significant differences between the low-score and high-score groups across pretest and posttest conceptual knowledge scores in both groups. Lastly, procedural knowledge pretest-posttest performances for the low-score and high-score groups did not result in statistically significant differences.

The focus on paraphrasing refutation text and studying the influence on three different types of knowledge extends previous research on the effect of conceptual change-oriented learning strategies on learning outcomes.

An extensive literature search, using the key terms and all variations of “refutation text” and “paraphrase”, failed to produce studies that explicitly paraphrased refutation text. More importantly, no study was found that paraphrased refutation text and tested declarative knowledge performance, conceptual knowledge performance, and procedural knowledge performance simultaneously. This suggests that this may be the first study to test the effects of combining conceptual change-oriented refutation text with paraphrasing while simultaneously testing for three types of knowledge.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Figures</td>
<td>ix</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xi</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>xiv</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>xv</td>
</tr>
<tr>
<td>Chapter 1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Chapter Abstract</td>
<td>1</td>
</tr>
<tr>
<td>Background</td>
<td>2</td>
</tr>
<tr>
<td>Different types of Knowledge</td>
<td>2</td>
</tr>
<tr>
<td>Misconceptions and Learning</td>
<td>3</td>
</tr>
<tr>
<td>Instructional Strategies and Learning Outcomes</td>
<td>5</td>
</tr>
<tr>
<td>Refutation Text vs. Non-Refutation Text</td>
<td>7</td>
</tr>
<tr>
<td>Refutation Text and Conceptual Change</td>
<td>9</td>
</tr>
<tr>
<td>Paraphrasing Refutation Text</td>
<td>11</td>
</tr>
<tr>
<td>Problem Statement</td>
<td>13</td>
</tr>
<tr>
<td>Purpose of the Study</td>
<td>14</td>
</tr>
<tr>
<td>Research Questions</td>
<td>15</td>
</tr>
<tr>
<td>Question 1: Effects of Reading vs. Reading and Paraphrasing Refutation Text on Different Types of Knowledge</td>
<td>15</td>
</tr>
<tr>
<td>Question 2: Effects of Reading Refutation Text vs. Not Reading Refutation Text on Different Types of Knowledge</td>
<td>16</td>
</tr>
<tr>
<td>Question 3: Effects of Reading and Paraphrasing Refutation Text vs. Not Reading Refutation Text on Different Types of Knowledge</td>
<td>16</td>
</tr>
<tr>
<td>Question 4: Low-score vs. High-score Pretest-Posttest differences on Different Types of Knowledge</td>
<td>16</td>
</tr>
<tr>
<td>Research Hypotheses</td>
<td>17</td>
</tr>
<tr>
<td>Significance of the Study</td>
<td>18</td>
</tr>
<tr>
<td>Educational Justification</td>
<td>18</td>
</tr>
<tr>
<td>Research Justification</td>
<td>19</td>
</tr>
<tr>
<td>Definitions of Terms</td>
<td>20</td>
</tr>
<tr>
<td>Chapter Summary</td>
<td>23</td>
</tr>
<tr>
<td>Chapter 2 LITERATURE REVIEW</td>
<td>24</td>
</tr>
<tr>
<td>Chapter Abstract</td>
<td>24</td>
</tr>
<tr>
<td>Theoretical Foundations</td>
<td>24</td>
</tr>
<tr>
<td>Conceptual Model</td>
<td>24</td>
</tr>
<tr>
<td>Misconception Repair Theory</td>
<td>28</td>
</tr>
</tbody>
</table>
Synthetic Meaning Theory ................................................................. 29
Knowledge-in-Pieces Theory .............................................................. 29
Impact of Refutation Text and Paraphrasing on Learning .................. 30
Refutation Text .................................................................................. 30
Paraphrasing ...................................................................................... 32
Concept Learning and Misconceptions ............................................. 34
Concept Categorization ..................................................................... 36
Misconceptions and Learning ............................................................. 37
Types of Knowledge and Learning Outcomes ................................ 39
Declarative Knowledge ..................................................................... 40
Conceptual Knowledge .................................................................... 41
Procedural Knowledge ..................................................................... 41
Chapter Summary ............................................................................. 42

Chapter 3 RESEARCH METHODOLOGIES ..................................... 44

Introduction ..................................................................................... 44
Pilot Studies ...................................................................................... 44
Main study ....................................................................................... 45
Study Context ................................................................................... 45
Research Design .............................................................................. 46
Variables ......................................................................................... 47
Participants ..................................................................................... 48
Summary of Study Design ................................................................. 51
Materials .......................................................................................... 53
Sources of Data ............................................................................... 53
Refutation Text ............................................................................... 54
Mapping of Refutation Text to Test Items ......................................... 57
Measurement Instruments ................................................................. 60
Measure of Declarative Knowledge .................................................. 61
Measure of Conceptual Knowledge .................................................. 63
Measure of Procedural Knowledge .................................................... 66
Procedures after intervention ........................................................... 68
Reliability and Validity ...................................................................... 69
Procedures ....................................................................................... 71
Overall data collection procedure .................................................... 71
Procedures prior to intervention ....................................................... 73
Procedures during intervention ........................................................ 75
Data Analyses .................................................................................. 76
Statistics applied to research questions .......................................... 76
Assumptions Testing ........................................................................ 77
Chapter Summary ............................................................................. 78

Chapter 4 RESULTS ........................................................................ 80
LIST OF FIGURES

Figure 2-1. A framework for relating misconceptions and learning outcomes. ...........26
Figure 3-1. Variables used in main study. .................................................................47
Figure 3-2. Participant groups and treatment levels. .................................................49
Figure 3-3. A refutation text item. ............................................................................55
Figure 3-4. Refutation Text groups, concepts, and misconceptions. ......................59
Figure 3-5. A declarative knowledge question. ..........................................................62
Figure 3-6. Screenshot of a partially completed conceptual knowledge test. ..........65
Figure 3-7. Screenshot of the procedural knowledge test.........................................66
Figure 3-8. Study’s high level events timeline. .........................................................73
Figure 4-1. Sample of misconceptions......................................................................87
Figure 4-2. Comparison of RT Group One declarative knowledge post test scores for reading vs. reading and paraphrasing.........................................................92
Figure 4-3. Comparison of RT Group One conceptual knowledge post test scores for reading vs. reading and paraphrasing.........................................................92
Figure 4-4. Comparison of RT Group One procedural knowledge post test scores for reading vs. reading and paraphrasing.........................................................92
Figure 4-5. Comparison of RT Group Two procedural knowledge post test scores for reading vs. reading and paraphrasing.........................................................93
Figure 4-6. Comparison of RT Group Two conceptual knowledge post test scores for reading vs. reading and paraphrasing.........................................................93
Figure 4-7. Comparison of RT Group Two declarative knowledge post test scores for reading vs. reading and paraphrasing.........................................................93
Figure 4-8. Profile plot for RT Group One low-score/high-score pre and post for declarative knowledge.................................................................105
Figure 4-9. Profile plot for RT Group Two low-score/high-score pre and post for declarative knowledge.................................................................105
Figure 4-10. Profile plot for conceptual knowledge pre and post, low vs. high score groups for RT group one treatment. ................................................................. 112

Figure 4-11. Profile plot for conceptual knowledge pre and post, low vs. High score groups for RT group two................................................................. 112

Figure 4-12. Profile plot for procedural knowledge pre and post, low vs. high score groups for RT group one treatment................................................. 120

Figure 4-13. Profile plot for procedural knowledge pre and post, low vs. high score groups for RT group two treatment................................................. 120
LIST OF TABLES

Table 1-1. Summary of Declarative, Conceptual, and Procedural Knowledge ........... 2

Table 1-2. Difference between Non-Refutation and Refutation Text using the Concept of NULL ................................................................. 8

Table 1-3. Definition of Terms ............................................................................. 20

Table 3-1. Number of College Semesters ............................................................ 45

Table 3-2. Summary of Data Submission .............................................................. 50

Table 3-3. Participant Groups, Concepts, and Tests ............................................. 58

Table 3-4. Rubric for Scoring Procedural Knowledge Test .................................. 67

Table 3-5. Reliability Analyses ............................................................................ 69

Table 4-1. Summary of Demographics and Composition of Groups per Research Question ......................................................................................... 81

Table 4-2. Change in Declarative Knowledge Pre And Posttest Responses .......... 84

Table 4-3. Summary of Incorrect Responses to Conceptual Knowledge Tests ...... 85

Table 4-4. Correlation for RT Group One ............................................................. 88

Table 4-5. Correlation for RT Group Two ............................................................. 88

Table 4-6. Read vs. Read and Paraphrase Posttest Means for RT Group One ........ 90

Table 4-7. Read vs. Read and Paraphrase Posttest Means for RT Group Two ......... 90

Table 4-8. MANOVA Results for RT Group One: Reading vs. Reading and Paraphrasing ......................................................................................... 94

Table 4-9. MANOVA Results for RT Group Two: Reading vs. Reading and Paraphrasing ......................................................................................... 95

Table 4-10. ANOVA Results for RT Group Two: Reading vs. Reading and Paraphrasing ......................................................................................... 96

Table 4-11. Posttest Means for RT Group One Students who did Not Read RT Group Two Treatment vs. RT Group Two Students who Read RT Group Two Treatment ................................................................................. 98
Table 4-12. MANOVA for RT Group One Students who did Not Receive RT Set Two vs. RT Group Two Students who did Not Receive RT Set One .................. 99

Table 4-13. Posttest Means for RT Group One Students who did Not Read and Paraphrase RT Group Two Treatment vs. RT Group Two Students who Read and Paraphrased RT Group Two Treatment ........................................ 101

Table 4-14. MANOVA for RT Group Two Students who did Not Receive RT Set One vs. RT Group One Students who did Not Receive RT Set Two ................. 102

Table 4-15. Declarative Knowledge Pre and Posttest Means for Low Score vs. High Score Groups: RT Group One Treatment ........................................ 104

Table 4-16. Declarative Knowledge Pre and Posttest Means for Low Score Vs High Score Groups: RT Group Two Treatment ........................................ 104

Table 4-17. MANOVA for Pre and Post, Low and High Score Groups for Declarative Knowledge RT Group One ..................................................... 107

Table 4-18. MANOVA For Pre and Post, Low and High Score Groups for Declarative Knowledge RT Group Two ..................................................... 108

Table 4-19. Within-Subjects Tests for Pre and Post, Low and High Score Groups for Declarative Knowledge RT Group One ........................................ 109

Table 4-20. Within-Subjects Tests for Pre and Post, Low and High Score Groups for Declarative Knowledge RT Group Two ........................................ 110

Table 4-21. Conceptual Knowledge Pre and Posttest Means for Low Score vs. High Score Groups: RT Group One Treatment ........................................ 111

Table 4-22. Conceptual Knowledge Pre and Posttest Means for Low Score vs. High Score Groups: RT Group Two Treatment ........................................ 111

Table 4-23. MANOVA For Pre and Post, Low and High Score Groups for Conceptual Knowledge RT Group One ..................................................... 114

Table 4-24. MANOVA for Pre and Post, Low and High Score Groups for Conceptual Knowledge RT Group Two ..................................................... 115

Table 4-25. Within-Subjects Tests for Pre and Post, Low and High Score Groups for Conceptual Knowledge RT Group One ........................................ 116

Table 4-26. Within-Subjects Tests for Pre and Post, Low and High Score Groups for Conceptual Knowledge RT Group Two ........................................ 117
Table 4-27. Procedural Knowledge Pre and Posttest Means For Low Score Vs High Score Groups: RT Group One ................................................................. 119

Table 4-28. Procedural Knowledge Pre and Posttest Means For Low Score Vs High Score Groups: RT Group Two ................................................................. 119

Table 4-29. MANOVA For Procedural Knowledge Pre and Post, Low vs. High Score Groups for RT Group One Treatment ......................................................... 122

Table 4-30. MANOVA For Pre and Post, Low and High Score Groups For Procedural Knowledge RT Group Two ................................................................. 123

Table 4-31. Within-Subjects Tests For Pre and Post, Low and High Score Groups for Procedural Knowledge RT Group One ......................................................... 124

Table 4-32. Within-Subjects Tests for Pre and Post, Low and High Score Groups for Procedural Knowledge RT Group Two ......................................................... 125

Table 4-33. Summary of Statistical Results ........................................................................... 126

Table 5-1. Relationship of Groups, Concept Ratings, and Levels of Treatment per Group .............................................................................................................. 132
LIST OF ABBREVIATIONS

CK       Conceptual Knowledge  
DK       Declarative Knowledge  
PK       Procedural Knowledge  
RT       Refutation Text  
RQ       Research Question
ACKNOWLEDGEMENTS

To my committee members—Roy Clariana (advisor and chair), Simon Hooper, Irene Petrick, Edgar Yoder, and Kyle Peck—for their expert opinions and advice: ngiyabonga (thank you). Special thanks to Ladislaus Semali for valuable research design suggestions and to Ellysa Stern Cahoy for indispensable guidance navigating the library system. To the College of Information Sciences and Technology students who participated in pilots and the main study, and the instructors—Irene Petrick, Erika Poole, Elizabeth Thiry, and Nathan Aileo—who granted access to their students: ngiyabonga.

To the support from the faculty, staff, and administration of the Department of Learning and Performance Systems at Penn State University’s College of Education: ngiyabonga.

On a personal note, this work is dedicated to my ancestors, the villages that raised me, a family full of pure and simple love, an incredible father, and the most precious mothers: Lomthandazo Sophie Dlamini and Qobhile Concelia Ntshalintshali. Your countless sacrifices will never be forgotten. Thank you for letting me stand on your shoulders. Finally, this is dedicated to the deferred dream that has propelled me since my preteen years: Makhosini Constantine “Madansane” Hlophe. May this work and the hope it embodies help find you Mabhengu, Khonkosi, mfana wa gogo.
Chapter 1

INTRODUCTION

Chapter Abstract

This chapter highlights the need for incremental learning, from grasping essential facts and relating concepts to applying knowledge. It provides an overview of why the need to correctly recall, understand, and apply knowledge means misconceptions must be identified and repaired for increased success as students who are expected to solve problems. The link between misconceptions and instructional strategies is discussed. While instructional strategies play a pivotal role in learning and application of correct conceptions (since this results in improved performance), this chapter discusses why instructional strategies play an equally important role in identifying misconceptions.

One of the instructional strategies—refutation text—is then introduced as a practical and readily available option for repairing misconceptions. Lastly, the goal of the study, research questions, research hypotheses, and significance of the study are developed. The chapter concludes by defining terms used in the main study.
Background

Different types of Knowledge

The ability to acquire subject-specific facts, relate concepts that belong to a particular knowledge domain, and apply appropriate techniques to solve problems is a necessary quality for students because it results in improved performance. Knowledge of facts is declarative knowledge, the ability to compare and classify them is conceptual knowledge, and the ability to apply declarative and conceptual knowledge in solving a problem is procedural knowledge. All three types of knowledge are essential to mastery and application of concepts (Jonassen, Beissner, & Yacci, 1993). Table 1-1 summarizes the three types of knowledge tested in the main study (Grotzner, 2002; Jonassen et al., 1993; McCormick, 1997; Rittle-Johnson, Star, & Durkin, 2009).

Table 1-1

<table>
<thead>
<tr>
<th>Declarative Knowledge</th>
<th>Conceptual Knowledge</th>
<th>Procedural Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Knowledge of facts</td>
<td>• Knowledge of classification</td>
<td>• Knowledge of techniques</td>
</tr>
<tr>
<td>• Precise detail</td>
<td>• Ability to compare, relate, summarize</td>
<td>• Ability to identify and apply appropriate methods</td>
</tr>
<tr>
<td>• Ability to name, list, describe</td>
<td>• Indicates grasp of declarative knowledge</td>
<td>• Indicates ability to apply declarative and conceptual knowledge</td>
</tr>
<tr>
<td>• Knowledge of properties</td>
<td>• Knowledge of principles</td>
<td>• Less specific than declarative and conceptual knowledge</td>
</tr>
<tr>
<td>• Illustrated by recall</td>
<td>• Illustrated by ability to categorize</td>
<td>• Illustrated by ability to solve open-ended problems</td>
</tr>
<tr>
<td>• Important to learning conceptual and procedural knowledge</td>
<td>• Important to learning procedural knowledge</td>
<td></td>
</tr>
</tbody>
</table>
As Table 1-1 illustrates, declarative knowledge is about facts. With declarative knowledge, the type of knowledge components processed are more distinct and therefore easier to identify (Son, Park, & Kim, 2011) than relatively higher order knowledge components such as procedural knowledge (reviewed in a subsequent section). Beyond the mere acquisition of isolated facts, connecting and applying concepts is necessary for using increasingly complex knowledge to solve problems.

**Misconceptions and Learning**

A misconception is a unique kind of error that is often difficult to identify and correct with formal instruction (Chi, Roscoe, Slotta, Roy, & Chase, 2012). Van den Broek and Kendeou (2008) refer to misconceptions as knowledge that is not accurate, while Hancock (1940) says misconceptions are false beliefs that cannot be rationalized or justifiably held up to normative standards of a particular knowledge domain. Unlike a typical error, a misconception persists so that traditional instruction sometimes fails to correct it (Keleş, Çepni, Haşíloğlu, & Aydin, 2011). Misconceptions therefore result in poor performance by students. For example, Chi and Roscoe (2002) identified learners who imagined electricity only as something that could be stored and therefore a substance that can leak (based on their understanding of batteries). According to their grasp of the topic, if electricity is a substance, then it can be stored in boxes and cans. In the main study, the knowledge domain area tested was relational database design.
To understand misconceptions, the definition of concept requires clarity; a concept is an entity that represents a part of knowledge or cognitive skill held by a learner (Reye, 1996). In addition, a concept can vary in scope from a word or definition to a topic, section, or even a chapter (Schank & Cleary, 1995).

The fact that misconceptions can happen at all levels of the knowledge continuum means that whether it is a false belief in erroneous facts, perpetual failure to associate two or more related ideas, or consistent failure to apply subject-specific rules, they are still regarded as misconceptions. These misconceptions require instructional strategies that will result in improved performance by students.

According to Hammer (1996), a misconception consists of several features: one, the misunderstanding of knowledge is deep-rooted and therefore persists over time; two, a misconception clashes with normative understanding or correct conception; three, the firmly entrenched misunderstanding prevents the correct understanding of more complex knowledge; and four, to reach an expert-level grasp of knowledge, misconceptions must first be repaired or removed completely. Misconceptions are therefore intuitive errors that are well-entrenched but false, hence they limit the ability of the learner to correctly understand knowledge at a factual level, conceptual level, or application level (Karl, Toonen, Grant, & Bowen, 2012).

But how do misconceptions develop among learners, and why are they difficult to identify and repair? A dominant view is that misconceptions develop when learners incorrectly categorize new or updated information when assimilating knowledge (Chi, Nokes-Malach, & Gadgil, 2012). This results in views that clash with the standards of a
particular knowledge domain. For example, intuitive errors about inertia (resistance of objects to change in motion) often lead to different interpretations of falling objects when students interpret laws of motion (Hynd & Alvermann, 1985). As a result, people sometimes hold multiple conceptions of the same idea within a knowledge domain, even when the knowledge is contradictory. As a result, cognitive tension may occur in learners’ minds when their understanding contradicts accepted norms of a subject domain (Tsang & Williams, 2012). But cognitive tension does not merely happen without reason; it is usually activated by introducing new knowledge that challenges previous knowledge.

Opportunities to appropriately apply knowledge—for example, to solve problems—are reduced when misconceptions relating to particular subject matter adversely affect a knowledge domain’s normative understanding (Mayer, 2002). According to Karl et al. (2012), misconceptions are likely to interfere with comprehension of subject matter because they tend to persist and hinder successful learning. The hindrance of learning by misconceptions suggests that successfully achieving subject-specific learning outcomes may be at risk if misconceptions are not identified and repaired in part or removed entirely if recognized as completely false.

**Instructional Strategies and Learning Outcomes**

Different instructional strategies are tailored for identifying and correcting misconceptions at various levels of a knowledge continuum that involves acquiring facts,
identifying and relating concepts, and using learned rules to solve problems. The achievement of learning outcomes is therefore directly related to the choice of instructional approach (Omoniyi & Adedapo, 2012). A learning outcome is the fulfillment of specific knowledge, skill, or disposition related to subject matter (Seel, 2012d). According to Mayer (1989), it involves a process that results in the acquisition of knowledge. In turn, what one does with a learning outcome is perform; that is, performance is the substantiation of a learning outcome (Mayer, 1989). Examples of performances that results from learning outcomes are declarative knowledge performance, conceptual knowledge performance, and procedural knowledge performance (Benbunan-Fich & Arbaugh, 2006; Trigwell & Prosser, 2011).

Declarative or factual knowledge is defined as knowledge of terminology and specific details, conceptual knowledge as knowledge of classifying concepts (indicates schema or knowledge or related ideas), and procedural knowledge as knowledge of skills and techniques applied to a particular area of knowledge (Clariana & Prestera, 2009; Hallett, Nunes, & Bryant, 2010; Kumaran, Summerfield, Hassabis, & Maguire, 2009). Examples of activities involved in these learning outcomes include the following: identifying terms as a way of demonstrating declarative knowledge, comparing or relating concepts to demonstrate conceptual knowledge, and implementing rules or procedures as a way of demonstrating knowledge application (Krathwohl, 2002). In addition, different levels of knowledge occur in a continuum; that is, cognitive complexity (Anderson, Dewhurst, & Nash, 2012) increases from declarative knowledge
performance to conceptual knowledge performance and through to procedural knowledge performance.

The successful fulfillment of learning outcomes is linked to the choice of instructional strategy. Examples of instructional strategies include providing feedback during instruction, comparing examples with non-examples, experimenting with examples, using imagery, and involving learners during instruction (Booth, Lange, Koedinger, & Newton, 2013; Clariana & Koul, 2006; Shen, Lee, Tsai, & Ting, 2012). In addition, proponents of conceptual change posit that instructional strategies that specifically target misconceptions may be most effective in correcting them (Chi, 2008). Conceptual change therefore concerns itself with altering misconceptions (Van den Broek & Kendeou, 2008). For example, conceptual change-oriented instructional approaches exist to expose valuable knowledge that needs to be learned or highlight misconceptions as a first step to repairing or removing them (Diakidoy, Mouskounti, & Ioannides, 2011; Tippett, 2010). One of these strategies is called refutation text.

**Refutation Text vs. Non-Refutation Text**

Refutation text is text designed to target misconceptions by highlighting their existence and then confronting a learner with an alternative version of text regarded as normative to a knowledge domain (Guzzetti, 2000). It is persuasive in nature; that is, it is meant to argue for a correct concept while simultaneously displacing or repairing a misconception (Van den Broek & Kendeou, 2008). On the other hand, non-refutation
text (often referred to as expository text) merely highlights important concepts that should be learned without directly targeting naïve or layman’s understandings and theories held by learners. By emphasizing consistency and highlighting ideas that would otherwise be unclear to a reader, refutation text improves memorization and understanding of text (Van den Broek & Kendeou, 2008). Consider an example of the concept of NULL in information organization. Table 1-2 illustrates how NULL is communicated as refutation text and non-refutation or expository text.

Table 1-2

*Difference between Non-Refutation and Refutation Text using the Concept of NULL*

<table>
<thead>
<tr>
<th>Concept NULL</th>
<th>Non-refutation (expository) Text</th>
<th>Refutation Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>NULL means the value is unknown, inapplicable, or withheld.</td>
<td>Some students confuse the meaning and use of NULL. Their understanding is that NULL means nothing, or worse still, zero.</td>
<td>But this is not true. NULL does not mean nothing or zero. It means the value is unknown, or the value is inapplicable (e.g. # of pregnancies for a man), or the value is being withheld (the boss’s salary). Unknown is especially tricky, but consider it like you would a half-truth.</td>
</tr>
</tbody>
</table>

As Table 1-2 illustrates, the main difference between refutation and non-refutation text is that refutation text explicitly tells the learner that their existing understanding is erroneous, whereas non-refutation text contains only the correct
information without stating that it is addressing a particular misconception (Broughton, Sinatra, & Reynolds, 2010). Non-refutation text does not mention misconceptions (Van den Broek & Kendeou, 2008). The wording in refutation text is intended to inform learners that what they will read next will likely conflict with their main study understanding of a concept. According to cognitive psychology research, text akin to But this is not true (see Table 1-2) may indicate to the reader that the next text they will encounter is important (Broughton et. al., 2010).

According to Broughton et al. (2010), the central aim of conceptual change-oriented refutation text is to activate cognitive conflict (raise awareness of an error in understanding) by focusing directly on the learner’s misunderstanding. Refutation text is effective because it co-actives conflicting versions of knowledge (therefore causing the reader to re-think) and then explains the conflict concisely (Van den Broek & Kendeou, 2008). The co-activation of prior knowledge and new knowledge helps learners realize misunderstandings and makes it possible to repair misconceptions. It is therefore necessary for challenging and addressing the concepts some learners misunderstand (Baralt, 2013). By highlighting intuitive errors, students’ performance can be improved.

**Refutation Text and Conceptual Change**

Conceptual change-oriented refutation text targets misconceptions so that when learners encounter advanced knowledge, intuitive errors can be revised or replaced. Chi (2008) summarizes learning into three types: one, new knowledge can be added where no
prior knowledge exists; two, knowledge can be added where prior knowledge is incomplete; and three, new knowledge is introduced to change incorrect prior knowledge. The main study focuses on changing incorrect prior knowledge and defines learning as such. By challenging learners’ misconceptions in a particular knowledge domain, refutation text coerces learners to examine their main study understanding and consider proposed alternatives. According to Çalik, Kolomuç, and Karagölge (2010), interacting with refutation text helped learners correct some long-held misconceptions about fluid pressure. A key feature of refutation text, according to Van den Broek and Kendeou (2008) and Posner, Strike, Hewson, and Gertzog (1982), is co-activation: the simultaneous activation of prior knowledge and new knowledge presented to a reader. By co-activating incorrect and main studyly accepted knowledge, refutation text contributes toward modifying concepts.

However, co-activation does not guarantee correction of misconceptions. According to Chi et al. (2012), the ease with each misconceptions are successfully refuted depends on several conditions: learners insist on their prior knowledge (robustness); learners display the same misconception in different contexts (consistency); learners repeat the same misconception from lower to higher education levels (persistency); learners pursuing different study paths display related incorrect conceptions (homogeneity); learners from different eras, e.g., “medieval scientists and contemporary naïve students” (Chi et al. 2012, p. 35) display the same incorrect conception (recapitulation); learners display incorrect conceptions that systematically upholds an empirically unproven theory (systematic). In short, learning strategies
intended to modify concepts have differing chances of success depending a wide variety of factors.

Because refutation text is a promising conceptual change-oriented instructional strategy, the main study focused on the effects of conceptual change-oriented refutation text as a possible solution to repairing misconceptions, and therefore achieving targeted learning outcomes. Since some students carry misconceptions about subject matter, further empirical investigation is needed to clarify their nuanced nature. With this in mind, the overarching question is how do we more effectively align instruction with learning to repair misconceptions among students? By implication, how should we attempt to reveal and repair misconceptions in order to optimize learning outcomes given busy college course schedules?

One way that conceptual change-oriented refutation text can be utilized, beyond merely reading, is to paraphrase or restate in different forms for clarity. In addition to clarity, paraphrasing can be used to determine if students received the message they were supposed to receive, that is, if the refutation text itself needs to be improved or changed. The next section reviews how paraphrasing can be applied to refutation text.

Paraphrasing Refutation Text

To paraphrase means to restate or reword text for clarity (Kletzien, 2009). Compared with quoting directly, paraphrasing shows that a learner has assimilated meaning and is able to reproduce the original text according to his or her own
understanding (Schumaker, Denton, & Deshler, 1984). For example, by using a paraphrasing verbs strategy, “tear gas” is paraphrased to “gas that causes tears” (Nakov & Hearst, 2013). In another example, one of the refutation text items used in the main study concerns knowing the difference between logical and physical database modeling. Part of the writing in the refutation text (also paraphrased from other sources) reads thus:

*Physical modeling depends on the specific RDBMS being used, but logical modeling deals with business requirements.* One way to paraphrase this text would be to re-write:

*Logical database modeling, unlike physical modeling which tends to be more database system-specific, focuses on business needs.* This paraphrase changed sentence structure, replaced some words with synonyms (e.g., replaced requirements with needs), and generalized the acronym RDMBS (relational database management system) to database. Additional paraphrasing options such as using antonyms (Hagaman, Casey, & Reid, 2012) could also have been used. The purpose is to demonstrate one’s own understanding without altering the core meaning of the source text. The main study tested the impact that paraphrasing refutation text had on repairing misconceptions. Declarative knowledge, conceptual knowledge, and procedural knowledge tests were used to assess the influence of the refutation text.

To better understand how subject matter misconceptions or intuitive errors can be most efficiently repaired, different levels of knowledge in terms of cognitive complexity—from declarative to conceptual to procedural—were considered. With this in mind, the main study sought to test the misconception repair theory of conceptual change by comparing refutation text reading and refutation text paraphrasing in terms of
declarative knowledge performance, conceptual knowledge performance, and procedural knowledge performance.

**Problem Statement**

Misconceptions of subject matter that persist undetected in students are a source of concern because they negatively impact students’ learning outcomes (Chi, et al., 2012). Because misconceptions sometimes continue undetected, there is a need for instructional strategies to align with learners’ prior knowledge before efforts can be made to reduce long-held intuitive errors.

Misconceptions are partly responsible for causing learning problems. Chi (2008) and diSessa (2006) studied underlying causes of misconceptions, examined how misconceptions likely occur in the mind and why they persist even when targeted guidance is applied. Despite the attention allotted to misconceptions, there has been limited focus on examining the different types of knowledge that learners possesses, how the knowledge can be elicited in practical and economical ways without disrupting teaching environments, and how the learners’ knowledge profiles affect learning outcomes (Chi, et al., 2012).

Instructional strategies that target students’ misconceptions might be regarded as yet another administrative task. To address such concerns, Coll, Çalik, and Ayas (2007) used traditional multiple-choice and short-paragraph questions to reveal and repair learners’ misconceptions during a study about fluid pressure. The study highlighted
practicality while testing, identifying, and repairing learners’ misconceptions. Proponents of simpler, non-intrusive methods (as opposed to more intrusive methods such as eye-tracking and think aloud strategies) have recommended instructional strategies such as refutation text to repair misconceptions without noticeably disrupting existing learning and teaching environments (Al khawaldeh & Al Olaimat, 2010); therefore, the main study considered refutation text as an example of a practical instructional strategy for reducing misconceptions among college students.

**Purpose of the Study**

The purpose of this research was to examine the effects of conceptual change-oriented refutation text on different types of knowledge among college students studying information organization. Within information sciences, the knowledge domain area tested was relational database design.

While instructional strategies such as learner involvement, use of imagery, and experimentation of examples have been used to address misconceptions (Anderson et al., 2012; McLaren et al., 2012; Shen et al., 2012), the main study focused on refutation text.

To minimize cognitive complexity (avoid confusion) while engaging with conceptual change-oriented refutation text, the refutation text items are kept as few and brief as possible; for example, in a study by Diakidoy, Kendeou, and Ioannides (2003) about energy, only 4 refutation text items were used, and all 4 items targeted misconceptions about the concept of energy. The implication is that there may be a
cognitive tipping point whereby too many refutation text items may have adverse effects on students in the form of excessive confusion (cognitive complexity may help explain results). The main study purposefully followed the same cognitive principle by limiting the number of refutation text items that each participant received. However, the study did not analyze whether or not five refutation text items created any more or less confusion than all ten would have.

The study examined the misconception repair theory of conceptual change by comparing learners who engaged with refutation text at different levels before taking knowledge tests. Specifically, the study analyzed the differential effects of conceptual change-oriented refutation text on students’ declarative knowledge performance, conceptual knowledge performance, and procedural knowledge performance when reading refutation text versus when reading and paraphrasing refutation text, and in a control condition.

Research Questions

**Question 1: Effects of Reading vs. Reading and Paraphrasing Refutation Text on Different Types of Knowledge**

How does reading refutation text differ from reading and paraphrasing refutation text in terms of declarative knowledge performance, conceptual knowledge performance, and procedural knowledge performance?
Question 2: Effects of Reading Refutation Text vs. Not Reading Refutation Text on Different Types of Knowledge

How do refutation text group one students who did not read refutation text set two differ from refutation text group two students who read refutation text set two in terms of declarative knowledge performance, conceptual knowledge performance, and procedural knowledge performance?

Question 3: Effects of Reading and Paraphrasing Refutation Text vs. Not Reading Refutation Text on Different Types of Knowledge

How do refutation text group one students who did not read and paraphrase refutation text set one differ from refutation text group two students read and paraphrased refutation text set two in terms of declarative knowledge performance, conceptual knowledge performance, and procedural knowledge performance?

Question 4: Low-score vs. High-score Pretest-Posttest differences on Different Types of Knowledge

Are low-score group vs. high-score group differences statistically significant in terms of pretest-posttest changes in declarative knowledge, conceptual knowledge, and procedural knowledge performance?
Research Hypotheses

**H₀₁**: No statistically significant differences exist between *reading refutation text* and *reading and paraphrasing refutation text* in terms of *declarative knowledge performance*, *conceptual knowledge performance*, and *procedural knowledge performance*.

**H₀₂**: No statistically significant differences exist between *refutation text group one students who did not read refutation text set two* and *refutation text group two students who read refutation text set two* in terms of *declarative knowledge performance*, *conceptual knowledge performance*, and *procedural knowledge performance*.

**H₀₃**: No statistically significant differences exist between *refutation text group one students who did not read and paraphrase refutation text set two* and *refutation text group two students who read and paraphrased refutation text set two* in terms of *declarative knowledge performance*, *conceptual knowledge performance*, and *procedural knowledge performance*.

**H₀₄**: No statistically significant *low-score group vs. high-score group* differences exist in terms of *pretest-posttest* changes in *declarative knowledge*, *conceptual knowledge*, and *procedural knowledge performance*. 
Significance of the Study

Educational Justification

While conceptual change-oriented instructional strategies such as expository (non-refutation) text have been used to highlight important concepts to learners (Diakidoy et al., 2011), fewer studies have considered the more direct refutation text as an alternative approach for highlighting and repairing students’ misconceptions (Keleş et al., 2011; Tsang & Williams, 2012). While using refutation text is cognitively more intrusive since it seeks to activate cognitive conflict (highlight misunderstanding), it is still less intrusive than approaches such as analyzing students’ eye movements (Ariasi & Mason, 2011) or assessing their think aloud method. If instructional strategies can be developed and tailored to more deliberately target learning outcomes according to the levels of cognitive complexity at which they are expected to occur, there may be more successes at repairing misconceptions.

Due to early proliferation and maturation of theories in the biological and physical sciences, research on misconceptions historically has an established base in the physical sciences (Alwan, 2011; Gräsel, Parchmann, & Beerenwinkel, 2011; Hynd & Alvermann, 1985). To this point, Chi, Slotta, and De Leeuw (1994) and Tippett (2010) stated that since identifying misconceptions is more readily acknowledged in the physical sciences, limited research in other fields should present opportunity for further investigation. As a result, the knowledge domain area examined in the main study was relational database design.
Furthermore, since there are different levels of knowledge, a study about misconceptions must consider the fact that false beliefs also occur at different levels of cognition. This partly explains why the initial view of conceptual change advanced by Posner et. al. (1982) has evolved to several views such as Chi and Roscoe's (2002) misconception repair theory of conceptual change. Notwithstanding, subsequent extensions of Posner et. al.’s (1982) conceptual change theory maintain its core stance: to effect conceptual change learners must experience dissatisfaction with the target concept; there must be a viable alternative concept that should be demonstrable; and the learner must regard the new concept as beneficial and applicable (Van den Broek & Kendeou, 2008). Despite growing awareness of different types of knowledge, there is limited research (Peterson, Ridenour, & Somers, 1990; Robinson & Dube, 2007) studying multiple types of knowledge simultaneously.

**Research Justification**

For the past two decades, research related to conceptual change has encountered different interpretations and definitions. As a result of the uncertainty, the problem of how to effectively repair misconceptions remains insufficiently studied (An & Wu, 2012; diSessa, 2006).

Tsang and Williams (2012) highlighted the need to analyze factors influencing misconceptions and application of knowledge, both of which require a grasp of how learning happens. In turn, understanding the learning process entails understanding
conceptual change as a phenomenon, but this has been difficult to achieve (Chi et al., 2012; Kowalski & Taylor, 2009).

The connection between misconceptions and learning outcomes has not been examined to the extent that learning science researchers can effectively extend theory to teaching and learning environments (Chi, Nokes-Malach, & Gadgil, 2012). Conceptual change research remains scant in detail and practical recommendations that address concerns about cost effectiveness, time implications, and practicality of implementation in real-life educational settings (Coll et al., 2007).

**Definitions of Terms**

Table 1-3 defines terms as they are applied in the main study.

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cognition</strong></td>
<td>The way in which information is processed; for example, processes include memory, attention, emotion, and perception (Seel, 2012a).</td>
</tr>
<tr>
<td><strong>Cognitive Complexity</strong></td>
<td>Cognitive functioning that ranges from using few knowledge constructs to many constructs, with the use of constructs ranging from simple to complex (Baralt, 2013).</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Concept</strong></td>
<td>“Labeled sets of objects, symbols, or events that share common characteristics or critical attributes. A concept is a mental construct or representation of a category that allows one to identify examples and non-examples of the category” (Schunk, 2008)</td>
</tr>
<tr>
<td><strong>Conceptual Change</strong></td>
<td>“Modification of misconceptions” (Van den Broek &amp; Kendeou, 2008, p. 336) or replacement of flawed understanding of concepts with empirically verified understanding (Posner, Strike, Hewson, &amp; Gertzog, 1982).</td>
</tr>
<tr>
<td><strong>Conceptual Knowledge</strong></td>
<td>Knowledge about the organization, classification, and relationships of the various parts of concepts that make up a particular domain knowledge (de Jong &amp; Ferguson-Hessler, 1996)</td>
</tr>
<tr>
<td><strong>Declarative Knowledge</strong></td>
<td>Knowledge-What or explicit knowledge of a fact, concept, event, or object (Fantl, 2012)</td>
</tr>
<tr>
<td><strong>Learning</strong></td>
<td>To gain knowledge, comprehension, or mastery through experience or study (Seel, 2012f)</td>
</tr>
<tr>
<td><strong>Learning Outcome</strong></td>
<td>Product, effect, or result gained from receiving instruction or acquiring knowledge (Seel, 2012d)</td>
</tr>
<tr>
<td><strong>Misconception</strong></td>
<td>Misinterpretations and false or unfounded beliefs that occur when ideas are miscategorized (Hancock, 1940) or formed without</td>
</tr>
<tr>
<td><strong>Term</strong></td>
<td><strong>Definition</strong></td>
</tr>
<tr>
<td>--------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>regard to “true” knowledge or experience</strong> (Emerite, 2012)</td>
<td></td>
</tr>
<tr>
<td><strong>Misconception Repair</strong></td>
<td>The incorporation of knowledge into a set of related concepts (schema) in such a way that the resulting or changed concept set is coherent with a particular knowledge domain (Rodrigues, Gabbay, &amp; Russo, 2011).</td>
</tr>
<tr>
<td><strong>Paraphrase</strong></td>
<td>Process of restating text in order to communicate meaning in a different form, often intended for further clarity of understanding (Hagaman et al., 2012). It is a form “memory representation” (Broughton, Sinatra, &amp; Reynolds, 2010, p. 408).</td>
</tr>
<tr>
<td><strong>Performance</strong></td>
<td>“What students can do as a result of learning” (Mayer, 1989, p. 47)</td>
</tr>
<tr>
<td><strong>Procedural Knowledge</strong></td>
<td>Knowledge about how to perform a specific task: also known as “knowledge-how” or knowledge of subject-specific skills (Seel, 2012e).</td>
</tr>
<tr>
<td><strong>Refutation Text</strong></td>
<td>Text constructed to directly confront and dispute learners’ misconceptions and then provide the correct conception (Tippett, 2010)</td>
</tr>
<tr>
<td><strong>Schema</strong></td>
<td>A set of related concepts, cognitive structures, or knowledge structures (Rumelhart &amp; Ortony, 1976)</td>
</tr>
</tbody>
</table>
Chapter Summary

Chapter 1 presented misconceptions and their expected negative impact on learning outcomes, highlighted their prevalence in learning environments, and emphasized the difficulties associated with identifying and repairing them. Refutation text and its role in challenging misconceptions were presented.

The next chapter reviews literature pertinent to conceptual change as a theoretical base. Concept learning, misconceptions, and refutation text are also synthesized. In view of the multi-faceted approach that teaching and learning assumes, literature linking learning outcomes to the environments within which they occur is analyzed.
Chapter 2

LITERATURE REVIEW

Chapter Abstract

This chapter reviews theoretical foundations of conceptual change, aspects of learning aligned with conceptual change theory, instructional strategies in general and refutation text in particular, and types of knowledge. With conceptual change theory as a foundation, the potential impact of paraphrasing refutation text is reviewed.

To directly address intuitive errors found in some students’ knowledge base, reviews of refutation text suggest that it may be more effective than expository or non-refutation text in repairing misconceptions. Thus, several other instructional strategies are reviewed, and so is the relationship between learning and developing misconceptions. Different types of knowledge are synthesized. The review concludes that the ability to identify and test for different types of knowledge may help elicit misconceptions.

Theoretical Foundations

Conceptual Model

To understand how refutation text may repair conceptions and affect conceptual change, the main study was conducted through the theoretical lens of conceptual change
theory (Posner et al., 1982). Posner et al. (1982) based their conceptual change theory on Kuhn’s (1970) notion of scientific revolutions and Piaget’s (1977) stance on conceptual conflict. From the original conceptual change theory that focused mostly on cognitive aspects of learning, Strike and Posner (1992) and others (Pintrich, Marx, & Boyle, 1993) revised it to include additional elements such as motivation and the role of social and environmental issues. The latter conceptual change theory was considered more holistic, that is, it recognized that learners’ ability to think (cognition) is influenced by a variety of factors.

To this point (cognitive or conceptual conflict as the main consideration of the conceptual change theory), misconceptions held by learners were tested and attempts made to activate cognitive conflict, with the end goal being to effect conceptual change by reducing misconceptions.

The following rationale is applied: Misconceptions are assumed to not only be isolated abstract representations of knowledge, but for each domain area—database design in the main study—the abstract representations relate to each other in varying degrees of strength (from strong to weak relations).

Furthermore, the main study considers misconception repair theory as one of several that clarify the meaning of conceptual change (Chi & Roscoe, 2002; DiSessa, 2000; Ivarsson, Schoultz, & Säljö, 2002; Vosniadou & Brewer, 1992). A framework for understanding the holistic environment within which misconceptions occur is illustrated in Figure 2-1.
Figure 2-1. A framework for relating misconceptions and learning outcomes.

Figure 2-1 illustrates a framework that comprises conceptual change theory and schema theory as the theoretical basis employed in the main study. In this framework, selected instructional strategies influence acquisition of facts, concepts, and procedures along a knowledge continuum that increases in cognitive complexity. The interplay of different types of knowledge results in misconceptions, in normative knowledge, or in a combination of both. The presence of misconceptions or correct conceptions (normative knowledge) influences learning outcomes.

In some instances, both misconceptions and correct conceptions exist concurrently (within the same schema or set of related knowledge structures). The theories shown in Figure 2-1 help explain not only the multifaceted influences of
teaching and learning, but also the reasons misconceptions occur alongside knowledge regarded as normative.

To repair misconceptions with a view to optimizing knowledge application requires that learners’ foundational knowledge is identified; thus, what learners are able to declare, or compare and contrast in terms of conceptual knowledge helps reveal intuitive errors (Domin & Bodner, 2012). Achieving the ability to apply subject-specific skills and learned rules is often linked to knowledge gains made at the declarative and conceptual levels. Solid acquisition of facts, terminology, principles, and theories of a domain area usually translates to success with increasingly challenging tasks such as relating and applying knowledge (Bloom, 1956).

According to Domin and Bodner (2012), there is significant overlap between conceptual understanding and knowledge application. By repairing misconceptions, learners can more successfully recognize and integrate related concepts. Chi et al. (2012) explain this notion further by stating that the ability to achieve a level of conceptual understanding required to apply rules to unfamiliar tasks calls for identification and repair of misconceptions. Unless intuitive errors are confronted cognitively, learners may continue carrying misconceptions that limit their ability to relate, integrate, and apply appropriate skills.

Following Posner et al.’s (1982) seminal contribution to the understanding of conceptual change, several theories have become notable for their different and often conflicting approaches to addressing conceptual change: Chi and Roscoe's (2002) misconception repair or belief revision theory, DiSessa's (2000) knowledge-in-pieces
theory, Vosniadou and Brewer's (1992) synthetic meaning theory, and Ivarsson et al.'s (2002) sociocultural theory. The first three theories are reviewed: the fourth theory—sociocultural theory of conceptual change (Ivarsson et al., 2002)—falls outside the scope of the main study since it does not consider prior knowledge as a catalyst or a necessary obstacle to learning. Elements of stigmergy (indirect interplay of people and environmental factors), whereby knowledge occurs based partly on others’ knowledge (Marsh & Onof, 2008), did not fall within the scope of this research.

**Misconception Repair Theory**

The misconception repair theory of conceptual change advocates transforming flawed schema, frameworks, or mental models to knowledge considered normative (Chi & Roscoe, 2002). Flawed schemata equate to misconceptions, while accurate knowledge holds up to standards accepted as correct by a field of study. The notion of misconception repair means there is existing knowledge that ought to undergo a process of partial revision or complete replacement. The knowledge targeted for review and repair is either incorrect, or correct but incomplete. In short, the target of revision are ideas that are considered inaccurate (Van den Broek & Kendeou, 2008). It is therefore a misconception that requires repair before additional knowledge is gained via instruction. Misconception repair occurs in the mind and its main feature is the replacement of previous misunderstandings. Learners must therefore fix their own misconceptions (Chi, 2008; Mayer, 2002).
Synthetic Meaning Theory

A second theory of conceptual change, synthetic meaning, depicts change as a process of blending or combining knowledge representations known as mental models (Vosniadou & Brewer, 1992), metaphorically described as ‘adding rooms to the house’. Unlike Chi and Roscoe (2002) who view misconceptions as flawed mental models or faulty schema that must be removed and replaced (i.e., ‘tear down the old rooms’), the synthetic meaning theory considers misconceptions as incomplete models that require synthesizing. Up to the point that misconceptions are activated, ignorance is viewed as incoherent knowledge. Change towards correct conception occurs as incoherent mental models are re-synthesized into coherent mental models. Prior knowledge therefore serves an important function in the synthetic meaning theory: it is both a necessary obstacle as well as a catalyst for change. The synthetic meaning process of conceptual change also occurs in the mind and its main feature is synthesis. In this view learners must synthesize previously meaningless or incoherent information into meaningful knowledge (Mayer, 2002).

Knowledge-in-Pieces Theory

Knowledge-in-pieces theory of conceptual change holds that misconceptions are unorganized pieces of information that require structure (diSessa, 2000). In this case, change comprises transformation of knowledge from formless to structured, with learners themselves as central participants of the organization process (diSessa, 2000). Similar to
Vosniadou and Brewer's (1992) synthetic meaning view, knowledge-in-pieces also uses prior knowledge as a catalyst for change. This approach also considers change to occur in the mind, and its main feature is categorizing or systematizing. In knowledge-in-pieces, learners must re-organize knowledge so that cognitive structures become related knowledge pieces (Mayer, 2002).

From the different approaches to conceptual change, the main study adopted the misconception repair theory since it goes beyond merely highlighting misconceptions held by learners: it advocates strategies that seek to explicitly identify and repair intuitive errors that may influence future learning.

**Impact of Refutation Text and Paraphrasing on Learning**

**Refutation Text**

In a study conducted by Hynd and Alvermann (1985), two groups of physics students were tested for their conceptual understanding of motion theory. Once misconceptions were elicited, refutation text was given to one group and non-refutation or expository text (which states the correct conception without highlighting errors in understanding) was given to the control group. Results revealed significant repair of misconceptions of the refutation text group when compared with non-refutation text group. The non-refutation text group’s misconceptions about motion theory remained mostly unrepaird.
Van den Broek and Kendeou (2008) reported two empirical studies where groups of students read refutation text and non-refutation text online. Students who read refutation text experienced positive conceptual change (based on pre and post test results), while the students given non-refutation text did not experience significant conceptual change. Another study investigated students’ experiences of refutation text and non-refutation text (Hynd, 2001). The students stated that they preferred refutation text because it coerced them into considering new knowledge that challenged their prior knowledge.

The efficacy of refutation text was validated by Van den Broek and Kendeou (2008) and Tippett (2010) who advanced the notion that refutation text can be regarded as expert knowledge because it provides both arguments (two contradictory sides of a proverbial coin) that lead to the learning of normative knowledge. In short, refutation text raises cognitive awareness of intuitive errors (Tippett, 2010).

Refutation text is more likely to highlight misconceptions because the normative and erroneous concept are presented simultaneously and close to each other (Van den Broek & Kendeou, 2008); that is, a learner reads about a misconception and the correct concept at the same time, hence cognitive conflict activates both instances of knowledge and causes a learner to re-consider prior knowledge.

It seems that refutation text could be a viable and practical instructional strategy option for repairing misconceptions. The next section considers other instructional strategies that also effect conceptual change, albeit with different results.
Paraphrasing

Paraphrasing is a strategy intended to improve learning by increasing understanding of text or any other form of information (Hagaman et al., 2012). As a learning strategy, paraphrasing has links to Wittrock's (1974) generative process model which posits that learning occurs more effectively when based on prior knowledge. Generative learning therefore places the learner at the heart of the learning process by letting the learner construct his or her own meaning. Generative learning is thus substantiated by strategies such as paraphrasing.

Paraphrasing and Refutation Text

Paraphrasing means to restate information in order to make it more understandable (Kletzien, 2009). It demonstrates how the person paraphrasing understands text or any other type of information. There are numerous ways paraphrasing can be accomplished: synonyms, antonyms, and phrasal verbs are some of the strategies that can be used to understand a piece of information clearer (Nakov & Hearst, 2013). For example, by the acronym DDL (Data Definition Language) can be paraphrased to language that defines data if the phrasal verb strategy is used. Other paraphrasing options such as using antonyms (Hagaman et al., 2012) are equally applicable.

Paraphrasing involves reading, thinking, understanding, pausing to consider what has been read, re-writing (preferably without looking at the source text), comparing the
source and paraphrased text, and finally citing the source (Hagaman et al., 2012; Nakov & Hearst, 2013). Ultimately, the main idea is to clarify what the source text is communicating, and then reproduce the intended meaning by rewriting it.

Refutation text is ideal for paraphrasing because interpreting it requires careful thought about what it is trying to communicate. To paraphrase the concept of schema, for instance, requires a reader to think about related concepts such as structure and units of organization (in the database design knowledge domain). Hence paraphrasing should result in a clearer, simpler, and more easily understood version of the source text.

**How Paraphrasing Works**

When learners paraphrase, they think about and use their own words based on their interpretation of the original text. Paraphrasing entails engaging with text at a deeper level than merely reading at the surface (Nakov & Hearst, 2013); that is, surface reading can be accomplished without necessarily seeking to thoroughly understand text in a meaningful way. When readers paraphrase they do more than just summarize text: they engage with prior knowledge and past experiences to produce a more personal version of the original text (Madnani & Dorr, 2010). It becomes personal because the paraphrase reflects the writer’s own way of expressing the original text. Consequently, paraphrasing encourages learners to be aware of their own level of understanding the source text. Lastly, by encouraging learners to gauge the extent of their understanding, paraphrasing
increases influences learning when they relate the source text to related knowledge they already possess.

In addition to paraphrasing, there are several other instructional and learning strategies that influence learning. One is direct learner involvement to enrich understanding of content (Piaget, 1964). Another strategy uses imagery to encourage formation and understanding of concepts beyond low-level, fact-based comprehension (Anderson & Kulhavy, 1972). Yet another strategy uses examples to clarify material that is considered difficult to comprehend, as well as prompt learners to re-use learned knowledge in different contexts (Booth et al., 2013). Moreover, providing feedback during instruction is one more approach that facilitates a continuous feedback loop between teacher and student (Clariana & Koul, 2006). Lastly, Merrill and Tennyson (1978) suggested that meaningful learning partly depends on using attributes (characteristics of concepts or knowledge elements) to establish and relate categories to one another.

From the approaches summarized, paraphrasing was considered for the main study.

Concept Learning and Misconceptions

Concept learning is an important cognitive activity of the learning sciences. The acquisition of concepts is integral to understanding how learning occurs, how misconceptions are formed, and ultimately how instruction can be tailored to repair
misconceptions and maximize the quality of learning outcomes. However, concept learning as an activity continues to face problems of lack of awareness and lack of alternative categories during instruction (Chi et al., 2012). This means: 1) instructors and learners may not be aware that learners’ conception of a particular knowledge area is error-prone, and 2) even when there is an awareness of misconceptions, it is often challenging to identify the process of transitioning to correct conceptualization. After all, a misconception is still a “conception” (a misconception in one setting is a correct conception in another setting) that is learned and that may sometimes persist alongside a correct conception. As a result of these problems, measuring concept learning—especially during instruction—presents its own set of challenges.

In their seminal work, *A Study of Thinking*, Bruner, Goodnow, and Austin (1986) defined concept learning as “the search for and listing of attributes that can be used to distinguish exemplars from non-exemplars of various categories” (p. 233). The attributes can be thought of as characteristics or traits that help define knowledge items—objects, events, or ideas—in part or as a whole. This process occurs at the abstract level as mental or conceptual entities, so it is a way of thinking. The process of identifying the characteristics of the knowledge items entails a deliberate effort of categorization which is made difficult by the occurrence of knowledge as interconnected and overlapping mental abstracts instead of individually identifiable knowledge objects (Bruner et al., 1986). The decision-making process is therefore in the form of identifying differences and similarities among networked mental entities or concepts that may be related to each other.
Concept Categorization

Learning is not only about acquiring factual knowledge, but it also involves the ability to categorize or classify knowledge elements so that they become part of a meaningful relationship with each other (Rumelhart & Ortony, 1976). Conceptualizing knowledge therefore includes categorization which is how a learner identifies and interprets information, ideas, and groups of knowledge to attain specific meaning. According to Tennyson, Woolley, and Merrill (1972), a concept is a cognitive structure that exists within or outside a domain of knowledge.

Concept formation is a systematic and formal process that leads to better comprehension and activation of long term memory structures (Atkinson, Derry, Renkl, & Wortham, 2000; Rosch, 1975). A study about categorizing birds (Wahlheim, Finn, & Jacoby, 2012) where the participants’ judgments of their own learning experiences (metacognition) was evaluated in terms of repetition and variability, concluded that lack of appreciation of related concepts adversely influenced holistic understanding of the main concept.

The concept identification process requires prior familiarization of some of the characteristics that partially identify the target concept. Once related pieces of information are adequately understood, learners are able to apply the acquired knowledge in various situations or contexts. At this level learners are able to generate and utilize perceptions (Curran & Schacter, 2001) to further their understanding of a particular knowledge domain. Regarding the notion of concepts, some studies posit that schema are
pivotal constructs that explain concept formation (Kumaran, Summerfield, Hassabis, & Maguire, 2009; Rosch, 1975).

While there have been different investigations about the categorization of ideas during concept learning, the conclusions drawn emphasize the important role of categorization in learning (Kumaran, Summerfield, Hassabis, & Maguire, 2009; Rittle-Johnson, Star, & Durkin, 2009). Furthermore, the notion of structure underscores how categorization ensures that the attributes identified and processed are an integral part of the main concept.

The interpretation of a hierarchical relationship in concept formation is important (Chi & Roscoe, 2002): the example of cobra -> poisonous snake -> snake -> reptiles -> living things as a hierarchy traces the attributes whose culmination is the formation of snake as a concept. But the relationship does not always have to be hierarchical: a lateral buildup from raw fact to attribute to concept is equally effective for concept attainment.

**Misconceptions and Learning**

Concept learning faces a problem of misconceptions or intuitive errors by learners as they attempt to repair false beliefs and link networks of information considered normative. These intuitive errors are partly due to lack of awareness (of the misconception itself) and lack of alternative concepts during instruction. Cognitive dissonance or conflict (Festinger, 1957) may or may not occur when the correct conception is presented even though it ‘conflicts’ with the misconception. According to
(Festinger, 1957), a set of knowledge elements in the mind is in dissonance if, based on their meaning and potential use as a combined entity, they do not agree with one another. For example, an intuitive error prevalent with some database designers is the interpretation and use of the commands delete, truncate, and drop (Kroenke & Auer, 2009; Ullman, Garcia-Molina, & Widom, 2001). While delete removes specified rows from a table, truncate removes all rows from a table, and drop removes a table from a database (Kroenke & Auer, 2009). These nuanced differences in the three commands is often missed or completely misunderstood by students and practitioners alike in the information science and technology field and result in misapplication and mistakes when designing databases. An entrenched misconception regarding these related commands can result in unplanned loss of data.

Misconceptions are also caused by a lack of alternative categories. Many learners categorize whales as fish instead of as mammals because, intuitively, whales do look like fish (Chi & Roscoe, 2002) and few mammals live in the ocean. This illustrates the view that misconceptions or intuitive ideas are persistent and therefore resistant to correction. While it is easy to see why a student would develop and cling to this view (that whales are fish), the more pressing issue is how can educational systems identify and repair such misconceptions before they become too firmly entrenched in learners’ minds.

With relevant instructional guidance and, more importantly, continuous assessment and measurement of knowledge at different levels of knowledge acquisition, more miscategorization problems can be minimized (Vasilyeva, Coley, & Muratore, 2009). The categorization skills garnered during early stages of learning ought to carry
learners through to the assimilation and learning of more complex knowledge, some of which is life-threatening if mishandled; for example, the formation of a wrong concept relating to a medical procedure can cost a patient’s life.

In order to identify and remove misconceptions, instructional strategies should target specific learning outcomes. To attempt to address misconceptions, a number of strategies have been studied with varying degrees of success. The next chapter considers several of these approaches.

Types of Knowledge and Learning Outcomes

The main study focused on the effects of conceptual change-oriented refutation text on different types of knowledge (learning outcomes) among college students studying information organization, with relational database design serving as the knowledge domain area tested.

The choice of instructional strategy directly influences achievement of specific learning outcomes (Trigwell & Prosser, 2011). A learning outcome is the fulfillment of specific knowledge or skill related to subject matter. Learning outcomes cover predetermined content or periods of time such as a week-long module, a semester course, or an entire program (Trigwell & Prosser, 2011). Examples of learning outcomes are declarative knowledge performance, conceptual knowledge performance, and procedural knowledge performance (Benbunan-Fich & Arbaugh, 2006; Trigwell & Prosser, 2011). Declarative or factual knowledge is defined as knowledge of terminology and specific
details, conceptual knowledge as knowledge of classifying concepts, and procedural knowledge as knowledge of skills and techniques applied to a specific area of knowledge (Clariana & Prestera, 2009; Jonassen, Beissner, & Yacci, 1993).

Because learning outcomes are directly linked to learners’ interaction with knowledge, it is crucial that the appropriate types of knowledge are targeted, particularly for the identification and removal of misconceptions. The following section reviews three knowledge types—declarative, conceptual, and procedural—that are integral components of learning outcomes and therefore important areas for misconception repair.

**Declarative Knowledge**

Declarative or factual knowledge is knowledge of terms that pertain to a knowledge domain (Andre & Ding, 1991). It is illustrated by learners’ ability to recall and describe what they know (Jonassen & Wang, 1993). A learner with effective declarative knowledge is also able to state properties or characteristics about subject-specific knowledge. According to Jonassen, Beissner, and Yacci (1993), declarative knowledge is as important to learning as conceptual and procedural knowledge are. Without the initial assimilation and processing of declarative knowledge, the learning of concepts would be difficult to achieve.
Conceptual Knowledge

Conceptual knowledge refers to the ability to categorize information elements (McCormick, 1997). Conceptual knowledge is an important learning outcome because it indicates learners’ grasp of classification, principles, and theories related to a particular knowledge domain. In addition, high levels of conceptual knowledge indicate an equally good grasp of facts and specific details (declarative knowledge). Conceptual knowledge is also necessary as a base upon which higher-order types of knowledge—procedural and meta-cognitive knowledge—can be established and executed (Hallett et al., 2010).

A more direct way of understanding knowledge classifications and categories is to understand how knowledge elements relate to each other based on similarity. The knowledge of how the elements are arranged is called structural knowledge (Jonassen et al., 1993). To understand how knowledge is arranged or related in a learner’s mind provides clues about how to best tailor specific instructional strategies to target specific learning outcomes.

Procedural Knowledge

The ability to apply factual and conceptual knowledge is an indicator of procedural knowledge. According to Grotzer (2002), Jonassen (1993), and McCormick (1997), procedural knowledge is both knowing techniques as well as knowing when to apply the correct technique to solve a problem. It is a test of skills for a specific
knowledge domain, and the ability to determine the appropriate criteria to apply to a given problem.

Procedural knowledge is therefore more abstract than conceptual and declarative knowledge because its elements are less specific, that is, it is not about displaying knowledge of specific terminology or principles. Together with declarative and conceptual knowledge, procedural knowledge has been identified as essential for long-term learning (Pellegrino & Hilton, 2012).

**Chapter Summary**

This chapter highlighted Posner et al.’s (1982) conceptual change theory as the foundation for the main study. Within this framework, alternative theories advanced to deal with misconceptions were compared.

The chapter also reviewed the potential impact of paraphrasing refutation text in order to effect conceptual change. In addition, problems such as lack of ways to measure misconceptions and lack of alternative approaches to repair them were reviewed. The nature of concepts, concept learning, and concept formation were also synthesized.

It was noted that unless misconceptions are identified and repaired as early as possible during the learning process, students may acquire more complex knowledge based on a foundation of misconceptions. The result may be a limited ability to define concepts, relate concepts, or apply knowledge in ways considered normative within each subject domain.
The next chapter reviews research methodologies as applied to the main study.
Chapter 3

RESEARCH METHODOLOGIES

Introduction

The main study analyzed the effects of conceptual-change oriented refutation text on students’ declarative knowledge performance, conceptual knowledge performance, and procedural knowledge performance. Within information sciences, the knowledge domain area tested was relational database design. It was conducted in two real classrooms with a mix of college freshmen and sophomores as participants. The study compared reading refutation text, reading and paraphrasing refutation text, and a control condition of not engaging with refutation text. Pilot studies helped refine the treatments.

Pilot Studies

Two pilot studies preceded the main study research. The purpose of both pilot studies was to refine the treatments by testing levels of understanding of database concepts among college students. Implications were that knowledge at one level does not automatically guarantee similar levels of knowledge at other levels, that is, students performed differently between tests that assessed different types of knowledge. Lastly, pilot work suggested that confronting intuitive errors more directly (beyond merely reading) may be required to influence conceptual change and repair the misconceptions.
Main study

Study Context

Test data were collected from 94 freshmen and sophomore college students (48 and 46 students in each class) pursuing an Information Sciences and Technology (IST) bachelor’s degree program at a large northeastern university. The students were enrolled in two sections of a 200-level college class, with different instructors teaching from the same syllabus. Based on the students’ indication of a range of number of semesters completed, some would have had well-entrenched misconceptions while others may have encountered database design concepts for the first time at the time the main study was conducted. Table 3-1 shows the number of semester each student had spent at college during the time the main study was conducted.

Table 3-1

<table>
<thead>
<tr>
<th>Number of College Semesters</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>11</td>
<td>17</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>27</td>
<td>41</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>66</td>
<td>100</td>
</tr>
</tbody>
</table>
As Table 3-1 indicates, students’ exposure to knowledge ranged from 2 to 11 semesters (inclusive). 41% of the students were on their fourth semester, 17% on their second semester, and 15% on their sixth semester. Very few students had spent more than six semesters at college.

The researcher recruited the students by approaching two instructors that taught the two sections of the class. Before requesting access to the students, the researcher checked the syllabus to ensure that the course prerequisites included the subject matter targeted for testing – relational database design.

**Research Design**

The main study employed a quantitative pre and posttest, experimental design. The independent variable (refutation text) was manipulated and different levels assigned to participants. This criterion, according to Tabachnick and Fidell (2013), distinguishes experimental research from correlational research and other study designs. From two intact classes, students were combined before each student was randomly assigned to one of four subgroups. Two subgroups engaged with refutation text treatment group one (RT group one), while the other two subgroups engaged with refutation text treatment group two (RT group two). All four student subgroups took the same pre and post-tests. The dependent variables, namely declarative knowledge performance, conceptual knowledge
performance, and procedural knowledge performance, were measured twice (before and after the refutation text treatment).

It sought to determine whether paraphrasing refutation text, as the intervention, influenced a group of students differently from another group of students. The design was regarded as a true experiment because the allocation of participants to groups was completely random (Kerlinger & Lee, 2000). Participant groups received different refutation text items; that is, participant groups received refutation text items from either RT group one or RT group two.

**Variables**

The variable landscape shown in Figure 3-1 has the independent variable—refutation text—split into three levels; read only, read and paraphrase, and a control level in the form of RT group one participants not engaging with RT group two’s refutation text, and vice versa. The dependent variables were declarative knowledge performance, conceptual knowledge performance, and procedural knowledge performance.

![Figure 3-1. Variables used in main study.](image-url)
Figure 3-1 also shows moderating variables, namely pretest declarative knowledge, pretest conceptual knowledge, and pretest procedural knowledge. The pretests were designed to assess prior knowledge that was expected to change or moderate the nature of the relationship between refutation text (independent variable) and the three dependent variables (Creswell, 2012). In short, it was anticipated that prior declarative, conceptual, and procedural knowledge was going to help determine whether or not engaging with refutation text affects students’ learning outcomes.

In addition to the variables shown in Figure 3-1, personal characteristics data (self-efficacy, critical thinking, and metacognitive self-regulation) that was expected to influence how refutation text affected declarative, conceptual, and procedural knowledge performance, was collected for potential use in subsequent studies. Time-on-task (total, not for individual sections of tests), gender, and class standing data were also collected for potential use in subsequent studies.

Lastly, learners who engaged with one set of refutation text (read or paraphrased) did not engage with the second set of refutation text. Specifically, RT group one participants did not read or paraphrase RT set two, and RT group two participants did not read or paraphrase RT set one.

Participants

Data were collected on two occasions: pretest and posttest. Prior to the pretest, two classes of 48 and 46 students per class were combined to form an initial set of 94
possible participants. The combined list of students was scrambled and randomly mixed using Microsoft Excel’s randomization function. After randomizing, the first quarter of the student list was labeled as RT group one, the second quarter was labeled as RT group two, the third quarter was labeled as group three, and the last quarter was labeled as group four. Each of the four groups therefore consisted of students from both classes. Moreover, each of the 4 groups was randomly assigned to one of the treatments; for example, RT group one was assigned to read refutation text without paraphrasing, whereas RT group two was assigned to read and paraphrase refutation text.

During data collection, 85 students submitted data during pretest followed by 76 students’ submissions during posttest. For inclusion in analysis, matching pretest and posttest data needed to be present; therefore, a student who participated in only one of the two tests had their submission discounted (see Table 3-1 summary of data submission). Based on this criterion, 66 students’ data remained valid for analysis after validation.

Figure 3-2 summarizes the randomized group assignments.

---

**Figure 3-2.** Participant groups and treatment levels.
Figure 3-2 shows two blocks, RT group one and RT group two, which were created to split the ten refutation text items into five items per group. Subsequently, data collected were analyzed within each group as well as by comparing across groups. RT group one therefore engaged with one set of five refutation text items while RT group two engaged with a different set of five refutation text items. The levels were reading refutation text versus reading and paraphrasing refutation text. In addition to the first two levels, RT treatment group one \( (N = 31) \) did not engage with RT treatment group two \( (N = 35) \). The reverse is also true.

Of the 66 students, groups created randomly ranged in size from 14 to 19 students per group. Table 3-2 shows the initial group sizes, the number of pretest and posttest submissions, plus valid and invalid submissions after data validation. Timeline information is also included. For data analysis, only data regarded as valid was considered.

Table 3-2

*Summary of Data Submission*

<table>
<thead>
<tr>
<th>RT group</th>
<th>Treatment Level</th>
<th>Initial group size</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Valid (after cleanup)</th>
<th>Invalid</th>
</tr>
</thead>
<tbody>
<tr>
<td>one</td>
<td>Read</td>
<td>23</td>
<td>19</td>
<td>20</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Read and Paraphrase</td>
<td>24</td>
<td>24</td>
<td>18</td>
<td>17</td>
<td>7</td>
</tr>
<tr>
<td>two</td>
<td>Read</td>
<td>24</td>
<td>22</td>
<td>19</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Read and Paraphrase</td>
<td>23</td>
<td>20</td>
<td>19</td>
<td>19</td>
<td>4</td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td>94</td>
<td>85</td>
<td>76</td>
<td>66</td>
<td>28</td>
</tr>
</tbody>
</table>

Text before and after data collection: Day 1 & 2, Day 8 & 9, After data collection.
As Table 3-2 illustrates, 28 potential participants were excluded during validation (mostly due to missing pre or post assessments). Several sanity checks were applied in determining valid versus invalid data. One of the main criteria was availability of both pretest and posttest data. If a participant’s data could not be matched, it was discarded since pre and post analysis was not possible. The reduction from 94 to 66 useable data points (see Table 3-2) was mainly due to this criterion. A further sanity check was time taken to complete the study. Of the 66 pre and post data points that remained, all were checked for time-to-task of under one minute. There was a solitary zero-minute time-to-task from one of the participants that would have otherwise met all other criteria for inclusion, so that was discarded. The remaining 66 data points passed the sanity check.

**Summary of Study Design**

The main study tested how conceptual change-oriented refutation text influenced declarative, conceptual, and procedural knowledge. A possible cause and effect relationship between refutation text and learning outcomes was investigated.

The refutation text treatment was split into read only versus read and paraphrase. In addition, RT group one participants did not read or paraphrase RT set two, and RT group two participants did not read or paraphrase RT set one. Each of the initial 94 students was randomly assigned to one of 4 groups, after combining students from two classes. The classes were combined in order to reduce variability (increase equivalence) that may have existed among the students (Kerlinger & Lee, 2000) due to factors such as
different class times, instructors’ teaching styles, and students’ personal reasons for class choice. Each of the participants’ outcome measures included declarative, conceptual, and procedural knowledge performance test scores. The final data analysis was based on 66 students after data cleaning.

Declarative knowledge performance, conceptual knowledge performance, and procedural knowledge performance questions (tests of learning outcomes) were aligned with the refutation text items. The process of aligning the refutation text items to the tests is explained in the section “Alignment of Refutation Text with Test Questions”.

To minimize cognitive complexity (scrambling the mind) while engaging with refutation text, the refutation text items were kept as few and brief as possible. The refutation text treatment was split into two groups for the following reasons: one, allocating five instead of ten refutation text items to a student was intended to limit excessive cognitive conflict (scrambling the mind); and two, comparing two studies also implied a more meaningful generalizability check because results observed from students engaging with only five refutation text items would not be as generalizable. Engaging with only one treatment set would implicitly imply that engaging with two different sets of refutation text treatment achieved the same results, which may not necessarily be the case. The “cost” of that approach was that each of the posttests aligned only with half of the refutation text items.

In place of analyzing individual students’ misconceptions, and allocating refutation text items only to those confirmed as holding misconceptions, a “shotgun” approach was applied: all students were given refutation text items. The main reason for
this was efficiency: instead of extra work that an instructor would be required to conduct to determine which individual students required refutation text and which ones did not, the main study’s approach meant any instructor could apply this today.

**Materials**

**Sources of Data**

A study by Guzmán, Conejo, and Gálvez (2010) concluded that data organization concepts such as entities, relationships, keys (especially the primary key) are often misunderstood by learners. These concepts were added to a list for testing. Suraweera and Mitrovic (2002), in a study that tested participants’ database design skills, showed that some students’ conception of the Entity-Relationship (ER) data model and the meaning of database schema was faulty or not in line with normative standards. E-R modelling and database schema were added to a list for testing.

After reviewing studies that highlight misconceptions in the information organization knowledge domain (see previous steps), and after applying an approach used by Guzmán et al. (2010) in a study about eliciting misconceptions in a programming course, more concepts were identified in a 300-level “Information and Organizations” course: primary key, entity, and attribute. These concepts were required prior knowledge for a case study about information organization.

The next set of concepts were identified and sourced from a pre-requisite course (required for the first pilot study participants’ class) that taught data organization. From
this class the following concepts were cross-checked with previous studies (described in step 1) and added to the list: logical modeling, logical schema, physical modeling, view, normalization, and constraints.

Finally, the concepts identified were cross-checked with prescribed college reference material such as Kroenke and Auer’s (2009) *Database Concepts, 4th Edition* and Ullman et al.’s (2001) *Database Systems: The Complete Book*.

**Refutation Text**

The refutation text treatment was designed to not only attempt to point out learners’ misconceptions about relational database design, but to also refute the misconceptions. The treatment attempted to achieve that by focusing on concepts that are commonly misunderstood in the database design domain (Kroenke & Auer, 2009; Ullman et al., 2001). Refutation text was given to participants as a response to misconceptions that they likely may hold.

Figure 3-3 displays refutation text designed to address a misconception in relational database design (See Appendix A for the complete set of refutation text items that were delivered to the participants). There were a total of ten refutation text items. A refutation text item similar to Figure 3-3 was shown to each learner one at a time (on a computer monitor).
The first part of refutation text (see Figure 3-3) identifies the potential misconception. Text on the first section of the refutation text items used in the main study averages 20 words. The second part is designed to prepare the learner for a correction of the possible misconception. A prompt (middle part in Figure 3-3) indicates that “this false belief must be resolved in the students’ minds”. Reading the prompt is expected to mentally prepare one to assimilate or accommodate the knowledge that refutes existing knowledge. The prompt is deliberately briefly worded to emphasize a possible challenge to the learner’s main study set of beliefs about a particular knowledge area. Text on the third section of the refutation text items used in the main study averages 4 words. Main study beliefs about subject matter are typically garnered via pretesting, as was the case in the main study.

The third and final part of the refutation text treatment (see Figure 3-3) presents correct (normative) conceptions or knowledge that attempts to refute existing misconceptions. Text on the third section of the refutation text items used in the main study averages 40 words.
Refutation text was designed to repair misconceptions via concisely worded refuting statements. Refutation text wording is not gentle in its approach; it is deliberately presented to learners in a declarative and authoritative format (Broughton et al., 2010). For example, in the main study learners were presented with refutation text that challenged misconceptions about physical database modeling (when compared with logical database modeling). The refutation text highlighted that physical modeling depends on the specific relational database management system; whereas, logical modeling is less dependent on any actual system. If the database designer previously believed that logical and physical modeling processes were interchangeable, reading refutation text challenged them to consider the normative version. This was tested by assessing how test performances of low-scoring students (more misconceptions) and high-scoring students (fewer misconceptions) changed after engaging with refutation text (pretest-posttest differences).

For text to be regarded as having the ability to refute misconceptions and therefore encourage conceptual change, three clearly distinguishable parts should be present (Tippett, 2010). One part is the non-refutation or expository part, the second uses phrases such as but that is not true to indicate change of direction, and the third part explicitly states the accepted view (normative) of the concept. Refutation text sequentially advances through a process of pointing out the likely misconception, declaring the misconception’s state (that it is an area that must be addressed), and finally providing the knowledge considered normative within the information systems field. Since the tests were designed to assess different types of knowledge about relational
database design, the refutation text treatment was also designed to target the most critical areas where misunderstanding would likely hinder assimilation or accommodation of more complex knowledge and its application (Elmadani, Mathews, & Mitrovic, 2012; Guzmán et al., 2010).

Since main study research is uncertain why the refutation text structure in particular seems to modify conceptual change better than non-refutation text (Van den Broek & Kendeou, 2008), the refutation text designed for the main study followed the same text flow pattern simply to maintain consistency.

**Mapping of Refutation Text to Test Items**

Every concept addressed by the ten refutation text items was tested. More importantly, each concept assessed in the declarative knowledge test was also assessed in the conceptual knowledge test. For example, both tests contained test items about the concept of keys. That made it possible to draw conclusions regarding students’ knowledge of concepts between the two different types of knowledge.

However, not all concepts were included in the single procedural knowledge test question. More specifically, the question incorporated six of the ten refutation text items (refer to Table 3-2). In short, comparison involving procedural knowledge performance is generalizable only to the extent of the concepts tested.
For further clarification, Table 3-3 shows the main components of the main study’s design. The two refutation text groups, randomly organized participant groups, corresponding refutation text items, and corresponding test items are shown.

Table 3-3

*Participant Groups, Concepts, and Tests*

<table>
<thead>
<tr>
<th>RT Group</th>
<th>Randomized Student Groups</th>
<th>Refutation Text Items</th>
<th>Declarative Knowledge (DK)</th>
<th>Conceptual Knowledge (CK)</th>
<th>Procedural Knowledge (PK)</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td></td>
<td>RT group one:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Read</td>
<td>RT_1_Entity-Rtp</td>
<td>DK_1_LogVisPhys</td>
<td>CK_1_Entity-Rtp</td>
<td>PK_Declare a relation</td>
</tr>
<tr>
<td></td>
<td>Read and Paraphrase</td>
<td>RT_3_Normalization</td>
<td>DK_2_Normalization</td>
<td>CK_2_Schema</td>
<td>(entity-Rtp,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RT_5_Name_Storage</td>
<td>DK_3_Entity-Rtp</td>
<td>CK_3_Key</td>
<td>normalization,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RT_7_Table_Row</td>
<td>DK_4_Key</td>
<td>CK_4_Table</td>
<td>schema, logical vs.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RT_9_Schema</td>
<td>DK_5_RdbSys (storage, key)</td>
<td>CK_5_Normalization</td>
<td>physical, key,</td>
</tr>
<tr>
<td>Two</td>
<td></td>
<td>RT group two:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Read</td>
<td>RT_2_LogVisPhys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Read and Paraphrase</td>
<td>RT_4_CompKeys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RT_6_PrimaryKey</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RT_8_Null</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RT_10_SQLcommand</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3-3 shows how the refutation text treatment items are mapped to each group and the participant groups. For instance, from the RT group one, the randomized group of participants labelled “Read” was tasked with reading refutation text items one, three, five, seven, and nine before answering all test questions. Similarly, from the RT group two, the randomized group of participants labelled “Read and Paraphrase” engaged with refutation text items two, four, six, eight, and ten before answering all test questions. All four groups took the same tests before and after engaging with varying levels of refutation text. To further clarify the alignment between refutation text items and posttest
items, Figure 3-4 uses arrows to show the mapping between each item from the refutation text groups (one and two).

Figure 3-4 shows how the two refutation text groups mapped to the items which, in turn, mapped to each of the test items. Incorrect responses from tests were then inferred as misconceptions. As Table 3-3 illustrates, all refutation text items are mapped to a test item in each of the three tests (DK, CK, and PK). The only exception is that only
six out of the ten refutation text items were tested in the open-ended procedural knowledge question: it would have been a stretch to test ten concepts in one question.

The tests assessed all concepts refuted in the treatment. Allocating five instead of all ten refutation text items to each student was intended to limit excessive cognitive conflict (scrambling the mind). It also enabled a more meaningful generalizability check; that is, results observed from students engaging with only five refutation text items would not be as generalizable as results from two groups of students. The results would implicitly imply that engaging with two different sets of refutation text treatment achieved the same results, which may not necessarily be the case. The “cost” of that approach was the each of the posttests aligned only with half of the refutation text items.

The next section reviews measurement instruments used in the main study.

Measurement Instruments

Three instruments were designed to measure three learning outcomes: declarative knowledge performance was measured with a multiple-choice test, conceptual knowledge performance was measured with a matching pairs test, and procedural knowledge performance was measured with an open-ended problem test. Examples of specific activities involved in these learning outcomes included the following: knowing the definition of a database candidate key (declarative knowledge outcome), relating the concepts table and entity (conceptual knowledge outcome), and applying knowledge of a primary key together with knowledge of relating key and unique to declare a relation
(procedural knowledge outcome). These are examples of concepts tested across all three knowledge tests.

The criterion measures consisted of pretest and similar posttests. The posttests were equivalent (not identical) to the pretests. For the multiple-choice and open-ended questions, sentence structures were slightly altered (subject and verb reversed). For the conceptual (concept relatedness) test, some posttest concepts were replaced with concepts that meant the same thing within the database design knowledge domain; for example, the database concepts table, entity, and relation are used interchangeably (Kroenke & Auer, 2009). Maintaining equivalence was designed to decrease memorization between pre and post-testing. Each of the instruments is reviewed in the next section.

**Measure of Declarative Knowledge**

Five (5) multiple choice questions called Type X or Multiple True/False were used to measure declarative knowledge. Each of the questions was essentially four true/false questions, that is, each multiple choice question was one question with four parts.

A declarative knowledge pretest was designed to measure learners’ knowledge of specific facts and terminology relating to relational database design. Research posits that knowledge of facts is pivotal to learners’ ability to develop schema and use their organization of knowledge to successfully complete tasks such as solving problems (McVee, Dunsmore, & Gavelek, 2005; Son et al., 2011). Figure 3-5 shows one of the
declarative knowledge questions developed for the main study (see Appendix B for all 5 declarative knowledge questions).

Figure 3-5. A declarative knowledge question.

From the multiple-choice declarative knowledge question shown in Figure 3-5, an incorrect response may suggest one or both of the following:

- A belief that keys, constraints, indexes, etc., are created when designing logical database models instead of when designing physical database models. This belief may result from a misconception about the difference between logical and physical database design.
- An understanding that one can create views during logical modeling and visualizes business activities and flow of data during physical modeling. Again, this choice may
result from a misconception about what logical database design entails when compared with physical database design.

Both the above statements are examples of possible misconceptions. Correcting misconceptions is important for comprehension of more complex relational database design subject matter.

The declarative knowledge test items were piloted in both the first and second pilots. Cronbach’s alpha reliability coefficient for the declarative knowledge performance posttest was .654 in the first pilot study and .725 in the second pilot study. Cronbach’s Alpha for the main study’s declarative knowledge instrument was .453 (refer to the “Reliability and Validity” section for the main study’s reliability results for all tests). The low reliability for this particular instrument suggests room for improvement; for example, increase the number of questions. Efforts to improve reliability after pilots included allocating different levels of treatment to all participants in order to minimize the impact of small samples.

Measure of Conceptual Knowledge

Ten (10) pairs of concepts were used to measure conceptual knowledge. Each concept uniquely matched one other item such that once a match was made for a particular concept, that concept would not correctly match any other concept. Marker and Clariana (2007) stated that concept relatedness or knowledge structure illustrates how
closely matched two or more knowledge items are to each other (e.g. table, entity, and relation clustered to indicate they are conceptually similar), and that the ability to make these connections is essential for tangential uses of knowledge such as applying subject-specific techniques to solve problems. Figure 3-5 displays a sample of pairs of concepts that participants were requested to match (see Appendix C for the complete test). It illustrates a conceptual knowledge test that was designed to elicit concept relatedness from learners. Similar to Marker and Clariana's (2007) study seeking to elicit concept relatedness, participants in the proposed study were asked to identify concepts that related the most to each other, and then select a radio button at the intersection of the concepts they regarded as most similar in meaning and/or function. They were further told to expect to make ten comparisons. Moreover, no concept matched with more than one other concept.

So, learners who understood the concept of table were be able to click the radio button representing the intersection of table and relation.
Figure 3-6. Screenshot of a partially completed conceptual knowledge test.

With reference to Figure 3-6, a learner whose intuitive grasp of a database table is incorrect would have struggled to relate the concepts table and relation. Figure 3-6 shows five out of ten pairs of concepts correctly matched. That was indication of a learner’s knowledge structure pertaining to each set of concepts.

The concepts were directly related to the refutation text that was given to participants. For example, concepts that required matching of table to relation (or to entity in the equivalent test) were contained in refutation text item number seven that exposed and refuted misconceptions about the concept of table. The pairs of concepts were similarly linked to both the declarative knowledge and procedural knowledge tests. The first and second pilot study’s Cronbach’s alpha reliabilities for conceptual knowledge performance were .555 and .692 respectively. The reliability score (Cronbach’s Alpha) for the conceptual knowledge instrument was .604 (refer to the
“Reliability and Validity” section for the main study’s reliability results for all tests).

According to Creswell (2012), the alpha for conceptual knowledge performance scores is considered acceptable for purposes of internal consistency since it exceeds .6.

**Measure of Procedural Knowledge**

One open-ended question was used to measure procedural knowledge. The question required knowledge of a combination of different but related concepts. Knowledge required to answer the question needed to be drawn from both declarative and conceptual knowledge. The procedural knowledge test measured how learners applied subject-specific skills or techniques to solve a problem. Figure 3-7 illustrates a procedural or application-level open-ended problem that the students were asked to solve.

*Figure 3-7. Screenshot of the procedural knowledge test*
In the procedural knowledge test (see Figure 3-7), the participants were asked to solve the given problem. As per a rubric (see Table 3-4) used by Domin and Bodner (2012), the participants’ responses were scored according to the following guidelines:

Table 3-4
Rubric for Scoring Procedural Knowledge Test

<table>
<thead>
<tr>
<th>Performance level</th>
<th>Performance</th>
<th>Points Allocation</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No attempt was made to solve the problem.</td>
<td>0</td>
<td>whose primary key is name and which has a foreign key that references presC# of relation</td>
</tr>
<tr>
<td>2</td>
<td>Attempts were made to solve the problem, but there is no demonstrable understanding of the relevant concepts.</td>
<td>2</td>
<td>name is FilmStudio’s primary key and FilmStudio is using foreign key presC# that references cert# of</td>
</tr>
<tr>
<td>3</td>
<td>Knowledge of the relevant concepts was demonstrated, but they were not integrated to solve the problem.</td>
<td>4</td>
<td>CREATE TABLE FILM STUDIO (: Name char(25) Not Null, / City char(25) Not Null, / presC# char(25) Not Null, / Constraint Name_PK Primary Key );</td>
</tr>
<tr>
<td>4</td>
<td>Problem solved incorrectly, but an understanding of the necessary concepts was demonstrated.</td>
<td>6</td>
<td>CREATE TABLE FilmStudio ( / name Char(25) PRIMARY KEY, / city Char(25) NOT NULL / presC# Int NOT NULL / add constraint FK _presC# references MovieExec / )</td>
</tr>
<tr>
<td>5</td>
<td>Problem solved correctly.</td>
<td>8</td>
<td>CREATE TABLE FilmStudio (name CHAR(30) PRIMARY KEY, address VARCHAR (255), presC# INT REFERENCES MovieExec(cert#)); Or CREATE TABLE FilmStudio (name CHAR(30) PRIMARY KEY, address VARCHAR (255), presC# INT, FOREIGN KEY (presC#) REFERENCES MovieExec(cert#));</td>
</tr>
</tbody>
</table>

In the rubric with examples (see Table 3-4), demonstrable understanding of the relevant concepts means learners illustrated specific techniques (procedural knowledge) required to solve the given problem. For example, attempting the problem shown in Figure 3-7 without referencing the stated foreign key does not demonstrate sufficient grasp of the ability to apply the concept of keys to solve an entity declaration problem. Similarly, to integrate relevant concepts to solve an open-ended problem such as the one shown in Figure 3-7 means a learner should know and apply several concepts, namely entity-relationships, normalization, schema, logical vs. physical design, keys, and the use of SQL syntax. Integrating these concepts earns a learner up to 8 points. Equal interval
scoring was applied to the procedural knowledge responses; that is, a student’s response was allocated points ranging from 0 to 8. Both the pretest and posttest procedural knowledge tests were graded with the same rubric (see Table 3-4). The reliability score (inter-rater) for the procedural knowledge instrument was .923 (refer to the “Reliability and Validity” section for the main study’s reliability results for all tests). According to Creswell (2012), the alpha for procedural knowledge performance scores is considered acceptable for purposes of internal consistency since it exceeds .6.

**Procedures after intervention**

After engaging with the refutation text, participants were asked if they had read the refutation text delivered to them. A Yes-or-No option was provided. Participants’ names, student IDs, and instructor names were requested for purposes of awarding extra credit points. Personal details (students’ usernames, full names, and email addresses) were never used for anything other than to ensure the correct students were awarded extra credit points. Pseudonyms were created, matched to the last four digits of students’ phone numbers across pretests and posttests, and then personal data were deleted from all further analysis.

The next day, the researcher went to the second class of participants and administered the same process as described above. After one week, the researcher returned to the first class of students for post-testing. Procedures similar to day one as described above were applied. On the last day of the procedure, the researcher returned
to the second class of students for post testing, using the same process as described above for the first day.

**Reliability and Validity**

**Reliability**

Results of reliability analyses for the main study’s declarative knowledge, conceptual knowledge, and procedural knowledge tests are shown in Table 3-5.

Table 3-5

*Reliability Analyses*

<table>
<thead>
<tr>
<th>Type of reliability</th>
<th>Instrument</th>
<th>No. of times instrument administered</th>
<th>No. of different versions of instrument</th>
<th>No. of students</th>
<th>No. of Items</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal Consistency</td>
<td>DK Test</td>
<td>Pre-post</td>
<td>Two (equivalent)</td>
<td>66</td>
<td>16 (multi options, multiple choice)</td>
<td>.453</td>
</tr>
<tr>
<td>(Cronbach’s Alpha)</td>
<td>CK Test</td>
<td>Pre-post</td>
<td>Two (equivalent)</td>
<td>66</td>
<td>9 (matching pairs)</td>
<td>.604</td>
</tr>
<tr>
<td>Inter-rater</td>
<td>PK test</td>
<td>Pre-post</td>
<td>Two (equivalent)</td>
<td>66</td>
<td>2 (open-ended question)</td>
<td>.923</td>
</tr>
</tbody>
</table>

Note: Some table headings adopted from Creswell (2009)

As Table 3-5 shows, procedural knowledge scores garnered high inter-rater reliability of .923, followed by conceptual knowledge scores with internal consistency reliability of .604, and finally declarative knowledge scores with internal consistency reliability of .453. According to Creswell (2012), the alphas for conceptual knowledge and procedural knowledge performance scores are considered acceptable since they exceed .6, while the alpha for declarative knowledge performance scores is not
acceptable for purposes of internal consistency. For the open-ended procedural knowledge tests, two raters independently scored both the pretest and posttest before inter-rater reliability was calculated. Post test scores were used for determining the main study’s inter-rater reliability.

According to Kerlinger and Lee (2000), the instruments’ reliability can be increased by adding questions that assess the same concepts. Issues of trade-off between covering as much knowledge base and administering longer tests during limited class time are reviewed in the discussion and interpretation chapter.

Validity

Several types of validity applied to the main study. Kerlinger and Lee (2000) highlight the most important: content, criterion-related, and construct validity.

For content validity, subject matter experts were asked to examine the themes and concepts that the researcher intended to present to students for testing. The experts specifically commented on the relevance of the relational database design concepts to their students’ background knowledge.

A distinguished professor of educational psychology (H. Suen, personal communication, March 18, 2013) validated test format and metrics by reviewing the declarative, conceptual, and procedural knowledge test questions. In his analyses of the multiple choice declarative knowledge questions, he stated that “this type of question is called either Type X or Multiple True/False. There are many different possible ways of
scoring such an item. It is essentially four true/false questions. You need to determine whether you wish to treat it as four questions or one question with four parts” (H. Suen, personal communication, March 18, 2013). Professor Suen also validated the conceptual knowledge test. Lastly, Professor Suen provided face validation by commenting that the “overall assessment looks good to me” (H. Suen, personal communication, March 18, 2013).

Another type of validity applicable to the main study was ecological: learners were two classes consisting of approximately the same number of learners that can be considered to be representative of the entry level students of the selected content. In addition, the two instructors gave the researcher permission to conduct the study because the content to be tested aligned with students’ expected prior knowledge (from prerequisite courses).

**Procedures**

**Overall data collection procedure**

Prior to commencing the collection of data, the level of significance was set at .05, which is the “happy medium” between Type I and Type II errors (Huck, Cormier, & Bounds, 1974): there was no compelling reason to set the alpha level at a more or less lenient level than .05 (Tabachnick & Fidell, 2013).
The procedure consisted of a pretest that spanned two days (data collection in classes held on different days), engagement with refutation text treatment, and a posttest equivalent to the pretest that also spanned a period of two days.

Data were collected on four different occasions: learners answered pretests on day one and day two (half of the sample between the two classes) and concluded on day eight and day nine with posttests. Pretesting lasted longer since it included once-only collection of demographic data (gender and class standing). Different levels of refutation text were given to participants after pre-testing. Students in RT group one received different refutation text items from students in RT group two. Specifically, RT group one participants did not read or paraphrase RT set two, and RT group two participants did not read or paraphrase RT set one. The baseline treatment was reading refutation text. From this baseline, some students were requested to paraphrase the refutation text items.

Data were collected from all participants at pretest and posttest. Unlike the pretests, the posttests consisted only of declarative, conceptual, and procedural knowledge performance tests (minus requests for demographic, motivation, and learning strategies data which was collected during pretesting). Appendix E contains a detailed, step by step description of data collection events.

Figure 3-8 shows how the research proceeded, from identifying and collecting instrument data, to executing two pilot studies and finally administering the main study.
The study’s research procedures are described in the following sections:

Procedures prior to intervention

Obtain IRB Approval

The Penn State Institutional Review Board (IRB) first granted ethics approval for data collection and publication in mid-October 2012 for the first pilot study. The approval applied to future related studies conducted by the same researcher. IRB approval was renewed in early spring 2013 for the second pilot as well as the main study.
Obtain Consent from Participants

Before participating in the two pilot studies and the main study, rules relating to privacy were explained to potential participants. Before beginning each data collection session, the researcher presented a one-slide summary of the rules governing participation. The researcher emphasized that participation was completely voluntary. Students indicated consent after reading the informed consent form (see Appendix D) by proceeding with the research study exercise.

If students chose not to participate, they were asked to email the researcher (email address provided in informed consent form) in order to be given an alternative task that also earned them extra credit points. The alternative task was to summarize, in one page or more, a database design topic of their own choosing. None of the students took up this alternative offer.

Adhere to Ethical Guidelines

Throughout the duration of the main study studies, ethical research guidelines were adhered to. That included readiness to resolve issues or problems in a manner that aligned with Penn State Institutional Review Board’s mandate. Fortunately, no such issues were encountered before, during, or after the data collection period.

A potential issue in research involving human participants is the Hawthorne Effect. The Hawthorne Effect is a phenomenon about research participants changing their normal behavior as a reaction to awareness of being studied (Kerlinger & Lee,
To minimize it, the researcher emphasized anonymity of data by telling participants exactly how personally identifiable data would be discarded, and that their instructors would never have access to their individual test responses.

Adhering to IRB rules also included emphasizing participants’ privacy by describing exactly how personal data would be handled, and which aspects of their responses would be analyzed and publicized. The researcher emphasized indifference to participants’ personally identifying data.

The researcher reminded potential participants what they stood to gain from participating: an opportunity to revise their understanding of database design aspects that they may not have been thinking about, or that were not clear to them.

To put potential participants at ease, the researcher minimized formal communication to potential participants. The researcher presented himself as one of the students, and not as someone above them by virtue of being a doctoral student. Lastly, the researcher dressed semi-formal instead of formal when appearing in person on data collection days.

**Procedures during intervention**

After completing the declarative knowledge, conceptual knowledge, and procedural knowledge pretests, refutation test items were presented to participants’ computer screens.
The screens showed the first of five refutation texts for each participant. Because engagement with refutation texts was different, the hyperlinks sent to the participants took them to 4 separate online testing areas within the main study’s test area in the Qualtrics system. Participants were then asked to either read each of the refutation text items without paraphrasing, or read and paraphrase each of the items presented on their screens. For the participants required to paraphrase, a multi-line free text field was presented below the refutation text item on the same computer screen. In addition to ensuring that students actually read the refutation text (order to paraphrase), the free text was collected for potential use in subsequent studies.

Data Analyses

Statistics applied to research questions

For research questions one, two, and three, MANOVA (and follow-up ANOVAs) was applied. Based on the four hypotheses developed for the main study (refer to the “Research Hypotheses” section) regarding reading refutation text and reading and paraphrasing refutation text in terms of declarative knowledge performance, conceptual knowledge performance, and procedural knowledge performance, mean differences were reviewed first to obtain an overall impression of direction and magnitude of scores.

For research question four, a mixed within-between subjects statistical analysis (also called split-plot or mixed factorial ANOVA) was applied (Kerlinger & Lee, 2000).
This design combines elements of repeated measures and between-subjects effects
(Kerlinger & Lee, 2000).

A median split was used to create a low score group and a high score group. 31
students from RT group one and 35 students from RT group two were placed in a low
scoring or high scoring group based on their pre-test scores. Since there were three
different tests, a median split was done for each kind of test; that is, a median split was
done for declarative knowledge, conceptual knowledge, and procedural knowledge. The
analysis incorporates within-subjects testing (pre and post comparison) and between-
subjects testing (low score group vs. high score group). In short, the within-subjects
variables are the pretest-posttest changes while the between-subjects factor is the
low/high median split. Since two sets of refutation text treatment (five items per set)
were allocated to different groups of students, separate analyses were conducted for
refutation text group one and refutation text group two.

To answer all research questions, pre-analyses included verification of
assumptions. The remainder of this section describes important assumptions tested as
part of the data analysis that follows in the next chapter.

**Assumptions Testing**

To partly answer each of the research questions, the following assumptions and
checks were applied as part of preparing for MANOVA analysis (Creswell, 2012; Huck
et. al., 1974; Kerlinger & Lee, 2000; Tabachnick & Fidell, 2013):
• Confirmed that the students were randomly assigned to each of the groups.
• Confirmed that all three dependent variables were measured as interval or ratio level data.
• Confirmed that no student appeared in more than one study group.
• Examined the data for accuracy. For example, each of the multiple-choice declarative knowledge questions contained more than one correct answer; hence the responses required recoding (this was mostly due to the way the Qualtrics survey software records responses).
• Tested each dependent variable for univariate normality. In addition to analyzing skewness and kurtosis, visual inspection of Normal Q-Q Plots and histograms were conducted.
• Confirmed that the levels of refutation text, namely reading only, reading and paraphrasing, and control, were categorical variables.
• Tested the assumption that sample variances were equivalent among all groups.

Chapter Summary

In the main study, two classes of 48 and 46 students per class were combined to form one group of 94 students. The students were enrolled in two sections of a 200-level college class, with different instructors teaching from the same syllabus. The researcher recruited the students by approaching two instructors that taught the two sections of the class. The test items assessed facts and definitions (declarative knowledge), concept
relatedness (conceptual or structural knowledge), and application of relational database
design knowledge (procedural knowledge). The study employed a quantitative pre- and
posttest, experimental design. From two intact classes, students were combined before
each student was randomly assigned to one of four subgroups. Two subgroups
participated in RT group one, and the other two subgroups participated in RT group two.
The dependent variables, namely declarative knowledge performance, conceptual
knowledge performance, and procedural knowledge performance, were measured once.

The next chapter provides an analysis of the data collected during a ten-day period.
Chapter 4

RESULTS

There are three main sections in this chapter: the first part describes frequencies and correlation coefficients, the second part presents statistical results for each of the four research questions, and the third part summarizes the statistical findings. Unless stated otherwise, the alpha level for main effects and interactions is set at .05 (Huck et. al., 1974). Furthermore, all statistical calculations include effect sizes. Data analysis was conducted using SPSS version 21.0.

Study Results

The next section presents overall descriptive statistics, checks assumptions, sets up data analyses, and then presents descriptive and main statistical results for each research question.

Summary of demographics and group composition

A summary of demographics and composition of groups per research question are presented in Table 4-1. Participants \(N = 66\) included 52 (79%) males and 14 (21%) females. RT group one participants had attended an average of 4.77 semesters \((SD = 1.73)\) while RT group two participants had attended an average of 4.11 semesters \((SD = \)
Lastly, RT group one participants spent an average of 24.94 minutes on the pretest and 10.74 minutes on the posttest, compared with RT group two participants who spent an average of 20.16 minutes on the pretest and 11.34 minutes on the posttest.

Table 4-1

| Summary of Demographics and Composition of Groups per Research Question |

<table>
<thead>
<tr>
<th>Variable</th>
<th>RT group one</th>
<th></th>
<th>%</th>
<th>RT group two</th>
<th></th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>24</td>
<td>77</td>
<td></td>
<td>28</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>7</td>
<td>23</td>
<td></td>
<td>7</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Read only vs. read &amp; paraphrase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Read only</td>
<td>14</td>
<td>45</td>
<td></td>
<td>16</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>Read and paraphrase</td>
<td>17</td>
<td>55</td>
<td></td>
<td>19</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>No reading vs. read and paraphrase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No reading</td>
<td>17 (RT group one)</td>
<td></td>
<td>19 (RT group two)</td>
<td></td>
<td>19 (RT group two)</td>
<td></td>
</tr>
<tr>
<td>Read and paraphrase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>14 (RT group one)</td>
<td></td>
<td>16 (RT group two)</td>
<td></td>
<td>16 (RT group two)</td>
<td></td>
</tr>
<tr>
<td>RQ 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DK low-score vs. high-score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low DK</td>
<td>16</td>
<td>52</td>
<td></td>
<td>23</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td>High DK</td>
<td>15</td>
<td>48</td>
<td></td>
<td>12</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>CK low-score vs. high-score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low CK</td>
<td>25</td>
<td>81</td>
<td></td>
<td>24</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>High CK</td>
<td>6</td>
<td>19</td>
<td></td>
<td>11</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>PK low-score vs. high-score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low PK</td>
<td>15</td>
<td>48</td>
<td></td>
<td>22</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>High PK</td>
<td>16</td>
<td>52</td>
<td></td>
<td>13</td>
<td>37</td>
<td></td>
</tr>
</tbody>
</table>

Note. Total percentages may not sum to 100% due to rounding error.

To answer the research questions, data in each group (one and two) were considered from various perspectives as shown in Table 4-1: read only vs. read and paraphrase, RT group one students who did not read and paraphrase RT set two vs. RT group two students who did not read and paraphrase RT set one, RT group two students who did not receive RT set one vs. RT group one students who did not receive RT set two, and low scores vs. high scores in each test.
Furthermore, before correlation and subsequent statistical analysis could be conducted, the data were assessed for normality. Results from Kolmogorov-Smirnov tests indicated that all declarative knowledge, conceptual knowledge, and procedural knowledge distributions for RT group one and RT group two were normal (significance levels exceeded .05), hence null hypotheses were retained (Huck et. al., 1974) and correlation coefficients were calculated next.

In addition, time-on-task for RT group one ranged from 11 to 89 minutes during pretest, and 4 to 19 minutes during posttest, while RT group two time-on-task ranged from 9 to 42 minutes during pretest, and 3 to 26 minutes during posttest. Time-on-task for individual parts of the activities (e.g. time taken to write one paraphrase) was not recorded.

Analysis of qualitative responses

Five students in RT group one and nine students in RT group two answered “No” when asked if they had read the refutation text, yet they wrote paraphrases. It seems they may have misunderstood the “did you read the refutation text?” question. Slightly more students in the read-and-paraphrase groups than the read-only groups answered “no” to this question. Since the requested paraphrases were written, their data were included for analysis.

In addition, the participants who paraphrased seemed to have taken it seriously: a check of the typed paraphrases did not indicate anything to the contrary. Subsequently,
no data point was excluded due to a student not taking it seriously. Moreover, the typed-in paraphrases ranged in length from 15 to 55 words. The paraphrases ranged from wording close to the original text, to clear attempts at interpreting the refutation text entirely in an individual’s own words.

The first part of refutation text (see Figure 3-3) identifies the potential misconception. Text on the first section of the refutation text items used in the main study averages 20 words. The second part is designed to prepare the learner for a correction of the possible misconception. A prompt (middle part in Figure 3-3) indicates that “this false belief must be resolved in the students’ minds”. Reading the prompt is expected to mentally prepare one to assimilate or accommodate the knowledge that refutes existing knowledge. The prompt is deliberately briefly worded to emphasize a possible challenge to the learner’s main study set of beliefs about a particular knowledge area. Text on the third section of the refutation text items used in the main study averages 4 words. Main study beliefs about subject matter are typically garnered via pretesting, as was the case in the main study.

The third and final part of the refutation text treatment (see Figure 3-3) presents correct (normative) conceptions or knowledge that attempts to refute existing misconceptions. Text on the third section of the refutation text items used in the main study averages 40 words.
Misconception Inference

To further summarize responses at the individual participant level, changes (in terms responses to test questions) were quantified as illustrated in Table 4-2.

Table 4-2

<table>
<thead>
<tr>
<th>Level of engagement with Refutation Text</th>
<th>Additional number of correct responses</th>
<th>Additional number of incorrect responses</th>
<th>Quantity of change and direction of change</th>
<th>% change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read+Para Ref Text group one</td>
<td>31</td>
<td>27</td>
<td>4</td>
<td>15%</td>
</tr>
<tr>
<td>Read Ref Text group one</td>
<td>17</td>
<td>24</td>
<td>-7</td>
<td>-29%</td>
</tr>
<tr>
<td>Read+Para Ref Text group two</td>
<td>35</td>
<td>20</td>
<td>15</td>
<td>75%</td>
</tr>
<tr>
<td>Read Ref Text group two</td>
<td>26</td>
<td>25</td>
<td>1</td>
<td>4%</td>
</tr>
</tbody>
</table>

In Table 4-2, the change between correct and incorrect responses is shown for RT groups one and two. After the pretests (not shown in this table), and after engaging with conceptual change-oriented refutation text at various levels, it is interesting to note the differences in quantity and direction of change not only between the treatment and control levels, but also between the two RT groups. For example, while RT group one that paraphrased refutation text generated 31 and 27 correct and incorrect responses at posttest, the RT group two that also paraphrased refutation text resulted in 35 and 20 correct and incorrect responses respectively. That represented 15% more correct responses than incorrect responses for students in RT group one, and 75% more correct responses than incorrect responses for students in RT group two. Based on these results,
paraphrasing the 5 RT group two refutation text items resulted in a larger positive change than paraphrasing the 5 RT group one refutation text items. This suggests that RT group two students experienced more pronounced conceptual change than RT group one students. In addition, reading without paraphrasing resulted in more incorrect responses than correct responses for the RT group one group, and no correct response gains over incorrect responses in the equivalent RT group two group. In terms of percentage change, that represented 29% less correct responses than incorrect responses for students in RT group one, and 4% more correct responses than incorrect responses for students in RT group two. It seems that something unique to each RT group treatment resulted in what appears to be more positive conceptual change for the RT group two when compared with their RT group one counterparts.

Similar analysis of individual responses to the conceptual knowledge test was conducted. A summary of incorrect responses (misconceptions) for pre and posttests is presented in Table 4-3.

Table 4-3

Summary of Incorrect Responses to Conceptual Knowledge Tests

<table>
<thead>
<tr>
<th></th>
<th>Level of engagement with Refutation Text</th>
<th>Incorrect responses at pretest (%)</th>
<th>Incorrect responses at posttest (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT group one</td>
<td>Read and paraphrase</td>
<td>67%</td>
<td>35%</td>
</tr>
<tr>
<td></td>
<td>Read</td>
<td>78%</td>
<td>59%</td>
</tr>
<tr>
<td>RT group two</td>
<td>Read and paraphrase</td>
<td>80%</td>
<td>58%</td>
</tr>
<tr>
<td></td>
<td>Read</td>
<td>86%</td>
<td>74%</td>
</tr>
</tbody>
</table>
Table 4-3 shows learners’ performance by RT group in terms of responses to the conceptual knowledge pretest and posttest. Proportions of incorrect responses in percentages are shown (the focus is on misconceptions which are inferred from the incorrect responses). From Table 4-3, it can be seen that pretesting produced more incorrect responses. Based on inferences made from incorrect responses (Guzmán, Conejo, & Gálvez, 2010; Elmadani, Mathews, & Mitrovic, 2012), this performance suggested a high level of misconception of the concepts tested among the participants. The misconceptions decreased in varying degrees after learners read or paraphrased conceptual change-oriented refutation text. From both RT groups, learners who paraphrased refutation text responded with the largest drop in incorrect responses; 67% to 35% in RT group one and 80% to 58% in RT group two. Learners in RT group two who read without paraphrasing refutation text maintained the highest proportion of incorrect responses between pre and posttests; 86% and 74% of their responses were incorrect in their pre and posttest respectively.

Since students were tested at three different levels of knowledge, misconceptions held in each knowledge area are different. Figure 4-3 presents a sample of misconceptions inferred from students’ responses in each test.
Figure 4-1. Sample of misconceptions. These are derived from the questions designed to test concepts such as keys and entity relationships as mentioned in the declarative knowledge misconceptions section, or an inability to understand and integrate SQL syntax sufficiently enough to declare a relation, as mentioned in the procedural knowledge misconceptions section.

Figure 4-1 illustrates a sample of misconceptions (inferred from students’ responses to tests) from each of the three tests. An example of a declarative knowledge misconception was confusing the entity-relation (E-R) approach for a relational data model instead of design methodology. This suggests that sometime in the past, when prior knowledge was acquired, the concepts of entity-relationship, relational data model, and design methodology were misunderstood. Figure 4-1 also illustrates misconceptions inferred from the conceptual knowledge and procedural knowledge tests.

<table>
<thead>
<tr>
<th>Declarative Knowledge Misconceptions sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Misunderstanding of when keys, check constraints, indexes, etc., and business processes and data flows are created defined</td>
</tr>
<tr>
<td>Misunderstanding of the Third Normal Form being a disadvantage for retrieval but an advantage for updates</td>
</tr>
<tr>
<td>Misunderstanding of the E-R approach as a relational data model instead of a design methodology</td>
</tr>
<tr>
<td>Misunderstanding of keys as a set of attributes that uniquely identifies an entity</td>
</tr>
<tr>
<td>Misunderstanding of effects of relational databases on storage</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conceptual Knowledge Misconceptions sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inability to understand and link the concept of row to the concept of record</td>
</tr>
<tr>
<td>Inability to understand and link the concept of table to the concept of relation</td>
</tr>
<tr>
<td>Inability to understand and link the concept of schema to the concept of units of organization</td>
</tr>
<tr>
<td>Inability to understand and link the concept of primary key to the concept of unique</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Procedural Knowledge Misconceptions sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inability to demonstrate an understanding of the concept of a table/record and apply the understanding to create a table</td>
</tr>
<tr>
<td>Inability to demonstrate an understanding of the concept of a primary key and integrate it with the concept of table in declaring a relation</td>
</tr>
<tr>
<td>Inability to apply the concepts of tables, keys, and integrate with SQL syntax in declaring a relation</td>
</tr>
</tbody>
</table>
Analysis of correlation

As a precursor to conducting multivariate analysis of variance, relationships among variables were determined. A Pearson correlation matrix was created among all of the research variables at posttest. Table 4-4 and Table 4-5 show correlation coefficients for declarative knowledge, conceptual knowledge, and procedural knowledge scores.

Table 4-4
Correlation for RT Group One

<table>
<thead>
<tr>
<th>Correlations</th>
<th>DKpre_RT group_one</th>
<th>CKpre_RT group_one</th>
<th>PKpreTest</th>
<th>DKpost_RT group_one</th>
<th>CKpost_RT group_one</th>
<th>PKpostTest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td>.125</td>
<td>.177</td>
<td>.036</td>
<td>.234</td>
<td>.523</td>
<td>.044</td>
</tr>
<tr>
<td>Sig (2-tailed)</td>
<td>.031</td>
<td>.015</td>
<td>.076</td>
<td>.163</td>
<td>.138</td>
<td>.055</td>
</tr>
<tr>
<td>N</td>
<td>31</td>
<td>31</td>
<td>31</td>
<td>31</td>
<td>31</td>
<td>31</td>
</tr>
</tbody>
</table>

Table 4-5
Correlation for RT Group Two

<table>
<thead>
<tr>
<th>Correlations</th>
<th>DKpre_RT group_two</th>
<th>CKpre_RT group_two</th>
<th>PKpreTest</th>
<th>DKpost_RT group_two</th>
<th>CKpost_RT group_two</th>
<th>PKpostTest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td>.234</td>
<td>.448</td>
<td>.012</td>
<td>.361</td>
<td>.260</td>
<td>.523**</td>
</tr>
<tr>
<td>Sig (2-tailed)</td>
<td>.002</td>
<td>.000</td>
<td>.055</td>
<td>.523**</td>
<td>.150</td>
<td>.000</td>
</tr>
<tr>
<td>N</td>
<td>31</td>
<td>31</td>
<td>31</td>
<td>31</td>
<td>31</td>
<td>31</td>
</tr>
</tbody>
</table>

As Tables 4-4 and 4-5 show, both RT group one and RT group two correlation coefficients ranged from small (.1 ≤ r ≤ .23) such as the relationship between DK posttest RT group one and PK posttest, to medium (.24 ≤ r ≤ .36) and large (r ≥ 0.37) such as the
relationship between DK pretest RT group one and DK posttest RT group one. Another example was the significant relationship between PK pretest RT group two and posttest RT group two. For a significant positive correlation, when one variable increased, the other variable also increased. There were few exceptions: for example, CK pretest RT group one and CK posttest RT group one suggested a weak negative relationship, \( r = -0.039 \). More importantly, relationships among dependent variables (posttest scores) indicated medium to large positive (and negative) linear relationships that suggested that the variables could be used in MANOVA (Cohen, 1988).

**Results by Research Question**

Pre and post knowledge assessment test scores were collected for different categories of knowledge groups. The main statistical results are presented next.

**Question 1: Effects of Reading vs. Reading and Paraphrasing Refutation Text on Different Types of Knowledge**

How does reading refutation text differ from reading and paraphrasing refutation text in terms of declarative knowledge performance, conceptual knowledge performance, and procedural knowledge performance?

To examine research question one for each group, MANOVA was conducted to assess if there were differences in declarative knowledge performance, conceptual
knowledge performance, and procedural knowledge performance (the dependent variables) by reading vs. reading and paraphrasing (levels of the refutation text independent variable). Prior to analysis, the assumption of equality of covariance matrices was assessed with a Box’s M test. The results of the Box’s M test were not significant, $p = .836$ for RT group one and $p = .051$ for RT group two, suggesting that the assumption was met (Tinsley & Brown, 2000). The assumption of equality of variance was assessed with Levene’s tests. The results of the Levene’s tests were all not significant ($p$ ranged from .184 to .997 for all groups), which meant the dependent variable scores were evenly distributed across all student groups in both groups (Thompson, 1999), suggesting the assumption was met.

**Results (research question one)**

Descriptive statistics for RT group one and two are presented in Table 4-6 and Table 4-7 respectively.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RT group one</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DK posttest Read</td>
<td>2.54</td>
<td>.72</td>
<td>1.0</td>
<td>3.5</td>
<td>-0.69</td>
<td>0.17</td>
<td>14</td>
</tr>
<tr>
<td>Read + Para</td>
<td>2.53</td>
<td>.84</td>
<td>1.0</td>
<td>3.5</td>
<td>-0.38</td>
<td>-0.73</td>
<td>17</td>
</tr>
<tr>
<td>CK posttest Read</td>
<td>2.00</td>
<td>1.41</td>
<td>0.0</td>
<td>4.0</td>
<td>0.19</td>
<td>-1.21</td>
<td>14</td>
</tr>
<tr>
<td>Read + Para</td>
<td>2.59</td>
<td>1.42</td>
<td>0.0</td>
<td>5.0</td>
<td>0.10</td>
<td>-0.49</td>
<td>17</td>
</tr>
<tr>
<td>PK posttest Read</td>
<td>2.46</td>
<td>2.19</td>
<td>0.0</td>
<td>6.0</td>
<td>0.20</td>
<td>-1.55</td>
<td>14</td>
</tr>
<tr>
<td>Read + Para</td>
<td>3.79</td>
<td>2.69</td>
<td>0.0</td>
<td>7.0</td>
<td>-0.38</td>
<td>-1.67</td>
<td>17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RT group two</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DK posttest Read</td>
<td>2.60</td>
<td>.65</td>
<td>1.3</td>
<td>3.7</td>
<td>-0.51</td>
<td>-0.19</td>
<td>16</td>
</tr>
<tr>
<td>Read + Para</td>
<td>3.00</td>
<td>.46</td>
<td>2.0</td>
<td>3.7</td>
<td>-0.86</td>
<td>0.82</td>
<td>19</td>
</tr>
<tr>
<td>CK posttest Read</td>
<td>1.44</td>
<td>1.15</td>
<td>0.0</td>
<td>4.0</td>
<td>0.47</td>
<td>0.06</td>
<td>16</td>
</tr>
<tr>
<td>Read + Para</td>
<td>2.32</td>
<td>1.00</td>
<td>1.0</td>
<td>4.0</td>
<td>0.02</td>
<td>-1.07</td>
<td>19</td>
</tr>
<tr>
<td>PK posttest Read</td>
<td>3.06</td>
<td>2.10</td>
<td>1.0</td>
<td>5.5</td>
<td>-0.34</td>
<td>-1.48</td>
<td>16</td>
</tr>
<tr>
<td>Read + Para</td>
<td>4.00</td>
<td>2.27</td>
<td>1.0</td>
<td>7.0</td>
<td>-0.52</td>
<td>-1.69</td>
<td>19</td>
</tr>
</tbody>
</table>

*Note: Means per test based on different scales, i.e., DK means should not be compared with CK or PK means*
For RT group one (see Table 4-6), DK mean scores for the read-only group \((N = 14)\) were higher than scores for the DK read-and-paraphrase group \((N = 17)\), CK mean scores for the read-only group \((N = 14)\) were lower than scores for the CK read-and-paraphrase group \((N = 17)\), and PK mean scores for the read-only group \((N = 14)\) were lower than scores for the PK read-and-paraphrase group \((N = 17)\). For RT group two (see Table 4-7), DK mean scores for the read-only group \((N = 16)\) were lower than scores for the read-and-paraphrase group \((N = 19)\), CK mean scores for the read-only group \((N = 16)\) were lower than scores for the read-and-paraphrase group \((N = 19)\), and PK mean scores for the read-only group \((N = 16)\) were lower than scores for the read-and-paraphrase group \((N = 19)\).

To supplement the descriptive statistics shown in Table 4-6 and Table 4-7, Figures 4-2 to Figure 4-4 show categories of the scores obtained by participants in each of the three tests, for students who received refutation text group one items. Reading versus reading and paraphrasing are compared.
The graphic shows the differences in post test scores of declarative, conceptual, and procedural knowledge tests. For each test, the scores are grouped so that, for example, students who scored between 3 and 4 points in the declarative knowledge test were mostly from the group that read and paraphrased. In the conceptual and procedural knowledge tests, no student who read without paraphrasing scored in the top brackets (5 points for the conceptual knowledge test, and 6.1 to 8 points for the procedural knowledge test).

Figures 4-5 to Figure 4-7 show categories of the scores obtained by participants in each of the three tests, for students who received refutation text group two items. Reading versus reading and paraphrasing are compared.
93

**Figure 4-5.** Comparison of RT Group Two declarative knowledge post test scores for reading vs. reading and paraphrasing. Most students who read and paraphrased scored more points in the higher brackets.

**Figure 4-6.** Comparison of RT Group Two conceptual knowledge post test scores for reading vs. reading and paraphrasing. More students who read and paraphrased scored 1, 3, and 4 points than students who read without paraphrasing.

**Figure 4-7.** Comparison of RT Group Two procedural knowledge post test scores for reading vs. reading and paraphrasing. More students who read and paraphrased scored in the higher brackets than students who read without paraphrasing.

The graphic (see Figures 4-5 to Figure 4-7) shows the differences in post test scores of declarative, conceptual, and procedural knowledge tests. For each test, the scores are bracketed so that, for example, in the conceptual knowledge test six students who read without paraphrasing and five who read and paraphrased scored 2 points. In the procedural knowledge test, no student who read without paraphrasing scored in the top bracket (6.1 to 8 points). Additional statistical analysis of the descriptive profile follows.

Table 4-8 shows MANOVA results for RT group one, reading vs. reading and paraphrasing.
Table 4-8

**MANOVA Results for RT Group One: Reading vs. Reading and Paraphrasing**

**RT group one**

<table>
<thead>
<tr>
<th>Effect</th>
<th>Pillai's Trace</th>
<th>Wilks' Lambda</th>
<th>Hotelling's Trace</th>
<th>Roy's Largest Root</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>.937</td>
<td>.063</td>
<td>14.801</td>
<td>14.801</td>
</tr>
<tr>
<td></td>
<td>133.212</td>
<td>133.212</td>
<td>133.212</td>
<td>133.212</td>
</tr>
<tr>
<td></td>
<td>3.000</td>
<td>3.000</td>
<td>3.000</td>
<td>3.000</td>
</tr>
<tr>
<td></td>
<td>27.00</td>
<td>27.00</td>
<td>27.00</td>
<td>27.00</td>
</tr>
<tr>
<td></td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>.937</td>
<td>.937</td>
<td>.937</td>
<td>.937</td>
</tr>
<tr>
<td></td>
<td>399.635</td>
<td>399.635</td>
<td>399.635</td>
<td>399.635</td>
</tr>
<tr>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

a. Design: Intercept + readOnly_vsg_readAndPara  
b. Exact statistic  
c. Computed using alpha = .05

MANOVA results for RT group one (see Table 4-8) showed no significant difference between reading vs. reading and paraphrasing, $F(3, 27) = 1.00, p = .407, \eta^2 = .10$, suggesting that there were no significantly large differences between reading and reading and paraphrasing (Cohen, 1988). Owing to the non-significant MANOVA result for RT group one, no follow-up test (ANOVA) was conducted.

Table 4-9 shows the MANOVA results for RT group two, reading vs. reading and paraphrasing.
On the other hand, MANOVA results for RT group two (see Table 4-9) showed a statistically significant difference in declarative, conceptual, and procedural knowledge performances based on the level of engaging with refutation text, $F(3, 31) = 3.87$, $p = .018$, partial $\eta^2 = .272$, suggesting that there were large differences between reading and reading and paraphrasing. The multivariate effect size of .27 implies that engaging with refutation text accounted for approximately 27% of the variance in test scores.

Since MANOVA for RT group two was significant, a series of individual follow-up ANOVAs on each of the dependent variables was then conducted. The ANOVAs were run to determine how declarative knowledge performance, conceptual knowledge performance, and procedural knowledge performance differed for reading only vs. reading and paraphrasing. The ANOVAs presented below showed significant differences
between read only and read and paraphrase groups for all tests except procedural knowledge RT group two (see Table 4-10).

Table 4-10

ANOVA Results for RT Group Two: Reading vs. Reading and Paraphrasing

<table>
<thead>
<tr>
<th>Source</th>
<th>Dependent Variable</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig</th>
<th>Partial Eta Squared</th>
<th>Noncent Parameter</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>DKpost_RTgroup_two</td>
<td>1.361</td>
<td>1</td>
<td>1.361</td>
<td>4.469</td>
<td>.042</td>
<td>.119</td>
<td>4.469</td>
<td>.537</td>
</tr>
<tr>
<td></td>
<td>CKpost_RTgroup_two</td>
<td>6.700</td>
<td>1</td>
<td>6.700</td>
<td>5.812</td>
<td>.022</td>
<td>.150</td>
<td>5.812</td>
<td>.648</td>
</tr>
<tr>
<td></td>
<td>PKpostTest</td>
<td>7.634</td>
<td>1</td>
<td>7.634</td>
<td>1.585</td>
<td>.217</td>
<td>.046</td>
<td>1.585</td>
<td>.231</td>
</tr>
<tr>
<td>Intercept</td>
<td>DKpost_RTgroup_two</td>
<td>272.789</td>
<td>1</td>
<td>272.789</td>
<td>895.850</td>
<td>.000</td>
<td>.964</td>
<td>895.850</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>CKpost_RTgroup_two</td>
<td>122.357</td>
<td>1</td>
<td>122.357</td>
<td>106.138</td>
<td>.000</td>
<td>.763</td>
<td>106.138</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>PKpostTest</td>
<td>433.234</td>
<td>1</td>
<td>433.234</td>
<td>89.952</td>
<td>.000</td>
<td>.732</td>
<td>89.952</td>
<td>1.000</td>
</tr>
<tr>
<td>readOnly_vs_readAndPara</td>
<td>DKpost_RTgroup_two</td>
<td>1.361</td>
<td>1</td>
<td>1.361</td>
<td>4.469</td>
<td>.042</td>
<td>.119</td>
<td>4.469</td>
<td>.537</td>
</tr>
<tr>
<td></td>
<td>CKpost_RTgroup_two</td>
<td>6.700</td>
<td>1</td>
<td>6.700</td>
<td>5.812</td>
<td>.022</td>
<td>.150</td>
<td>5.812</td>
<td>.648</td>
</tr>
<tr>
<td></td>
<td>PKpostTest</td>
<td>7.634</td>
<td>1</td>
<td>7.634</td>
<td>1.585</td>
<td>.217</td>
<td>.046</td>
<td>1.585</td>
<td>.231</td>
</tr>
<tr>
<td>Error</td>
<td>DKpost_RTgroup_two</td>
<td>10.040</td>
<td>33</td>
<td>.305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CKpost_RTgroup_two</td>
<td>38.043</td>
<td>33</td>
<td>1.153</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PKpostTest</td>
<td>158.938</td>
<td>33</td>
<td>4.816</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>DKpost_RTgroup_two</td>
<td>289.556</td>
<td>35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CKpost_RTgroup_two</td>
<td>173.000</td>
<td>35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PKpostTest</td>
<td>613.000</td>
<td>35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>DKpost_RTgroup_two</td>
<td>11.410</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CKpost_RTgroup_two</td>
<td>44.743</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PKpostTest</td>
<td>166.571</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. R Squared = .119 (Adjusted R Squared = .093)
b. R Squared = .150 (Adjusted R Squared = .124)
c. R Squared = .046 (Adjusted R Squared = .017)
d. Computed using alpha = .05

Table 4-10 shows that the level of engagement with refutation text has a statistically significant effect on declarative knowledge performance \( (F(1, 33) = 4.47, p = .042, \text{partial } \eta^2 = .12) \), a statistically significant effect on conceptual knowledge performance \( (F(1, 33) = 5.81, p = .022, \text{partial } \eta^2 = .15) \), but the effect on procedural knowledge performance was not statistically significant \( (F(1, 33) = 1.59, p = .217, \text{partial } \eta^2 = .10) \).
partial $\eta^2 = .046$). This is in line with the differences in means shown in Table 4-7 (Read vs. Read+Paraphrase posttest means for RT group two).

**Question 2: Effects of Reading Refutation Text vs. Not Reading Refutation Text on Different Types of Knowledge**

How do *refutation text group one students who did not read refutation text set two* differ from *refutation text group two students who read refutation text set two* in terms of *declarative knowledge performance, conceptual knowledge performance, and procedural knowledge performance*?

To examine research question two, a MANOVA was conducted to assess if there were differences in declarative knowledge, conceptual knowledge, and procedural knowledge scores by reading refutation text (RT group two students who did not receive RT set one) vs. not reading refutation text (RT group one students who did not receive RT set two). Total scores were used for this analysis (since students in RT group one served as controls for students in RT group two, and vice versa). Prior to analysis, the assumption of equality of covariance matrices was assessed with a Box’s M test. The results of the Box’s M test were not significant, $p = .908$, suggesting that the assumption was met (Tinsley & Brown, 2000). The assumption of equality of variance was assessed with Levene’s tests. The results of Levene’s tests were all not significant ($p$ ranged from .569 to .984 for all tests), which meant the dependent variable scores were evenly
distributed across all student groups (Thompson, 1999), suggesting the assumption was met.

**Results (research question two)**

Descriptive statistics for posttests comparing the RT group one who did not receive RT set two vs. the RT group two who did not receive RT set one are presented in Table 4-11.

<table>
<thead>
<tr>
<th>Table 4-11</th>
<th>Posttest Means for RT Group One Students who did Not Read RT Group Two Treatment vs. RT Group Two Students who Read RT Group Two Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Descriptive Statistics</strong></td>
<td></td>
</tr>
<tr>
<td>RT group one students who did not read RT set two</td>
<td>12.36</td>
</tr>
<tr>
<td>RT group two students who read RT set two</td>
<td>12.13</td>
</tr>
<tr>
<td>Total</td>
<td>12.23</td>
</tr>
<tr>
<td>RT group one students who did not read RT set two</td>
<td>4.07</td>
</tr>
<tr>
<td>RT group two students who read RT set two</td>
<td>2.56</td>
</tr>
<tr>
<td>Total</td>
<td>3.2</td>
</tr>
<tr>
<td>RT group one students who did not read RT set two</td>
<td>2.46</td>
</tr>
<tr>
<td>RT group two students who read RT set two</td>
<td>3.06</td>
</tr>
<tr>
<td>Total</td>
<td>2.78</td>
</tr>
</tbody>
</table>

*Note: Means per test based on different scales, i.e., DK means should not be compared with CK or PK means*

The descriptive statistics in Table 4-11 shows declarative knowledge mean posttest scores for RT group one students who did not receive RT set two ($N = 14$) vs.
scores for RT group two students who did not receive RT set one \((N = 16)\), CK mean scores for RT group one students who did not receive RT set two \((N = 14)\) were higher than scores for RT group students two who did not receive RT set one \((N = 16)\), and PK mean scores for RT group one students who did not receive RT set two \((N = 14)\) were lower than scores for RT group two students who did not receive RT set one \((N = 16)\).

Table 4-12 shows the MANOVA results comparing RT group one students who did not receive RT set two vs. RT group two students who did not receive RT set one.

Table 4-12

<table>
<thead>
<tr>
<th>MANOVA for RT Group One Students who did Not Receive RT Set Two vs. RT Group Two Students who did Not Receive RT Set One</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Multivariate Tests</strong>&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Effect</td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>Wilks' Lambda</td>
</tr>
<tr>
<td>Hotelling's Trace</td>
</tr>
<tr>
<td>Roy's Largest Root</td>
</tr>
<tr>
<td>RTgroup1_notReadRTset2 vs. RTgroup2_readRTset2</td>
</tr>
<tr>
<td>Wilks' Lambda</td>
</tr>
<tr>
<td>Hotelling's Trace</td>
</tr>
<tr>
<td>Roy's Largest Root</td>
</tr>
</tbody>
</table>

<sup>a</sup> Design: Intercept + RTgroup1_notReadRTset2 vs_RTgroup2_readRTset2

<sup>b</sup> Exact statistic

<sup>c</sup> Computed using alpha = .05

As shown in Table 4-12, results for the MANOVA did not show significant differences between RT group two who did not receive RT set one and RT group one who did not receive RT set two, \(F(3, 26) = 2.68, p = .068\), partial \(\eta^2 = .24\), suggesting that there were no differences between declarative knowledge, conceptual knowledge, and procedural knowledge scores by RT group two who did not receive RT set one vs. RT
group one who did not receive RT set two. Since the MANOVA was not significant, individual ANOVAs were not examined further. In addition, since significance was not found, the null hypotheses was not rejected at alpha = .05.

**Question 3: Effects of Reading and Paraphrasing Refutation Text vs. Not Reading Refutation Text on Different Types of Knowledge**

How do *refutation text group one students who did not read and paraphrase refutation text set one* differ from *refutation text group two students read and paraphrased refutation text set two* in terms of *declarative knowledge performance, conceptual knowledge performance, and procedural knowledge performance*?

To examine research question three, a multivariate analysis of variance was conducted to assess if there were differences in DK performance, CK performance, and PK performance scores when comparing RT group two who did not read and paraphrase RT set one vs. RT group one who did not read and paraphrase RT set two. The total scores were used for this analysis (since students in RT group one served as controls for students in RT group two, and vice versa). Prior to analysis, the assumption of equality of covariance matrices was assessed with a Box’s M test. The results of the Box’s M test were not significant, $p = .514$, suggesting that the assumption was met (Tinsley & Brown, 2000). The assumption of equality of variance was assessed with Levene’s tests. The results of the Levene’s tests were all not significant ($p$ ranged from .164 to .543 for all
tests), which meant the dependent variable scores were evenly distributed across all student groups (Thompson, 1999), suggesting the assumption was met.

### Results (research question three)

Descriptive statistics for posttests comparing RT group two who did not read and paraphrase RT set one vs. RT group one who did not read and paraphrase RT set two are presented in Table 4-13.

<table>
<thead>
<tr>
<th>Table 4-13</th>
<th>Posttest Means for RT Group One Students who did Not Read and Paraphrase RT Group Two Treatment vs. RT Group Two Students who Read and Paraphrased RT Group Two Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Descriptive Statistics</td>
</tr>
<tr>
<td></td>
<td>RT_group_1 _not_Read_Para_RT_set_2 _vs_ RT_group_2 _read_Para_RT_set_2 _Mean _Std. _Deviation _N</td>
</tr>
<tr>
<td>DKpostTest</td>
<td>RT group one students who did not read and paraphrase RT set two</td>
</tr>
<tr>
<td></td>
<td>13.24</td>
</tr>
<tr>
<td></td>
<td>RT group two students who read and paraphrased RT set two</td>
</tr>
<tr>
<td></td>
<td>14.26</td>
</tr>
<tr>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>CKpostTest</td>
<td>RT group one students who did not read and paraphrase RT set two</td>
</tr>
<tr>
<td></td>
<td>4.59</td>
</tr>
<tr>
<td></td>
<td>RT group two students who read and paraphrased RT set two</td>
</tr>
<tr>
<td></td>
<td>4.21</td>
</tr>
<tr>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>PKpostTest</td>
<td>RT group one students who did not read and paraphrase RT set two</td>
</tr>
<tr>
<td></td>
<td>3.79</td>
</tr>
<tr>
<td></td>
<td>RT group two students who read and paraphrased RT set two</td>
</tr>
<tr>
<td></td>
<td>4.00</td>
</tr>
<tr>
<td></td>
<td>Total</td>
</tr>
</tbody>
</table>

*Note: Means per test based on different scales, i.e., DK means should not be compared with CK or PK means*  

The descriptive statistics in Table 4-13 shows that DK mean posttest scores for RT group one who did not read and paraphrase RT set two \((N = 17)\) were lower than scores for RT group two who did not read and paraphrase RT set one \((N = 19)\), CK mean
scores for RT group one who did not read and paraphrase RT set two \((N = 17)\) were higher than scores for RT group two who did not read and paraphrase RT set one \((N = 19)\), and PK mean scores for RT group one who did not read and paraphrase RT set two \((N = 17)\) were lower than scores for RT group two who did not read and paraphrase RT set one \((N = 19)\).

Table 4-14 shows the MANOVA results for comparison of RT group two students who did not receive RT set one vs. RT group one students who did not receive RT set two.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Pillai's Trace</th>
<th>Wilks' Lambda</th>
<th>Hotelling's Trace</th>
<th>Roy's Largest Root</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(.978 )</td>
<td>(.022 )</td>
<td>(.44837 )</td>
<td>(.44837 )</td>
</tr>
<tr>
<td></td>
<td>(478.25^p)</td>
<td>(478.25^p)</td>
<td>(478.25^p)</td>
<td>(478.25^p)</td>
</tr>
<tr>
<td></td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
</tr>
<tr>
<td></td>
<td>32.00</td>
<td>32.00</td>
<td>32.00</td>
<td>32.00</td>
</tr>
<tr>
<td></td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>.978</td>
<td>.978</td>
<td>.978</td>
<td>.978</td>
</tr>
<tr>
<td></td>
<td>1434.772</td>
<td>1434.772</td>
<td>1434.772</td>
<td>1434.772</td>
</tr>
<tr>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

As shown in Table 4-14, results for the MANOVA did not show significant differences between RT group two students who did not read and paraphrase RT set one, and RT group one students who did not read and paraphrase RT set two, \(F(3, 32) = 0.62, p = .608\), partial \(\eta^2 = .06\), suggesting that there were no significant differences between
declarative knowledge performance, conceptual knowledge performance, and procedural knowledge performance scores by RT group two who did not read and paraphrase RT set one vs. RT group one who did not read and paraphrase RT set two (group of 17 students in RT group one). Since the MANOVA was not significant, individual ANOVAs were not examined further. Moreover, since significant differences were not found, the null hypotheses could not be rejected.

**Question 4: Low-score vs. High-score Pretest-Posttest differences on Different Types of Knowledge**

Are low-score vs. high-score differences statistically significant in terms of pretest-posttest changes in declarative knowledge, conceptual knowledge, and procedural knowledge performance?

**Results (research question four)**

For all scores analyzed to answer research question four, all results are presented for both RT group one and RT group two. (Visuals are presented side by side for easier comparison).
Declarative Knowledge:

Table 4-15 and Table 4-16 summarize descriptive statistics based on the participants’ declarative knowledge test scores. RT group one and RT group two results are presented.

Table 4-15

<table>
<thead>
<tr>
<th>Declarative Knowledge Pre and Posttest Means for Low Score Vs High Score Groups: RT Group One Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Descriptive Statistics</strong></td>
</tr>
<tr>
<td><strong>DKpre_RTgroup_one</strong></td>
</tr>
<tr>
<td>Low DK</td>
</tr>
<tr>
<td>High DK</td>
</tr>
<tr>
<td>Total</td>
</tr>
<tr>
<td><strong>DKpost_RTgroup_one</strong></td>
</tr>
<tr>
<td>Low DK</td>
</tr>
<tr>
<td>High DK</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Table 4-16

<table>
<thead>
<tr>
<th>Declarative Knowledge Pre and Posttest Means for Low Score Vs High Score Groups: RT Group Two Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Descriptive Statistics</strong></td>
</tr>
<tr>
<td><strong>DKpre_RTgroup_two</strong></td>
</tr>
<tr>
<td>Low DK</td>
</tr>
<tr>
<td>High DK</td>
</tr>
<tr>
<td>Total</td>
</tr>
<tr>
<td><strong>DKpost_RTgroup_two</strong></td>
</tr>
<tr>
<td>Low DK</td>
</tr>
<tr>
<td>High DK</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Table 4-15 and Table 4-16 show different pre and post means of declarative knowledge scores for groups one and two. For both groups, low-score and high-score means are split across pretest and posttest declarative knowledge scores. Random assignment accidently created non-equivalent groups, and as a result, RT group one started out less able than RT group two. Specifically, the DK pre-test low-score and high-score groups for RT group two were higher than the equivalent RT group one scores. For RT group one, pre and post means scores for students in the low-score group (N = 16) decreased from 3.13 points to 2.34 points. Pre and post means scores for RT group one students in the high-score group (N = 15) decreased from 5.93 points to 2.73
points. For RT group two, pre and post means scores for students in the low-score group (N = 23) decreased sharply from 6.13 points to 2.78 points. Pre and post means scores for RT group two students in the high-score group (N = 12) followed a similar trend and decreased sharply from 9.08 points to 2.89 points.

Figure 4-8 and Figure 4-9 are profile plots that visualize the relationship of the pre and post declarative knowledge scores (marginal means) of the low-score and high-score groups, for both RT group one and RT group two.

The profile plots as shown in Figure 4-8 and Figure 4-9 represent pre and post means for low-score and high-score groups. RT group one and RT group two plots are shown. The vertical axis shows the mean scores while the horizontal axis displays the
two times (pre and post) that the declarative knowledge test was taken. In RT group one, mean scores of the high-score group seem to have fallen more sharply from pre to post than scores of the low-score group. Mean scores for students who started with more incorrect responses therefore decreased less than mean scores for students who started with less incorrect responses. In RT group two, mean scores of both the low-score and high-score groups appear to have decreased sharply from pre to post testing. In this case, both groups of students started with high correct responses, but then dropped sharply to end at posttest with high incorrect responses. Further testing was required to determine whether or not the observed differences were statistically significant, that is, to determine if the different pre and post scores produced by the low and high-score groups were significant.

Prior to analysis, the assumption of equality of covariance matrices was assessed with a Box’s M test. The results of the Box’s M test were not significant ($p = .412$ for RT group one and $p = .707$ for RT group two), suggesting that the assumption was met (Tinsley & Brown, 2000) for both groups. In short, the correlations at pretest and posttest were found to be similar across both low-score and high-score groups. In addition, the assumption of equality of variance was assessed with Levene’s tests. The results of Levene’s tests were all not significant ($p$ ranged from .137 to .974 for all tests in both groups 1 and 2), which meant the dependent variable scores were evenly distributed across all student groups (Thompson, 1999), and which suggested the assumption was met. MANOVA analysis was then conducted for RT group one and RT group two declarative knowledge.
Table 4-17 shows MANOVA results for pre and post, low-score vs. high-score groups for Declarative Knowledge RT group one.

Table 4-17
**MANOVA for Pre and Post, Low and High Score Groups for Declarative Knowledge RT Group One**

<table>
<thead>
<tr>
<th>Effect</th>
<th>Value</th>
<th>F</th>
<th>df</th>
<th>Error df</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
<th>Noncent Parameter</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre_post_DK_RTgroup1</td>
<td>.858</td>
<td>174.578</td>
<td>1.00</td>
<td>29.000</td>
<td>.000</td>
<td>.858</td>
<td>174.578</td>
<td>1.00</td>
</tr>
<tr>
<td>Wilks' Lambda</td>
<td>.142</td>
<td>174.578</td>
<td>1.00</td>
<td>29.000</td>
<td>.000</td>
<td>.858</td>
<td>174.578</td>
<td>1.00</td>
</tr>
<tr>
<td>Hotelling's Trace</td>
<td>6.020</td>
<td>174.578</td>
<td>1.00</td>
<td>29.000</td>
<td>.000</td>
<td>.858</td>
<td>174.578</td>
<td>1.00</td>
</tr>
<tr>
<td>Roy's Largest Root</td>
<td>6.020</td>
<td>174.578</td>
<td>1.00</td>
<td>29.000</td>
<td>.000</td>
<td>.858</td>
<td>174.578</td>
<td>1.00</td>
</tr>
<tr>
<td>pre_post_DK_RTgroup1 * low_high_DK</td>
<td>.690</td>
<td>64.437</td>
<td>1.00</td>
<td>29.000</td>
<td>.000</td>
<td>.690</td>
<td>64.437</td>
<td>1.00</td>
</tr>
<tr>
<td>Wilks' Lambda</td>
<td>.310</td>
<td>64.437</td>
<td>1.00</td>
<td>29.000</td>
<td>.000</td>
<td>.690</td>
<td>64.437</td>
<td>1.00</td>
</tr>
<tr>
<td>Hotelling's Trace</td>
<td>2.222</td>
<td>64.437</td>
<td>1.00</td>
<td>29.000</td>
<td>.000</td>
<td>.690</td>
<td>64.437</td>
<td>1.00</td>
</tr>
<tr>
<td>Roy's Largest Root</td>
<td>2.222</td>
<td>64.437</td>
<td>1.00</td>
<td>29.000</td>
<td>.000</td>
<td>.690</td>
<td>64.437</td>
<td>1.00</td>
</tr>
</tbody>
</table>

a. Design: intercept + low_high_DK
b. Exact statistic
c. Computed using alpha = .05

MANOVA results for RT group one (see Table 4-17) showed a statistically significant interaction effect, $F(1, 29) = 64.44$, $p = .000$, partial $\eta^2 = .69$. The MANOVA results also showed a significant main effect for pretest-posttest declarative knowledge, $F(1, 29) = 174.58$, $p = .000$, partial $\eta^2 = .86$, suggesting that there were large differences between pretest and posttest scores (Cohen, 1988).

Similar MANOVA analysis was conducted for RT group two. Table 4-18 shows MANOVA results for pre and post, low-score vs. high-score groups for Declarative Knowledge RT group two.
Table 4-18

**MANOVA For Pre and Post, Low and High Score Groups for Declarative Knowledge RT Group Two**

**RT group two**

<table>
<thead>
<tr>
<th>Effect</th>
<th>Value</th>
<th>F</th>
<th>Hypothesis df</th>
<th>Error df</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
<th>Noncent. Parameter</th>
<th>Observed Power²</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre_post_DK_R group 2</td>
<td>.947</td>
<td>592.586</td>
<td>1.00</td>
<td>33.000</td>
<td>.000</td>
<td>.947</td>
<td>592.586</td>
<td>1.000</td>
</tr>
<tr>
<td>Wilks' Lambda</td>
<td>.053</td>
<td>592.586</td>
<td>1.00</td>
<td>33.000</td>
<td>.000</td>
<td>.947</td>
<td>592.586</td>
<td>1.000</td>
</tr>
<tr>
<td>Hotelling's Trace</td>
<td>17.957</td>
<td>592.586</td>
<td>1.00</td>
<td>33.000</td>
<td>.000</td>
<td>.947</td>
<td>592.586</td>
<td>1.000</td>
</tr>
<tr>
<td>Roy's Largest Root</td>
<td>17.957</td>
<td>592.586</td>
<td>1.00</td>
<td>33.000</td>
<td>.000</td>
<td>.947</td>
<td>592.586</td>
<td>1.000</td>
</tr>
<tr>
<td>pre_post_DK_R group 2 * low_high_DK</td>
<td>.615</td>
<td>52.736</td>
<td>1.00</td>
<td>33.000</td>
<td>.000</td>
<td>.615</td>
<td>52.736</td>
<td>1.000</td>
</tr>
<tr>
<td>Wilks' Lambda</td>
<td>.385</td>
<td>52.736</td>
<td>1.00</td>
<td>33.000</td>
<td>.000</td>
<td>.615</td>
<td>52.736</td>
<td>1.000</td>
</tr>
<tr>
<td>Hotelling's Trace</td>
<td>1.598</td>
<td>52.736</td>
<td>1.00</td>
<td>33.000</td>
<td>.000</td>
<td>.615</td>
<td>52.736</td>
<td>1.000</td>
</tr>
<tr>
<td>Roy's Largest Root</td>
<td>1.598</td>
<td>52.736</td>
<td>1.00</td>
<td>33.000</td>
<td>.000</td>
<td>.615</td>
<td>52.736</td>
<td>1.000</td>
</tr>
</tbody>
</table>

a. Design: intercept + low_high_DK
b. Exact statistic
c. Computed using alpha = .05

MANOVA results for RT group two (see Table 4-18) showed a statistically significant interaction effect, $F(1, 33) = 52.74$, $p = .000$, partial $\eta^2 = .62$. MANOVA results also showed a significant main effect for pretest-posttest declarative knowledge, $F(1, 33) = 592.59$, $p = .000$, partial $\eta^2 = .95$, suggesting that there were large differences between pretest and posttest scores (Cohen, 1988).

Tests of within-subject effects were conducted, and results based on the participants’ declarative knowledge test scores are presented in Table 4-19 and Table 4-20 for RT group one and RT group two respectively.

Table 4-19 shows the results of assessments conducted to test whether or not the mean score differences reported in Table 4-15 and Table 4-16, and plotted in Figure 4-2 and Figure 4-3, were statistically significant. To determine this, the main result
considered is the \textit{pre\_post\_DK\_group\_1} by \textit{low\_high\_DK} group interaction (refer to Table 4-19). This interaction addresses the question of whether or not the low-score group increased more than the high-score group between pretests and posttests. In short, the magnitude of performance changes between the low-score and high-score students is contingent upon the pretest-posttest factor.

Table 4-19

\textit{Within-Subjects Tests for Pre and Post, Low and High Score Groups for Declarative Knowledge RT Group One}

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
<th>Noncent. Parameter</th>
<th>Observed Power*</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre_post_DK_RTgroup1</td>
<td>Sphericity Assumed</td>
<td>61.356</td>
<td>1</td>
<td>61.356</td>
<td>174.578</td>
<td>.000</td>
<td>.858</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>61.356</td>
<td>1.000</td>
<td>61.356</td>
<td>174.578</td>
<td>.000</td>
<td>.858</td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>61.356</td>
<td>1.000</td>
<td>61.356</td>
<td>174.578</td>
<td>.000</td>
<td>.858</td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>61.356</td>
<td>1.000</td>
<td>61.356</td>
<td>174.578</td>
<td>.000</td>
<td>.858</td>
</tr>
<tr>
<td>pre_post_DK_RTgroup1 * low_high_DK</td>
<td>Sphericity Assumed</td>
<td>22.647</td>
<td>1</td>
<td>22.647</td>
<td>64.437</td>
<td>.000</td>
<td>.690</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>22.647</td>
<td>1.000</td>
<td>22.647</td>
<td>64.437</td>
<td>.000</td>
<td>.690</td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>22.647</td>
<td>1.000</td>
<td>22.647</td>
<td>64.437</td>
<td>.000</td>
<td>.690</td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>22.647</td>
<td>1.000</td>
<td>22.647</td>
<td>64.437</td>
<td>.000</td>
<td>.690</td>
</tr>
<tr>
<td>Error(pre_post_DK_RTgroup1)</td>
<td>Sphericity Assumed</td>
<td>10.192</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>10.192</td>
<td>29.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>10.192</td>
<td>29.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>10.192</td>
<td>29.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Computed using alpha = .05

For RT group one, the interaction is statistically significant, F(1, 29) = 64.44, p = .000, partial $\eta^2 = .69$, suggesting that there were substantial differences between the low-score and high-score groups across pretest and posttest declarative knowledge scores (Cohen, 1988).
Tests of within-subject effects for participants’ declarative knowledge RT group two are presented in Table 4-20.

Table 4-20

*Within-Subjects Tests for Pre and Post, Low and High Score Groups for Declarative Knowledge RT Group Two*

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig</th>
<th>Partial Eta Squared</th>
<th>Noncent. Parameter</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre_post_DK_RTgroup2</td>
<td>Sphericity Assumed</td>
<td>359.017</td>
<td>1</td>
<td>359.017</td>
<td>592.586</td>
<td>.000</td>
<td>.947</td>
<td>.592.586</td>
</tr>
<tr>
<td>Greenhouse-Geisser</td>
<td>359.017</td>
<td>1.000</td>
<td>359.017</td>
<td>592.586</td>
<td>.000</td>
<td>.947</td>
<td>.592.586</td>
<td>1.000</td>
</tr>
<tr>
<td>Huynh-Feldt</td>
<td>359.017</td>
<td>1.000</td>
<td>359.017</td>
<td>592.586</td>
<td>.000</td>
<td>.947</td>
<td>.592.586</td>
<td>1.000</td>
</tr>
<tr>
<td>Lower-bound</td>
<td>359.017</td>
<td>1.000</td>
<td>359.017</td>
<td>592.586</td>
<td>.000</td>
<td>.947</td>
<td>.592.586</td>
<td>1.000</td>
</tr>
<tr>
<td>pre_post_DK_RTgroup2 * low_high_DK</td>
<td>Sphericity Assumed</td>
<td>31.950</td>
<td>1</td>
<td>31.950</td>
<td>52.736</td>
<td>.000</td>
<td>.615</td>
<td>.52.736</td>
</tr>
<tr>
<td>Greenhouse-Geisser</td>
<td>31.950</td>
<td>1.000</td>
<td>31.950</td>
<td>52.736</td>
<td>.000</td>
<td>.615</td>
<td>.52.736</td>
<td>1.000</td>
</tr>
<tr>
<td>Huynh-Feldt</td>
<td>31.950</td>
<td>1.000</td>
<td>31.950</td>
<td>52.736</td>
<td>.000</td>
<td>.615</td>
<td>.52.736</td>
<td>1.000</td>
</tr>
<tr>
<td>Lower-bound</td>
<td>31.950</td>
<td>1.000</td>
<td>31.950</td>
<td>52.736</td>
<td>.000</td>
<td>.615</td>
<td>.52.736</td>
<td>1.000</td>
</tr>
<tr>
<td>Error(pre_post_DK_RTgroup2)</td>
<td>Sphericity Assumed</td>
<td>19.993</td>
<td>33</td>
<td>.606</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greenhouse-Geisser</td>
<td>19.993</td>
<td>33.000</td>
<td>.606</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Huynh-Feldt</td>
<td>19.993</td>
<td>33.000</td>
<td>.606</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower-bound</td>
<td>19.993</td>
<td>33.000</td>
<td>.606</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For RT group two, the interaction is also statistically significant, F(1, 33) = 52.74, p = .000, partial η2 = .615, suggesting that there were substantial differences between the low-score and high-score groups across pretest and posttest declarative knowledge scores (Cohen, 1988). Since a drop in mean scores for both low-score and high-score groups across pretests and posttests was reported (refer to Figure 4-2 and Figure 4-3), these interaction results suggest that for declarative knowledge, refutation text had a substantial negative impact on the low-score students, and a profound negative impact on the high-score students.
In short, the low score group’s pre and post mean performance improved at a lower rate than the high score group. In this case, refutation text does not seem to have improved a low performer (who started with more misconceptions) or a top performer (who started with fewer misconceptions). Since pre and post-performance worsened for declarative knowledge, it can be stated that refutation text as an intervention confused the students.

**Conceptual Knowledge:**

Table 4-21 and Table 4-22 summarize descriptive statistics based on the participants’ conceptual knowledge test scores. RT group one and RT group two results are presented.

**Table 4-21**

*Conceptual Knowledge Pre and Posttest Means for Low Score vs. High Score Groups: RT Group One Treatment*

<table>
<thead>
<tr>
<th></th>
<th>low</th>
<th>high</th>
<th>CK Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CKpre_RTgroup_one</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low CK</td>
<td>.56</td>
<td></td>
<td>.507</td>
<td></td>
<td>25</td>
</tr>
<tr>
<td>High CK</td>
<td>2.67</td>
<td></td>
<td>.816</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>.97</td>
<td></td>
<td>1.016</td>
<td></td>
<td>31</td>
</tr>
<tr>
<td><strong>CKpost_RTgroup_one</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low CK</td>
<td>2.44</td>
<td></td>
<td>1.417</td>
<td></td>
<td>25</td>
</tr>
<tr>
<td>High CK</td>
<td>1.83</td>
<td></td>
<td>1.472</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>2.32</td>
<td></td>
<td>1.423</td>
<td></td>
<td>31</td>
</tr>
</tbody>
</table>

**Table 4-22**

*Conceptual Knowledge Pre and Posttest Means for Low Score vs. High Score Groups: RT Group Two Treatment*

<table>
<thead>
<tr>
<th></th>
<th>low</th>
<th>high</th>
<th>CK Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CKpre_RTgroup_two</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low CK</td>
<td>.50</td>
<td></td>
<td>.511</td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>High CK</td>
<td>2.00</td>
<td></td>
<td>.000</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td>.97</td>
<td></td>
<td>.822</td>
<td></td>
<td>35</td>
</tr>
<tr>
<td><strong>CKpost_RTgroup_two</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low CK</td>
<td>1.83</td>
<td></td>
<td>1.239</td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>High CK</td>
<td>2.09</td>
<td></td>
<td>.944</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td>1.91</td>
<td></td>
<td>1.147</td>
<td></td>
<td>35</td>
</tr>
</tbody>
</table>
Table 4-21 and Table 4-22 show different pre and post means of conceptual knowledge scores for groups 1 and 2. For RT group one, pre and post means scores for students in the low-score group increased from 0.56 points to 2.44 points. On the other hand, pre and post means scores for RT group one students in the high-score group decreased from 2.67 points to 1.83 points. For RT group two, pre and post means scores for students in the low-score group increased sharply from 0.50 points to 1.83 points. Pre and post means scores for RT group two students in the high-score group increased from 2.00 points to 2.09 points. Figure 4-10 and Figure 4-11 are plots that visualize the relationship of the pre and post conceptual knowledge scores (marginal means) of the low-score and high-score groups, for both RT group one and RT group two.

**Figure 4-10.** Profile plot for conceptual knowledge pre and post, low vs. high score groups for RT group one treatment. The low-score group at pretest scored higher at posttest, whereas the pretest high-score group scored lower at posttest.

**Figure 4-11.** Profile plot for conceptual knowledge pre and post, low vs. High score groups for RT group two. Both pretest low-score and high-score groups scored higher at posttest.
The profile plots as shown in Figure 4-10 and Figure 4-11 represent pre and post marginal means for low-score and high-score groups. RT group one and RT group two plots are shown. The vertical axis shows the mean scores while the horizontal axis displays the two times (pre and post) that the conceptual knowledge test was taken. In RT group one, mean scores of the low-score increased sharply from pre to post while scores of the high-score group decreased. In RT group two, mean scores of the low-score group increased sharply and high-score groups increased from pre to post testing. In short, the low-score group’s scores increased more sharply than the high-score group’s scores. Further testing was required to determine whether or not the observed differences were significant. To achieve this, within-subject effects were analyzed.

Prior to analysis, the assumption of equality of covariance matrices was assessed with a Box’s M test. The results of the Box’s M test were not significant for RT group one ($p = .176$), suggesting that the assumption for RT group one was met (Tinsley & Brown, 2000). Box’s M test was not calculated for RT group two. In short, the correlations at pretest and posttest were found to be similar across both low-score and high-score groups for RT group one, but not for RT group two. In addition, the assumption of equality of variance was assessed with Levene’s tests. The results of Levene’s tests were significant ($p = .025$ for RT group one pretest scores), suggesting the assumption was not met since not all the dependent variable scores were evenly distributed across all student groups (Thompson, 1999). The results of Levene’s tests were not significant ($p = .975$ for RT group one pretest scores and $p = .147$ for RT group
two posttest scores), suggesting the assumption was met for these particular test scores.

The analysis proceeded as described below.

MANOVA analysis was conducted for RT group one and RT group two conceptual knowledge.

Table 4-23 shows MANOVA results for pre and post, low-score vs. high-score groups for Conceptual Knowledge RT group one.

Table 4-23

MANOVA For Pre and Post, Low and High Score Groups for Conceptual Knowledge RT Group One

| RT group one |

<table>
<thead>
<tr>
<th>Effect</th>
<th>Value</th>
<th>F</th>
<th>Hypothesis df</th>
<th>Error df</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
<th>Noncent. Parameter</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre_post_CK_RTgroup1</td>
<td>.082</td>
<td>2.585(^5)</td>
<td>1.000</td>
<td>29.000</td>
<td>.119</td>
<td>.082</td>
<td>2.585</td>
<td>.343</td>
</tr>
<tr>
<td>Wilks' Lambda</td>
<td>.918</td>
<td>2.585(^5)</td>
<td>1.000</td>
<td>29.000</td>
<td>.119</td>
<td>.082</td>
<td>2.585</td>
<td>.343</td>
</tr>
<tr>
<td>Hotelling's Trace</td>
<td>.089</td>
<td>2.585(^5)</td>
<td>1.000</td>
<td>29.000</td>
<td>.119</td>
<td>.082</td>
<td>2.585</td>
<td>.343</td>
</tr>
<tr>
<td>Roy's Largest Root</td>
<td>.089</td>
<td>2.585(^5)</td>
<td>1.000</td>
<td>29.000</td>
<td>.119</td>
<td>.082</td>
<td>2.585</td>
<td>.343</td>
</tr>
<tr>
<td>pre_post_CK_RTgroup1 * low_high_CK</td>
<td>.375</td>
<td>17.370(^6)</td>
<td>1.000</td>
<td>29.000</td>
<td>.000</td>
<td>.375</td>
<td>17.370</td>
<td>.981</td>
</tr>
<tr>
<td>Wilks' Lambda</td>
<td>.625</td>
<td>17.370(^6)</td>
<td>1.000</td>
<td>29.000</td>
<td>.000</td>
<td>.375</td>
<td>17.370</td>
<td>.981</td>
</tr>
<tr>
<td>Hotelling's Trace</td>
<td>.599</td>
<td>17.370(^6)</td>
<td>1.000</td>
<td>29.000</td>
<td>.000</td>
<td>.375</td>
<td>17.370</td>
<td>.981</td>
</tr>
<tr>
<td>Roy's Largest Root</td>
<td>.599</td>
<td>17.370(^6)</td>
<td>1.000</td>
<td>29.000</td>
<td>.000</td>
<td>.375</td>
<td>17.370</td>
<td>.981</td>
</tr>
</tbody>
</table>

a. Design: Intercept + low_high_CK
   Within Subjects Design: pre_post_CK_RTgroup1
b. Exact statistic
c. Computed using alpha = .05

MANOVA results for RT group one (see Table 4-23) showed a statistically significant interaction effect, \(F(1, 29) = 17.37, p = .000\), partial \(\eta^2 = .38\). On the other hand, the MANOVA results showed a non-significant main effect for pretest-posttest...
conceptual knowledge, $F(1, 29) = 2.59, p = .119$, partial $\eta^2 = .082$, suggesting that there were no large differences between pretest and posttest scores (Cohen, 1988).

Similar MANOVA analysis was conducted for RT group two. Table 4-26 shows MANOVA results for pre and post, low-score vs. high-score groups for Conceptual Knowledge RT group two.

Table 4-24

**MANOVA for Pre and Post, Low and High Score Groups for Conceptual Knowledge RT Group Two**

<table>
<thead>
<tr>
<th>Effect</th>
<th>Pillai's Trace Value</th>
<th>F(1, 29)</th>
<th>df Hypothesis</th>
<th>df Error</th>
<th>Sig</th>
<th>Partial Eta Squared</th>
<th>Noncent. Parameter</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre_post_CK_RTgroup2</td>
<td>2.97</td>
<td>13.932</td>
<td>1.000</td>
<td>33.000</td>
<td>.001</td>
<td>.297</td>
<td>13.932</td>
<td>.952</td>
</tr>
<tr>
<td>Wilks' Lambda</td>
<td>7.03</td>
<td>13.932</td>
<td>1.000</td>
<td>33.000</td>
<td>.001</td>
<td>.297</td>
<td>13.932</td>
<td>.952</td>
</tr>
<tr>
<td>Hotelling's Trace</td>
<td>422</td>
<td>13.932</td>
<td>1.000</td>
<td>33.000</td>
<td>.001</td>
<td>.297</td>
<td>13.932</td>
<td>.952</td>
</tr>
<tr>
<td>Roy's Largest Root</td>
<td>422</td>
<td>13.932</td>
<td>1.000</td>
<td>33.000</td>
<td>.001</td>
<td>.297</td>
<td>13.932</td>
<td>.952</td>
</tr>
</tbody>
</table>

$a$. Design: intercept + low_high_CK  
$\text{Within Subjects Design: pre_post_CK_RTgroup2}$  
$b$. Exact statistic  
$c$. Computed using alpha = .05

MANOVA results for RT group two (see Table 4-24) showed a statistically significant interaction effect, $F(1, 33) = 10.60, p = .003$, partial $\eta^2 = .243$. In addition, the MANOVA results showed a significant main effect for pretest-posttest conceptual knowledge, $F(1, 33) = 13.93, p = .001$, partial $\eta^2 = .297$, suggesting that there were large differences between pretest and posttest scores (Cohen, 1988).
Table 4-25 and Table 4-26 show within-subject effects based on the participants’ conceptual knowledge test scores for both RT group one and RT group two.

Table 4-25

*Within-Subjects Tests for Pre and Post, Low and High Score Groups for Conceptual Knowledge RT Group One*

<table>
<thead>
<tr>
<th>Tests of Within-Subjects Effects</th>
<th>CK_change_RTgroup1</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure</td>
<td>Type III Sum of Squares</td>
<td>df</td>
<td>Mean Square</td>
<td>F</td>
<td>Sig.</td>
<td>Partial η&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>Source</td>
<td>pre_post_CK_RTgroup1</td>
<td>2.650</td>
<td>1</td>
<td>2.650</td>
<td>2.585</td>
<td>.119</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>2.650</td>
<td>1.000</td>
<td>2.650</td>
<td>2.585</td>
<td>.119</td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>2.650</td>
<td>1.000</td>
<td>2.650</td>
<td>2.585</td>
<td>.119</td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>2.650</td>
<td>1.000</td>
<td>2.650</td>
<td>2.585</td>
<td>.119</td>
</tr>
<tr>
<td></td>
<td>pre_post_CK_RTgroup1 * low_high_CK</td>
<td>17.812</td>
<td>1</td>
<td>17.812</td>
<td>17.370</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>17.812</td>
<td>1.000</td>
<td>17.812</td>
<td>17.370</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>17.812</td>
<td>1.000</td>
<td>17.812</td>
<td>17.370</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>17.812</td>
<td>1.000</td>
<td>17.812</td>
<td>17.370</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Error(pre_post_CK_RTgroup1)</td>
<td>29.737</td>
<td>29</td>
<td>1.025</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sphericity Assumed</td>
<td>29.737</td>
<td>29</td>
<td>29.000</td>
<td>1.025</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>29.737</td>
<td>29</td>
<td>29.000</td>
<td>1.025</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>29.737</td>
<td>29</td>
<td>29.000</td>
<td>1.025</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>29.737</td>
<td>29</td>
<td>29.000</td>
<td>1.025</td>
<td></td>
</tr>
</tbody>
</table>

For RT group one, the interaction was statistically significant, $F(1, 29) = 17.37, p = .000$, partial $\eta^2 = .38$, suggesting that there were substantial differences between the low-score and high-score groups across pretest and posttest conceptual knowledge scores (Cohen, 1988). The high-score group’s performance decreased noticeably whereas the low-score group’s performance increased across pretests and posttests. The interaction suggests that the refutation text had a profound conceptual change effect on all students’ conceptual knowledge (in both groups).
For RT group two, the interaction was also statistically significant, $F(1, 33) = 10.60, p = .003$, partial $\eta^2 = .24$, suggesting that there were substantial differences between the low-score and high-score groups across pretest and posttest conceptual knowledge scores (Cohen, 1988).

Table 4-26

**Within-Subjects Tests for Pre and Post, Low and High Score Groups for Conceptual Knowledge**

**RT Group Two**

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>Mean Square</th>
<th>$F$</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
<th>Noncent. Parameter</th>
<th>Observed Power*</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre_post_CK_RTgroup2</td>
<td>Sphericity Assumed</td>
<td>7.650</td>
<td>1</td>
<td>13.932</td>
<td>.001</td>
<td>.297</td>
<td>13.932</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>7.650</td>
<td>1.000</td>
<td>13.932</td>
<td>.001</td>
<td>.297</td>
<td>13.932</td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>7.650</td>
<td>1.000</td>
<td>13.932</td>
<td>.001</td>
<td>.297</td>
<td>13.932</td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>7.650</td>
<td>1.000</td>
<td>13.932</td>
<td>.001</td>
<td>.297</td>
<td>13.932</td>
</tr>
<tr>
<td>pre_post_CK_RTgroup2 * low_high_CK</td>
<td>Sphericity Assumed</td>
<td>5.822</td>
<td>1</td>
<td>10.602</td>
<td>.003</td>
<td>.243</td>
<td>10.602</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>5.822</td>
<td>1.000</td>
<td>10.602</td>
<td>.003</td>
<td>.243</td>
<td>10.602</td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>5.822</td>
<td>1.000</td>
<td>10.602</td>
<td>.003</td>
<td>.243</td>
<td>10.602</td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>5.822</td>
<td>1.000</td>
<td>10.602</td>
<td>.003</td>
<td>.243</td>
<td>10.602</td>
</tr>
<tr>
<td>Error(pre_post_CK_RTgroup2)</td>
<td>Sphericity Assumed</td>
<td>18.121</td>
<td>33</td>
<td>.549</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>18.121</td>
<td>33.000</td>
<td>.549</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>18.121</td>
<td>33.000</td>
<td>.549</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>18.121</td>
<td>33.000</td>
<td>.549</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Computed using alpha = .05

Table 4-25 and Table 4-26 shows the results of assessments conducted to test whether or not the test score differences were statistically significant. To determine this, the main result considered was the `pre_post_CK_group_1 by low_high_CK` group interaction (refer to Table 4-25 and Table 4-26). This interaction addressed the question of whether or not the low-score increased more than the high-score group between pretests and posttests. In short, the magnitude of performance changes between the low-
score and high-score students was contingent upon the pretest-posttest factor. For conceptual knowledge performance, both RT group one and RT group two low-score groups improved from pretest to posttest, while the high-score groups either decreased (RT group one) in conceptual knowledge performance or increased (RT group two).

In short, the low score group’s pre and post mean performance improved more sharply than the high score groups in both RT group one and RT group two. Refutation text seems to have improved all low performers’ conceptual knowledge (who started with more misconceptions) more than the top performers (who started with less misconceptions). Since pre and post-performance improved for the low-score groups more than it did for the high-score group in terms of conceptual knowledge, it can be stated that refutation text appears to have a profound positive conceptual change effect on students with more misconceptions (an effective remediation strategy).

**Procedural Knowledge:**

Table 4-27 and Table 4-28 summarize descriptive statistics based on the participants’ procedural knowledge test scores. RT group one and RT group two results are presented.
Table 4-27
Procedural Knowledge Pre and Posttest Means For Low Score Vs High Score Groups: RT Group One

<table>
<thead>
<tr>
<th>Low PK</th>
<th>High PK</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>PKpreTest</td>
<td>1.433</td>
<td>1.347</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>High PK</td>
<td>5.719</td>
<td>.729</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3.645</td>
<td>2.419</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>PKpostTest</td>
<td>1.600</td>
<td>1.938</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>High PK</td>
<td>4.688</td>
<td>2.088</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3.194</td>
<td>2.528</td>
<td>31</td>
<td></td>
</tr>
</tbody>
</table>

Table 4-28
Procedural Knowledge Pre and Posttest Means For Low Score Vs High Score Groups: RT Group Two

<table>
<thead>
<tr>
<th>Low PK</th>
<th>High PK</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>PKpreTest</td>
<td>2.023</td>
<td>1.728</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>High PK</td>
<td>5.308</td>
<td>.480</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3.243</td>
<td>2.126</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>PKpostTest</td>
<td>2.886</td>
<td>2.143</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>High PK</td>
<td>4.731</td>
<td>1.877</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3.571</td>
<td>2.213</td>
<td>35</td>
<td></td>
</tr>
</tbody>
</table>

Table 4-27 and Table 4-28 show different pre and post means of procedural knowledge scores for groups 1 and 2. For RT group one, pre and post means scores for students in the low-score group increased from 1.43 points to 1.6 points. On the other hand, pre and post means scores for RT group one students in the high-score group decreased noticeably from 5.72 points to 4.69 points. For RT group two, pre and post means scores for students in the low-score group increased from 2.02 points to 2.89 points. Pre and post means scores for RT group two students in the high-score group decreased from 5.31 points to 4.73 points. Figure 4-6 and Figure 4-7 are plots that visualize the relationship of the pre and post procedural knowledge scores (marginal means) of the low-score and high-score groups, for both RT group one and RT group two.
The profile plots as shown in Figure 4-12 and Figure 4-13 represent pre and post marginal means for low-score and high-score groups. RT group one and RT group two plots are shown. The vertical axis shows the mean scores while the horizontal axis displays the two times (pre and post) that the procedural knowledge test was taken. In RT group one, mean scores of the low-score group increased from pre to post while scores of the high-score group decreased. In RT group two, mean scores of the low-score group also increased while the high-score group’s scores decreased from pre to post.
testing. Further testing was required to determine whether or not the observed differences were significant. To achieve this, within-subject effects were analyzed.

Prior to analysis, the assumption of equality of covariance matrices was assessed with a Box’s M test. The results of the Box’s M test were not significant \((p = .170)\) for RT group one and significant for RT group two \((p = .001)\), suggesting that the assumption was met for RT group one but not met for RT group two (Tinsley & Brown, 2000). In short, the correlations at pretest and posttest were found to be similar across both low-score and high-score groups in RT group one but not in RT group two. In addition, the assumption of equality of variance was assessed with Levene’s tests. The results of Levene’s tests were significant for RT group one pretest scores \((p = .014)\), not significant for RT group one posttest scores \((p = .939)\), significant for RT group two pretest scores \((p = .000)\), and not significant for RT group two posttest scores \((p = .280)\). This suggested that assumptions were met for RT group one posttest and RT group two posttest scores (scores were evenly distributed across all student groups), but not for RT group one and RT group two pretest scores (scores not evenly distributed across all student groups). The analysis proceeded.

MANOVA analysis was conducted for RT group one and RT group two procedural knowledge.

Table 4-29 shows MANOVA results for pre and post, low-score vs. high-score groups for Procedural Knowledge RT group one.
MANOVA results for RT group one (see Table 4-29) showed a statistically non-significant interaction effect, $F(1, 29) = 1.97, p = .171$, partial $\eta^2 = .064$. Similarly, the MANOVA results showed a non-significant main effect for pretest-posttest procedural knowledge, $F(1, 29) = 1.03, p = .319$, partial $\eta^2 = .034$, suggesting that there were no large differences between pretest and posttest scores (Cohen, 1988).

Similar MANOVA analysis was conducted for RT group two. Table 4-30 shows MANOVA results for pre and post, low-score vs. high-score groups for Procedural Knowledge RT group two.
Table 4-30

MANOVA For Pre and Post, Low and High Score Groups For Procedural Knowledge RT Group Two

RT group two

MANOVA results for RT group two (see Table 4-30) showed a statistically non-significant interaction effect, $F(1, 33) = 4.114$, $p = .051$, partial $\eta^2 = .111$. In addition, the MANOVA results showed a non-significant main effect for pretest-posttest conceptual knowledge, $F(1, 33) = .163$, $p = .689$, partial $\eta^2 = .005$, suggesting that there were no large differences between pretest and posttest scores (Cohen, 1988).

Tests of within-subject effects were conducted, and results based on the participants’ procedural knowledge test scores are presented in Table 4-31 and Table 4-32 for RT group one and RT group two respectively. Table 4-31 and Table 4-32 shows the results of assessments conducted to test whether or not the mean score differences were statistically significant. To determine this, the main result considered was the $pre_post_PK\_group\_1$ by $low_high_PK$ group interaction. This interaction addressed the question of whether or not the low-score group increased more than the high-score group.
between pretests and posttests. In short, the magnitude of performance changes between the low-score and high-score students was contingent upon the pretest-posttest factor.

Table 4-31
Within-Subjects Tests For Pre and Post, Low and High Score Groups for Procedural Knowledge

<table>
<thead>
<tr>
<th>RT group one</th>
</tr>
</thead>
</table>

RT group one

For RT group one, the interaction was not statistically significant, F(1, 29) = 1.97, p = .171, partial η2 = .06, suggesting that there were no substantial differences between the low-score and high-score groups across pretest and posttest procedural knowledge scores (Cohen, 1988). For RT group two (refer to Figure 4-7), the interaction was not statistically significant, F(1, 33) = 4.11, p = .051, partial η2 = .11, suggesting that there were differences, albeit not substantial, between the low-score and high-score groups across pretest and posttest procedural knowledge scores (Cohen, 1988). The high-score
group’s performance decreased and the low-score group’s performance increased across pretests and posttests, and there was no interaction effect.

Table 4-32

*Within-Subjects Tests for Pre and Post, Low and High Score Groups for Procedural Knowledge RT Group Two*

<table>
<thead>
<tr>
<th>Effect</th>
<th>Value</th>
<th>F</th>
<th>Hypothesis df</th>
<th>Error df</th>
<th>Sig</th>
<th>Partial Eta Squared</th>
<th>Noncent. Parameter</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre_post_PK</td>
<td>.005</td>
<td>.163</td>
<td>1.000</td>
<td>33.000</td>
<td>.005</td>
<td>.163</td>
<td>.068</td>
<td>.068</td>
</tr>
<tr>
<td>Wilks' Lambda</td>
<td>.995</td>
<td>.163</td>
<td>1.000</td>
<td>33.000</td>
<td>.005</td>
<td>.163</td>
<td>.068</td>
<td>.068</td>
</tr>
<tr>
<td>Hotelling's Trace</td>
<td>.005</td>
<td>.163</td>
<td>1.000</td>
<td>33.000</td>
<td>.005</td>
<td>.163</td>
<td>.068</td>
<td>.068</td>
</tr>
<tr>
<td>Roy's Largest Root</td>
<td>.005</td>
<td>.163</td>
<td>1.000</td>
<td>33.000</td>
<td>.005</td>
<td>.163</td>
<td>.068</td>
<td>.068</td>
</tr>
<tr>
<td>pre_post_PK * low_high_PK</td>
<td>.111</td>
<td>4.114</td>
<td>1.000</td>
<td>33.000</td>
<td>.051</td>
<td>.111</td>
<td>4.114</td>
<td>.504</td>
</tr>
<tr>
<td>Wilks' Lambda</td>
<td>.889</td>
<td>4.114</td>
<td>1.000</td>
<td>33.000</td>
<td>.051</td>
<td>.111</td>
<td>4.114</td>
<td>.504</td>
</tr>
<tr>
<td>Hotelling's Trace</td>
<td>.125</td>
<td>4.114</td>
<td>1.000</td>
<td>33.000</td>
<td>.051</td>
<td>.111</td>
<td>4.114</td>
<td>.504</td>
</tr>
<tr>
<td>Roy's Largest Root</td>
<td>.125</td>
<td>4.114</td>
<td>1.000</td>
<td>33.000</td>
<td>.051</td>
<td>.111</td>
<td>4.114</td>
<td>.504</td>
</tr>
</tbody>
</table>

a. Design: Intercept + low_high_PK
b. Exact statistic
c. Computed using alpha = .05

For RT group two, the high-score group’s performance decreased whereas the low-score group’s performance increased across pretests and posttests, but that was not sufficient to produce an interaction effect. This suggests that while refutation text improved the low-score groups’ procedural knowledge and decreased the high-score groups’ procedural knowledge (in both groups), the changes from pretest to posttest were not impactful enough to produce interaction effects.
Summary of Results

A summary of statistical results for all four research questions is presented in Table 4-33.

Table 4-33
Summary of Statistical Results

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RT Group One</strong></td>
<td><strong>RT Group Two</strong></td>
</tr>
</tbody>
</table>
| **Research Question 1**: Hypothesis: No statistically significant differences exist between reading refutation text and reading and paraphrasing refutation text in terms of declarative knowledge performance, conceptual knowledge performance, and procedural knowledge performance. | **MANOVA**: (Reading vs. Reading and Paraphrasing) 
\[ F(3, 27) = 1.00, p = .407, \text{ partial } \eta^2 = .10 \]. The null hypothesis was retained. | **MANOVA**: (Reading vs. Reading and Paraphrasing) 
\[ F(3, 31) = 3.87, p = .018, \text{ partial } \eta^2 = .272 \]. The null hypothesis was rejected. |
| **DK ANOVA**: (Reading vs. Reading and Paraphrasing) 
\[ F(1, 33) = 4.47, p = .042, \text{ partial } \eta^2 = .12 \]. The null hypothesis was rejected. | **CK ANOVA**: (Reading vs. Reading and Paraphrasing) 
\[ F(1, 33) = 5.81, p = .022, \text{ partial } \eta^2 = .15 \]. The null hypothesis was rejected. |
| **PK ANOVA**: (Reading vs. Reading and Paraphrasing) 
\[ F(1, 33) = 1.59, p = .217, \text{ partial } \eta^2 = .05 \]. The null hypothesis was retained. | **MANOVA** (RT group one students who did not read RT set two vs. RT group two students who read RT set two) 
\[ F(3, 26) = 2.68, p = .068, \text{ partial } \eta^2 = .24 \]. The null hypothesis was retained. |

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Results</th>
</tr>
</thead>
</table>
| **Research Question 2**: Hypothesis: No statistically significant differences exist between refutation text group one students who did not read refutation text set two and refutation text group two students who read refutation text set two in terms of declarative knowledge performance, conceptual knowledge performance, and procedural knowledge performance. | **MANOVA** (RT group one students who did not read RT set two vs. RT group two students who read RT set two) 
\[ F(3, 26) = 2.68, p = .068, \text{ partial } \eta^2 = .24 \]. The null hypothesis was retained. |
In Table 4-33, MANOVA results for RT group one showed no significant differences between reading vs. reading and paraphrasing. In contrast, results for RT group two showed a statistically significant MANOVA. Follow-up RT group two
ANOVAs showed significant differences between read only and read and paraphrase groups for all tests except for the procedural knowledge test.

For research question two, MANOVA results did not show statistically significance differences between RT group one students who did not read RT set two, and RT group two students who read RT set two. Consequently, no follow-up individual ANOVAs were conducted.

Similarly, for research question three, MANOVA results did not show statistically significance differences between RT group one students who did not read and paraphrase RT set two, and RT group two students who read and paraphrased RT set two, hence no individual ANOVAs were conducted.

To answer research question four, mixed factorial ANOVAs were conducted for each of the three different test scores. Each ANOVA compared low-score and high-score differences in terms of pretest-posttest changes. RT group one results were as follows: low-score/high-score differences for pretests and posttests were statistically significant for declarative and conceptual knowledge performance, but not statistically significant for procedural knowledge. On the other hand, RT group two results were as follows: low-score/high-score differences for pretests and posttests were also statistically significant for declarative and conceptual knowledge performance, but not statistically significant for procedural knowledge.

The next chapter discusses these results, implications, limitations, and recommendations for future work.
Chapter 5

DISCUSSION, LIMITATIONS, AND RECOMMENDATIONS

The main study investigated the effects of reading refutation text, and reading and paraphrasing refutation text on declarative knowledge performance, conceptual knowledge performance, and procedural knowledge performance (collectively called learning outcomes) among college students studying information science. Within information sciences, the knowledge domain area tested was relational database design.

The main study was administered at a large northeastern university’s college of information sciences and technology (College of IST). It was conducted in real-life teaching settings during normal class periods.

Test data were collected from 94 sophomore and freshmen college students (48 and 46 students in each section of a class) pursuing an Information Sciences and Technology (IST) bachelor’s degree program. The students were enrolled in two sections of a 200-level college class, with different instructors teaching from the same syllabus. Based on the students’ indication of a range of number of semesters completed, some would have had well-entrenched misconceptions while others may have encountered database design concepts for the first time at the time the main study was conducted.

Regarding the amount of subject matter that was already covered, the learners were expected to have covered basic relational database concepts in prior courses. In addition to prior knowledge, the main study classes covered topics such as data types,
data modeling, and an introduction to Structured Query Language (SQL). The two classes, collectively covering the same material and possessing similar subject matter, served as the ideal context for assessing prior declarative, conceptual, and procedural knowledge. That contrasted with most previous studies of conceptual change that considered young children learning new science concepts. While investigating the moment or place (or period of time) when misconceptions formed among the study participants, the question was worth asking: were some of the misconceptions learned in learners’ main study courses, or recent previous courses, or before college?

Data were collected from participants on four different occasions. Participants answered pretests on day one and day two and pretests on day eight and day nine. Pretesting included once-off collection of demographic data and standardized scale data. Participants received refutation text after pre-testing. A randomized group of participants were asked to paraphrase their understanding of refutation text.

Based on results, the differences between RT group one and RT group two warranted further analysis of the concepts that were tested.

Additional analysis of differences between treatment groups one and two

Since the design effectively contained two refutation text groups (RT group one and RT group two), differences in test scores between these two warranted further review. To understand differences in the refutation text that learners engaged with, the concepts that were tested in the main study were scrutinized. The concepts were
independently rated by two subject matter experts. Results showed variability due to
different levels of difficulty among the concepts; that is, the experts did not rate the
concepts as having similar levels of difficulty. Even though the concepts all fell under
the database design knowledge domain, they were rated at different levels of difficulty.
More specifically, the concepts that corresponded with RT group one were rated as easier
than concepts corresponding to RT group two. The complete expert ratings are shown in
Appendix F.

The differences in level of difficulty were seen as important to explaining some of
the differences in test scores. It seemed that cognitive conflict activation occurred at
different levels depending on the type of misconception confronted when engaging with
refutation text.

The independent ratings suggested that combining the two groups (instead of
treating them separately as RT group one and 2) could have potentially created a
confounding variable in the form of a non-homogenous refutation text treatment variable.
Because the level of difficulty of the concepts was dissimilar, it became clear that
considering RT groups one and two as two treatments retrospectively created two
treatment groups from which results could be more reliably generalized. Table 5-1
illustrates how the main study’s group divisions relate with the independent expert
ratings.
Table 5-1

Relationship of Groups, Concept Ratings, and Levels of Treatment per Group

<table>
<thead>
<tr>
<th>Blocks</th>
<th>Randomized Groups (conceptLevels_RefTextLevels)</th>
<th>N</th>
<th>Total N</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT Group One (Easier concepts)</td>
<td>• easierConcepts_readOnly</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• easierConcepts_read&amp;Paraphrase</td>
<td>17</td>
<td>31</td>
</tr>
<tr>
<td>RT Group Two (More difficult concepts)</td>
<td>• difficultConcepts_readOnly</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• difficultConcepts_read&amp;Paraphrase</td>
<td>19</td>
<td>35</td>
</tr>
<tr>
<td><strong>Total Number of Participants</strong></td>
<td><strong>66</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5-1 shows the impact of asking subject matter experts to rate the concepts. The concepts rated as easier and difficult corresponded with RT group one and RT group two respectively. From a data analysis perspective, the two separate groups were each considered more reliable when analyzed separately than when combined (Kerlinger & Lee, 2000).

Without the split, the concepts and refutation text treatment would be erroneously regarded as if they were of the same level of difficulty. In hindsight, it seems that in addition to splitting the tests to minimize cognitive complexity among learners, the act of creating the groups also served the purpose of homogenizing the treatment groups.
Discussion of Findings

Comparing Reading vs. Reading and Paraphrasing Refutation Text

First, the study hypothesized that no significant differences would exist between reading refutation text vs. reading and paraphrasing refutation text in terms of declarative knowledge performance. While group comparisons based on reading vs. reading and paraphrasing suggest that paraphrasing refutation text had a statistically significant effect on declarative knowledge performance, comparisons of low-score groups and high-score groups between pretest and posttest highlight a different scenario: misconceptions at the declarative knowledge increased after students engaged with refutation text. Notwithstanding, previous research has shown that conceptual change-oriented and generative learning strategies can impede lower forms of knowledge such as verbatim or declarative knowledge, and simultaneously improve higher forms of knowledge such as conceptual, structural, or procedural knowledge (Marker & Clariana, 2007; Mayer, 1989).

It was also interesting to note that reading versus reading and paraphrasing was not significantly different for RT group one, but significantly different in RT group two.

Second, the study further hypothesized that no significant differences would exist between reading refutation text vs. reading and paraphrasing refutation text in terms of conceptual knowledge performance. Reading and paraphrasing refutation text had a statistically significant positive effect on conceptual knowledge. It seems engaging with
refutation text by paraphrasing may have activated some cognitive conflict that resulted in fewer misconceptions.

It was interesting to note that reading versus reading and paraphrasing was significantly different for both RT group one and RT group two. For the conceptual knowledge tests, the different sets of refutation text did not lead to conflicting performances as was the case between declarative knowledge RT group one and RT group two. It seems there was something inherent in the two groups that caused learners to react positively (fewer misconceptions) or negatively (more misconceptions).

The improvement in conceptual knowledge performance and decline in a lower form of knowledge (declarative) is in line with previous research whereby lower forms of knowledge are negatively impacted and higher forms of knowledge are positively impacted (Marker & Clariana, 2007; Mayer, 1989).

Third, the study also hypothesized that no significant differences would exist between reading refutation text vs. reading and paraphrasing refutation text in terms of procedural knowledge performance. Refutation text levels had a statistically significant effect on procedural knowledge performance. Based on statistical significance, effect sizes, plus direction and magnitude of mean scores, it appears there was sufficient cognitive conflict activated after paraphrasing to result in misconceptions getting either repaired or replaced.

In short, reading versus reading and paraphrasing was significantly different for both procedural knowledge test scores. Since procedural knowledge effectively tests application of both conceptual and declarative knowledge, learners’ ability to solve the
open-ended problem can be regarded as indication of their conceptual and declarative knowledge.

Comparing RT group one students who did not read RT set two vs. RT group two students who read RT set two

The study hypothesized that no statistically significant differences would exist between refutation text group one students who did not read refutation text set two and refutation text group two students who read refutation text set two in terms of declarative knowledge performance, conceptual knowledge performance, and procedural knowledge performance. The null hypothesis was retained hence no further analysis was warranted.

Comparing RT group one students who did not read and paraphrase RT set two vs. RT group two students who read and paraphrased RT set two

The study hypothesized that no statistically significant differences would exist between refutation text group one students who did not read and paraphrase refutation text set two and refutation text group two students who read and paraphrased refutation text set two in terms of declarative knowledge performance, conceptual knowledge performance, and procedural knowledge performance. The null hypothesis was retained hence no further analysis was warranted.
Comparing low-score vs. high-score differences in terms of pretest-posttest changes

The hypothesized that no statistically significant low-score group vs. high-score group differences would exist in terms of pretest-posttest changes in declarative knowledge, conceptual knowledge, and procedural knowledge performance. The null hypotheses were rejected for both RT Group One and RT Group Two declarative knowledge performance and conceptual knowledge performance (procedural knowledge performance was inconclusive), hence further discussion was warranted.

For declarative knowledge performance, results were counter-intuitive. For RT group one as well as RT group two, pre and post mean scores for students in both low-score and high-score groups decreased. In short, the refutation text treatment had a profound negative effect on declarative knowledge performance, which also happened to be the lowest form of knowledge tested in this study (compared with conceptual and procedural knowledge). The differences between pretest and posttest were statistically significant. Previous research has shown that conceptual change-oriented and generative learning strategies can impede lower forms of knowledge such as verbatim (declarative) knowledge, and simultaneously improve higher forms of knowledge such as conceptual, structural, or procedural knowledge (Marker & Clariana, 2007; Mayer, 1989). Results in this study substantiated those previous findings.

For conceptual knowledge performance, both RT group one and RT group two low-score groups improved from pretest to posttest, while the high-score groups either decreased (RT group one) in conceptual knowledge performance or increased (RT group two). This suggested that for RT group one and RT group two, there were substantial
differences between the low-score and high-score groups across pretest and posttest conceptual knowledge scores. The improvement in conceptual knowledge performance is in line with previous research whereby lower forms of knowledge (declarative) are negatively impacted and higher forms of knowledge (e.g. conceptual and procedural) are positively impacted (Marker & Clariana, 2007; Mayer, 1989). Results in this study supported those previous findings.

Concepts assessed in the concept matching exercise (conceptual knowledge performance) came from the same relational database design concepts pool as the declarative knowledge multiple choice test. That implied that performance in conceptual knowledge testing and declarative knowledge testing may not be mutually exclusive; that is, to be able to classify and relate concepts, one must be able to identify and define them at the declarative knowledge level. The refutation text administered in the main study was designed to aid classifying and relating relational database design concepts.

The significant effect of refutation text on conceptual knowledge may indicate an ability of refutation text to help learners relate knowledge. To be able to relate different knowledge pieces requires extended knowledge of declarative knowledge already learned. The refutation text instrument used in the main study incorporated this requirement in its design. Furthermore, knowing how to relate different but related aspects of knowledge requires familiarity of the particular domain within which the knowledge applies. For example, to normalize assumes different meanings within and outside relational database design. The refutation text treatment attempted to be as explicit as possible in highlighting how specific concepts related to each other.
The low score group’s pre and post mean performance improved better than the high score group in both RT group one and RT group two. Refutation text seems to have improved a low performer (who started with more misconceptions) more than a top performer (who started with less misconceptions). Since pre and post-performance improved for the low-score groups more than the high-score group in terms of procedural knowledge, it can be stated that refutation text appears to have had marginal positive conceptual change impact on students’ procedural knowledge.

For procedural knowledge testing, learners who paraphrased were able to improve scores on a relatively difficult open-ended question. The question was more difficult than the conceptual and declarative knowledge questions because it required learners to apply a combination of knowledge pieces to solve the problem. For example, the mix of concepts included the entity-relationship (E-R) sub knowledge area. Refutation text items that contained the concept areas required to solve the procedural knowledge problem were likely the ones that helped learners apply E-R concepts.

On the other hand, learners who simply glossed over the text without pausing to think about it may be the ones who fared poorly. Here, cognitive conflict may have not activated sufficiently enough to “upset” existing misconceptions, so the posttest procedural knowledge question remained as difficult as during pretesting.
Theoretical and Practical Implications for Instructional Design

From a theoretical perspective, the presence of multiple conceptions exhibited by the main study’s participants is in line with the notion of multiple, and often conflicting, conceptions of subject matter (diSessa, 2000). Some of the conceptions are correct yet some are not, hence the presence of misconceptions that require refuting and changing. To this point, results from this research provide insight into the effects of conceptual change-oriented learning strategies such as the refutation text that students engaged with.

The main study’s practical implication for instructional design is that paraphrasing refutation text is likely to positively modify conceptual knowledge performance, negatively influence declarative knowledge, and have mixed influence on procedural knowledge performance. Yet before attempts to effect conceptual change are made, there remains the challenge of determining not only what it is that students do not know, but perhaps more importantly, how they do not know. For example, Chi (2008) states that if it is determined that students’ prior conceptual knowledge is lacking, then instruction ought to focus on learning activities such as categorizing, comparing, or classifying.

Based on statistical analyses that resulted in different performances between RT group one and RT group two, coupled with concept ratings by independent experts, it seems that the tests were of different levels of difficulty. That was serendipity: there was no deliberate allocation of refutation text items to one group instead of another except for a fortuitous odd-numbered and even-numbered split. It is interesting to note that RT group one’s refutation text (for easier concepts) seems to have been better suited for
conceptual knowledge than for declarative knowledge. By implication, refutation text aligned with more difficult concepts (RT group two) seems to have been more suited for declarative knowledge than for conceptual knowledge. In short, the implication is that paraphrasing refutation text appears to have boosted different types of knowledge in different and sometimes opposing ways: it boosted conceptual knowledge more positively than it did declarative knowledge. The implications may be far-reaching: it may mean that while conceptual change-refutation text repairs misconceptions with some types of knowledge, it may simultaneously activate levels of cognitive conflict that confuse learners further with other types of knowledge. This further implies that instructors should not assume that refutation text positively influences all types of knowledge.

Results indicate differential effects of conceptual change-oriented refutation text on different levels of knowledge; that is, combination of refutation text and paraphrasing seems to have had varying effects on declarative knowledge performance, conceptual knowledge performance, and procedural knowledge performance. The implications are that the influence of paraphrased refutation text manifests differently depending on the type of knowledge being tested.

Lastly, because pre and post-performance improved for the low-score groups more than it did for the high-score group in terms of conceptual knowledge, an implication is that the changes exhibited by the two groups are remediation targets.
Limitations of the Study

A limitation was allocating refutation text to all students instead of only those individually identified as holding the most misconceptions. While this approach risked annoying students who held less misconceptions, albeit no student achieved 100% in any of the tests, it ensured efficiency on the instructor’s part by making it possible for him or her to prepare refutation text based on main study literature, and attempt to refute maximum misconceptions with as little effort as possible. The main study was therefore limited to group-level responses as opposed to individual learner responses.

Results of the study cannot be generalized beyond college-level students since the concepts were tested on college students. The concepts tested were also contained in college-level syllabi. There may be unique environmental or behavioral features at other education levels that would need to be factored in to a similar study.

Furthermore, since participants’ semesters at college ranged from 2 to 11 semesters (inclusive), results of the main study cannot be generalized outside this range. More specifically, since 41% of the students were on their fourth semester, 17% on their second semester, and 15% on their sixth semester, the conclusions drawn in the main study are generalizable mainly to students who have spent two to six semesters at college.

In addition, results of the main study are generalized to those who read and paraphrased in both pretests and posttests as requested. Moreover, the results apply to those who did paraphrase despite a few students indicating that they did not read the refutation text. The sanity check of paraphrases that indicated knowledge of the refutation text makes it difficult to assume that a student somehow paraphrased correctly
without reading the refutation text. In future studies, this check should be programatically enforced.

Since the precise reason refutation text seems to activate cognitive conflict that leads to conceptual change remains unclear (Van den Broek & Kendeou, 2008), alternative explanations are rife: one is that the sheer interest that refutation text generates (in terms of structure and wording) peaks readers’ interest and makes it easier to remember (Broughton et al., 2010). The theoretical and practical implications stated in the main study (refer to the “Theoretical and Practical Implications for Instructional Design” section”) are therefore limited to conceptual change theory pronouncements as first advanced by Posner et. al. (1982) and later revised by Strike and Posner (1992) and others (Pintrich, Marx, & Boyle, 1993) to include additional elements such as motivation, social factors, and the environment.

Another limitation applies to assumptions testing: while assumptions for univariate normality were checked in the main study, less attention was paid to multivariate normality. According to Tabachnick and Fidell (2013), while univariate normality is a prerequisite for multivariate normality, the latter does not always result from confirmation of univariate normality. Results and conclusions made in the main study are therefore limited to confirmation of univariate normality, that is, the assumption that the three dependent variables were reasonably normally distributed (or approached reasonable normal distribution).

A further limitation is that the main study did not check if refutation text planted a misconception instead of repairing one. It is possible that in the cases that a student’s
score decreased from pretest to posttest, one of the reasons could have been the refutation text having an unintended effect. Future studies may include checking for this possibility.

While results of the main study applied to college-level students, most of the research design, including data collection procedures, can be generalized beyond college students. The concepts targeted for testing would need to be adjusted accordingly, but engaging with potential participants, administering treatment, and collecting responses following the main study’s design is not unique to college-level environments.

**Considerations for future replication of the main study**

The main study conducted group-level analysis. For future replication of a similar study, additional individual-level analysis is suggested. For example, a smaller sample (or more research time) may permit giving specific refutation text that is tailor-made for a specific student. The main study applied a shotgun approach in the interest of limited resources.

It seems that engaging with refutation text coerced students into considering new knowledge that challenged their prior knowledge. The degree to which conceptual change occurs when reading only versus when reading and paraphrasing is likely explained by more than scores on a test; however, the experimentation that is possible, such as the one conducted for the main study, provides insight into some of the conceptual change that does occur when prior knowledge is suddenly challenged.
Since declarative knowledge refers to facts, refutation text that refuted declarative knowledge would have had to activate cognitive conflict about misunderstood relational database design facts. Beyond observing changing scores after students engaged with the treatment, unanswered questions pertain to the actual process of cognitive conflict activation. Within the scope of the main study, only the scores contribute toward telling a part of the conceptual change story that seems to have occurred in learners.

It is possible that because the type of knowledge elements that comprise declarative knowledge are more distinct and easier to identify, refutation text is more easily able to target misconceptions of this type. This requires further scrutiny.

Future research related to the main study should consider gathering more time-on-task data. For example, knowing time-on-task for individual parts of the activities (e.g. time taken to write one paraphrase) can help with relating duration with test scores.

The paraphrases alone can be subjected to thorough qualitative analysis. They reveal the students’ interpretation of the refutation text, therefore analysis could consider test scores in light of themes drawn from the paraphrases.

Another aspect that should be considered in attempts to replicate the main study involves students’ responses to whether they read the refutation text or not, and tying that to the paraphrases. In the main study, only the presence of a paraphrase was tied to the sample exclusion process: the Yes/No responses were not tied to the sample exclusion process. Students who indicated they had not read the refutation text went on to write paraphrases that indicated knowledge of the refutation text. It seems a few students did not understand the “did you read the refutation text?” question, so in future replication
studies this should be programmatically tied whether or not a student actually paraphrased.

In addition, in cases where pretest high-score students scored lower in posttests and pretest low-score students scored higher in posttests, the regression to the mean phenomenon should be considered. A larger sample of participants would make checking for regression to the mean more possible.

Lastly, replication of this study should consider differences between electronic and paper versions of the refutation text treatment: it is possible that participants may produce significant differences when engaging with the treatment online, compared with refutation text printed on paper.

**Recommendations**

Based on result that suggest paraphrasing refutation text produces different levels of success depending on the type of knowledge, it is recommended that in future work, different types of knowledge be isolated and targeted with refutation text and paraphrasing.

Furthermore, there may be unique environmental or behavioral features that would need to be factored in to a similar study. Examples are personal characteristics such as self-efficacy and critical thinking. In larger samples, these factors may influence the learning outcomes examined in the main study differently. Data measuring self-efficacy, critical thinking, metacognitive self-regulation, and class standing (number of
semesters) were collected for possible future use. It is recommended that future studies consider these data.

The time taken for each participant to take a pretest and posttest was recorded. Time spent on individual test questions was not collected or recorded. While time-on-task analysis was not conducted in the main study, future work could include analyzing if the time it takes to read and paraphrase refutation text affects declarative knowledge, conceptual knowledge, and procedural knowledge performance. An interesting position regarding time spent on refutation text is that due to the interest sparked by its structure and tone, readers actually spend less time reading refutation text, but experience more conceptual change than when engaging with non-refutation text (Broughton et. al., 2010). Further studies that focus on this phenomenon are recommended.

In addition, since one of the intentions of the main study was to determine ease of use of refutation text by instructors, it is recommended that future work include improving scoring techniques. Regarding the conceptual knowledge test, a less complicated scoring technique would have been to remove mutual dependence of the paired concepts; that is, if a learner mismatched one pair, he or she would automatically mismatch a second pair (H. Suen, personal communication, March 18, 2013). The test could therefore have been made less difficult by removing the mutual dependencies of paired concepts (H. Suen, personal communication, March 18, 2013). An instructor with limited time would find it easier to score the tests and provide feedback in a timely manner.
To examine how some participants went through the process of paraphrasing refutation text effectively enough to significantly improve conceptual knowledge posttest performance requires insight into their choice of paraphrasing style. Although examining how participants paraphrased exceeded the scope of the main study, results suggest that paraphrasing increased understanding of learners’ conceptual knowledge. Nuanced examination of the paraphrasing process requires further study.

Finally, future work should attempt to identify features of students’ misconceptions in terms of what Chi et al. (2012) characterizes as one of the following: systematic, robust, consistent, persistent, homogeneous, and recapitulated (refer to the section “Refutation Text and Conceptual Change” in Chapter 1). Determining that misconceptions among a sample of students are robust, for instance, may require that intervening instruction be tailored toward refutation text that targets robust misconceptions most effectively.

**Conclusion**

Results of the main study indicated that refutation text had a negative influence on declarative knowledge performance and positive influence on conceptual knowledge performance. The influence of refutation text on procedural knowledge performance did not result in statistically significant differences. Combinations of significance levels and effect sizes suggest that paraphrasing refutation text improved some students’ learning outcomes. More specifically, results suggest that paraphrasing refutation text, when
compared with merely reading without paraphrasing, contributed toward repairing some learners’ misconceptions about relational database design.

Based on the main study’s results and previous studies, refutation text that succeeded in repairing misconceptions effectively activated cognitive conflict that led to improved scores. The transformation from misconception to a level of understanding that holds up to normative interpretations implies that something changed. Overall results can be partially attributed to engaging with refutation text in ways that successfully highlighted misconceptions and helped repair or remove them. Assuming an absence of additional behavioral and environmental confounding factors that may have led to the repair of misconceptions among learners, it can be postulated that paraphrasing refutation text played a significant role in repairing learners’ intuitive errors.

While paraphrasing as a learning strategy is commonplace, the main study demonstrated that combining paraphrasing with refutation text is a practical form of instructional design that may help activate cognitive conflict and reduce misconceptions of subject matter. An extensive literature search, using the key terms and all variations of “refutation text” and “paraphrase”, failed to produce studies that explicitly paraphrased refutation text. In addition, no study was found that paraphrased refutation text and tested declarative knowledge performance, conceptual knowledge performance, and procedural knowledge performance simultaneously. Commenting on one of the types of knowledge (conceptual) tested in the main study, Chi (2008) stated that “no research has investigated confrontation at the categorical level” (p. 77). It was ultimately concluded
that this may be the first study that combined conceptual change-oriented refutation text with paraphrasing and simultaneously tested for three types of knowledge.

In summary, this research extends conceptual change-oriented literature by showing how combining paraphrasing with refutation text reduces misconceptions. This assertion applies mainly to conceptual knowledge performance and, to a lesser extent, procedural knowledge performance. Secondly, the main study’s findings confirmed previous research that has lower forms of knowledge (e.g. declarative knowledge) negatively affected while higher forms of knowledge (e.g. conceptual knowledge and procedural knowledge) positively affected by conceptual change-based learning interventions.
REFERENCES


http://ir.canterbury.ac.nz/handle/10092/7399


APPENDIX A. REFUTATION TEXT

E-R approach and Relational Data Modeling

Some students believe the entity-relationship (E-R) approach is based on the relational data model, or even more worrying, that it is a data model.

But this belief is false.

The E-R approach has nothing to do with the relational data model. It is simply a design methodology that can be applied to any model, including the relational data model. The concept of "relation" means different things to the E-R approach and the relational data model.

Logical versus Physical Database Modeling

Some students go for years confusing logical and physical database modeling. They confuse which activities/processes are possible between the two models.

This false belief must be resolved in students' minds.

Physical modeling depends on the specific RDBMS being used, but logical modeling deals with business requirements, processes, and data flows that are independent of the RDBMS. While you can define keys, constraints, etc., during logical modeling, you actually create these based on the software and hardware.

Normalization

Some database designers believe 3NF must be achieved for a model to be considered functional, or even relational. They are taught that 3NF is the minimum requirement.

But this understanding is not true.

1NF is enough for a database model to be considered relational. While higher NFs are preferred, understand that there begins to be a trade-off between update (improves) and retrieval (slows) performances. So yes, redundancy CAN be desirable if it is CONTROLLED effectively. Optimization is the goal.
Composite Keys

Some students believe composite keys have to be unique. Some also stop after combining two attributes when creating composite keys.

But this is a misconception. Firstly, composite keys can be unique or non-unique, just like one-column keys. Secondly, composite keys can be a combination of two or more attributes. For example, if LastName and NameOfMajor are not unique, you can add a third attribute to make a unique composite key.

Impact of relational databases on storage

Some students believe the layout of a relational model after normalization automatically means more storage will be required.

But this is not true. Precisely due to efficiencies gained from normalization, a properly set up relational database eliminates data redundancy. In addition to resolving many other potential anomalies, relational database minimizes storage requirements.

Primary Key

Some students carry several misunderstandings about keys, especially the primary key.

But it is vital to understand keys. Every relation must have a key, even if the key belongs to another relation (in the case of weak entity sets). Crucially, no two rows can agree on all attributes, but they can have the same attributes in some of the columns.
Column or field or attribute?

Some students advance through the world of database design with wrong notions of concepts such as fields or relations. Unfortunately, these are common misconceptions very early in your relational database knowledge acquisition, understand what the different concepts mean. (Tables, files, relations) and rows, records, tuples) and (columns, field, attribute) are often mixed up. Learn to use them in their appropriate contexts.

Null Values

Some students often confuse the meaning and use of NULL. Their understanding is that null means nothing, or worse still, zero. But this is not true. Null does not mean nothing or zero. It means the value is unknown (but the value exists!), or the value is inapplicable (e.g. of pregnancies for a man), or the value has been withheld (the boss’s salary). Unknown is especially tricky, but consider it like you would a half-truth.

Schema

The concept of schema is often misunderstood. Some database designs do not quite understand the nature and boundaries of a schema. But you can’t be a good designer without understanding schemas. At a basic level, understand that a schema is a relation and its set of attributes. But schemas get more involved than this; the key is to think of organized structures such as catalogs (sets of schemas) and clusters (collections of catalogs). Understanding this makes you a better designer.
DDL or DML or TCL or DCL or…?

Some students blindly issue SQL commands without pausing to think how the statements are logically organized according to function. For some DROP is the same as DELETE and ALTER equivalent to UPDATE.

But these are inexcusable misconceptions.

Know exactly how the SQL commands you use are classified. By knowing the differences between TCL, DCL, DML, and DDL commands, you will be more efficient in your database work. You will avoid embarrassing situations of trying to rollback after committing your updates.
APPENDIX B. DECLARATIVE KNOWLEDGE TEST QUESTIONS

Which of the following statements is true regarding logical versus physical database modeling? (There may be more than one correct answer)

- Logical modeling focuses on business requirements, whereas physical modeling focuses on defining attacks at the schema level.
- Primary keys, foreign keys, check constraints, and indexes are created during logical modeling, but business processes and data are defined during physical modeling.
- Physical modeling visualizes business activities and relevant data, whereas logical modeling creates views from database tables.
- A physical model partly depends on the database software in use, but a logical model is created independently of a specific database system.

Which of the following statements is true regarding normalization? (There may be more than one correct answer)

- 3NF is required for a model to be considered relational.
- Relational databases must be in third normal form.
- Normalization involves identifying candidate keys and functional dependencies.
- The Third Normal Form can be a disadvantage for retrieval but an advantage for updates.

Which of the following statements is true regarding the Entity-Relationship Approach? (There may be more than one correct answer)

- The E-R approach is based on the relational data model.
- The E-R approach is a design methodology.
- The E-R approach is a data model.
- Entities and relationships are the two main components of the E-R approach.

Which of the following statements is true regarding keys? (There may be more than one correct answer)

- A database will not function unless composite keys are unique.
- Surrogate keys are more useful to users than primary keys.
- The primary key concept is unnecessary.
- A key is a set of attributes that uniquely identifies an entity.

Which of the following statements is true regarding relational systems? (There may be more than one correct answer)

- Relational databases consume too much storage due to redundancy.
- Data integrity is not possible with relational systems.
- Relational systems work for querying data during testing, but not in production.
- In a relational system, all rows have a primary key.
APPENDIX C. CONCEPTUAL KNOWLEDGE TEST

Your task is to identify concepts that relate the most to each other. Once you find a match, select a radio button at the intersection of the concepts you think are closest in meaning or function. You will make 18 such comparisons. No concept matches with more than one other concept.
APPENDIX D. INFORMED CONSENT FORM

Implied Informed Consent Form for Social Science Research
The Pennsylvania State University

Title of Research: Confronting Misconceptions: Instructional Effects of Refutation Text on College Students' Learning Outcomes

Principal Investigator:
General M. Ndshalitshali, Graduate Student
314 Keller Building
University Park, PA 16802
(917)209-9006; general@psu.edu

Advisor:
Roy Cranana, PhD
314 Keller Building
University Park, PA 16802
(814)965-1958; rbc4@psu.edu

1. Purpose of the Study: The purpose of this research is to study the effects of refutation text on learning outcomes of college students.

2. Procedures to be followed: You will be asked to answer a series of questions regarding database design concepts. In addition, you will be asked to answer questions regarding self-efficacy, critical thinking, and metacognitive self-regulation.

3. Duration/Time: The study is expected to take approximately 25 minutes.

4. Statement of Confidentiality: Your participation in this research is confidential. In the event of any publication or presentation resulting from the research, no personally identifiable information will be shared because your name is in no way linked to your responses.

5. Right to Ask Questions: Please contact General Ndshalitshali at general@psu.edu with questions or concerns about this study.

6. Payment for Participation: You will receive extra credit points as agreed by your instructor. There is another option to participating to receive the extra credit; please email general@psu.edu to be sent a database design topic and prepare a one-page reaction to it.

7. Voluntary Participation: Your decision to be in this research is voluntary. You can stop at any time. You do not have to answer any questions you do not want to answer.

You must be 18 years of age or older to take part in this research study.

Completion and return of the survey implies that you have read the information in this form and consent to take part in the research. Please keep this form for your records or future reference.
APPENDIX E. DETAILED DATA COLLECTION PROCESS

Beginning of Data Collection Period

Preparation:

- Obtained permission from two instructors to conduct research with their students
- Requested and obtained access to class roster
- Combined students from both classes to create one list of potential participants
- Used Microsoft Excel’s randomization function to divide the list of students into equal size groups

Day 1 began:
Activities in class:

- The researcher attended the first class to begin administering the study.

- The researcher briefed participants about the study. (A single slide explaining the study had already been emailed to the participants a few days prior to testing. In addition, the instructors had already mentioned the upcoming study to their respective classrooms.)

- After answering questions about the research process, all participants were asked to access their emails (via the institution’s Learning Management System) and ‘click on’ a hyperlink that took them to the university’s data collection online site (Qualtrics).

- The first thing that participants saw on their computer screens was a brief introductory sentence with a hyperlink to an informed consent form.

- After reading the informed consent form, all participants were given approximately 45 minutes to take the declarative, conceptual, and procedural knowledge tests.

- To identify each participant without using personal information that may be linked directly to a participant, one question asked for the last 4 digits of a participants’ cellphone (or telephone) number. They were asked for the same number in the posttest so that the two tests could be linked.

- After completing a short demographic survey (gender, number of semesters), the declarative knowledge test showed on the computer screen (one question at a time).

- After completing the declarative knowledge test, the conceptual knowledge test showed on the computer screen.
After completing the conceptual knowledge test, the procedural knowledge test showed on the computer screen.

End of Declarative, Conceptual, and Procedural Knowledge Test:

Engagement with Refutation Text Treatment:

After completing the DK/CK/PK tests:

- The next screen showed the first of five refutation texts for each participant.
- Because engagement with refutation texts was different, the hyperlinks sent to the participants took them to 4 separate online testing areas within the main study’s test area in Qualtrics. The differences were as follows:
  - Participants assigned to RT group one were asked to read five refutation text items aligned with RT group one.
  - Participants assigned to RT group two were asked to read and paraphrase 5 refutation text items aligned with RT group one.
  - Participants assigned to group 3 were asked to read 5 refutation text items aligned with RT group two.
  - Participants assigned to group 4 were asked to read and paraphrase five refutation text items aligned with RT group two.
- The next question asked participants if they had read the refutation text delivered to them. A YES/NO option was provided.
- For purposes of awarding extra credit points, participants’ names, student IDs, and instructor names were requested.
(These personal details were never used for anything other than for awarding extra credit points. They were omitted from all data sets analyzed for the main study)

Day 1 ended:

Day 2 began:
• The next day, the researcher went to the second class of +50 students
• The same process as described above (see Day 1 research procedures) was applied

Day 2 ended:

Day 8 began:
• After one week, the researcher returned to the first class of +50 students for posttesting
• The same process as described above (see Day 1) was applied

Day 8 ended:

Day 9 began:
• The researcher returned to the second class of +50 students for posttesting
• The same process as described above (see Day 1) was applied

Day 9 ended:

End of Data Collection Period
# APPENDIX F. CONCEPT RATINGS

<table>
<thead>
<tr>
<th>Concept</th>
<th>Expert 1</th>
<th>Expert 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entity-Relationship approach</td>
<td>Easy</td>
<td>Moderate</td>
</tr>
<tr>
<td>Relational Data Modeling</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
<tr>
<td>Logical Database Modeling</td>
<td>Moderate</td>
<td>Difficult</td>
</tr>
<tr>
<td>Physical Database Modeling</td>
<td>Moderate</td>
<td>Difficult</td>
</tr>
<tr>
<td>Normalization</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
<tr>
<td>Composite Keys</td>
<td>Difficult</td>
<td>Moderate</td>
</tr>
<tr>
<td>Storage</td>
<td>Easy</td>
<td>Easy</td>
</tr>
<tr>
<td>Primary Key</td>
<td>Easy</td>
<td>Easy</td>
</tr>
<tr>
<td>Column</td>
<td>Easy</td>
<td>Easy</td>
</tr>
<tr>
<td>Field</td>
<td>Easy</td>
<td>Easy</td>
</tr>
<tr>
<td>Attribute</td>
<td>Easy</td>
<td>Easy</td>
</tr>
<tr>
<td>NULL</td>
<td>Moderate</td>
<td>Easy</td>
</tr>
<tr>
<td>Schema</td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
<tr>
<td>DDL</td>
<td>Difficult</td>
<td>Moderate</td>
</tr>
<tr>
<td>DML</td>
<td>Difficult</td>
<td>Difficult</td>
</tr>
<tr>
<td>TCL</td>
<td>Difficult</td>
<td>Difficult</td>
</tr>
<tr>
<td>DCL</td>
<td>Difficult</td>
<td>Difficult</td>
</tr>
<tr>
<td>Table</td>
<td>Easy</td>
<td>Easy</td>
</tr>
<tr>
<td>Row</td>
<td>Easy</td>
<td>Easy</td>
</tr>
<tr>
<td>Entity</td>
<td>Easy</td>
<td>Easy</td>
</tr>
<tr>
<td>Record</td>
<td>Easy</td>
<td>Easy</td>
</tr>
</tbody>
</table>
APPENDIX G. PROCEDURAL KNOWLEDGE TEST

Declare the relation:

FilmStudio(name, city, presCD)
whose primary key is name and which has a foreign key presCD that references cert# of relation:

MovieExec(name, city, cert#, netWorth).


VITA

General M. Ntshalintshali general@intellerptual.com

Summary of Experience

Circa 13 years full-time Information Systems experience in senior management, specialist, and consultant roles in several industries: financial services, oil/gas manufacturing, rail transportation, mobile telecommunications, and educational testing services.

Professional Experience

Strategic Management
- Hired new staff and managed 17 direct reports, including IT administrators, architects, analysts, and project managers
- Provided systems integration for a bank’s geographically dispersed item-processing environment
- Managed external service provider relationships, commercial contracts, and SLAs

IT Service Delivery
- Participated in negotiating Operations & Maintenances Statement of Works on behalf of service provider
- Coordinated SharePoint platform management with client on behalf of service provider

Financial Management
- Administered an Information Systems department budget for 4 years
- Helped reduce item processing costs for a retail bank by 25%

Project Management
- Delivered a Reporting Services project to an educational testing enterprise
- Delivered integrated imaging, signature verification, and document management systems of an item processing project

IT Operations and Maintenance
- Administered enterprise knowledge management applications (SharePoint)
- Administered enterprise finance solutions, including databases (Oracle Financials)

Work History

- 2013 – current, SharePoint Consultant, Educational Testing Services, Princeton, NJ
- 2009 – 2013 (part time), Teaching Assistant/Instructional Designer (during PhD study), Stony Brook, NY, and State College, PA
- 1999 – 2001, MIS Manager, MetroRail, Pretoria, Gauteng

Education

- Ph.D. in Learning, Design, & Technology : Penn State University, University Park, PA : May 2014
- MBA : The University of the Witwatersrand, Johannesburg, Gauteng : June. 2010

Publications (sample)


