Pedagogical Efficacy of Argument Visualization Tools

by

Hui Niu

Master of Arts, Simon Fraser University, 2006
Master of Engineering, Harbin Institute of Technology, 1998
Bachelor of Arts, Harbin Institute of Technology, 1997
Bachelor of Engineering, Harbin Institute of Technology, 1996

Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

in the Educational Technology & Learning Design Program

Faculty of Education

© Hui Niu 2016

SIMON FRASER UNIVERSITY

Summer 2016

All rights reserved. However, in accordance with the Copyright Act of Canada, this work may be reproduced, without authorization, under the conditions for “Fair Dealing.” Therefore, limited reproduction of this work for the purposes of private study, research, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.
Approval

Name: Hui Niu
Degree: Doctor of Philosophy
Title: Pedagogical Efficacy of Argument Visualization Tools
Examing Committee: Chair: Alyssa Wise
                    Associate Professor

John Nesbit
Senior Supervisor
Professor

Phil Winne
Supervisor
Professor

David Kaufman
Internal/External Examiner
Professor

Perry Klein
External Examiner
Associate Professor
Faculty of Education
Western University

Date Defended/Approved: July 27th, 2016
Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Update Spring 2016
Abstract

The purpose of this research was to investigate the instructional effects of argument visualization tools (AVTs) by developing and evaluating the use of a tool called the Dialectical Map (DM). In a laboratory experiment, each of 125 participants was randomly assigned to one of three groups: a DM Group that studied with the DM and received training on argumentation concepts, an Argue Group that received argumentation training only, and a Control Group that received no training and did not use the DM. Pre-test data were collected on participants’ basal free-recall ability and judgment of learning. After studying an expository text on fracking, participants gave a judgment of learning and were tested on critical thinking, recall and comprehension, and argumentative writing. Studying with the DM increased confidence in learning, recall and comprehension, and the use of argumentation in a writing task. In addition to the laboratory experiment, three semester-long classroom implementations were conducted with undergraduate students. Both the instructor and students expressed positive attitudes with their DM experiences. Findings from this thesis provide an insight into prior research on educational benefits of AVTs and have instructional implications for incorporating effective AVTs, such as the DM, in classrooms. Future research will continue on gathering data from multiple settings to improve the design and applications of the Dialectical Map.

Keywords: Computer mediated learning; argumentation; argument visualization tools; the Dialectical Map; argumentative writing; pedagogical efficacy
To the brain cells that bravely gave their lives in the writing of this thesis.
Acknowledgements

Is getting a PhD a big deal?

It depends. On one hand, it is undeniably the highest and most prestigious degree one can accomplish in a given field. On the other hand, there is still so much I don’t know. The more I learn about it, the more I realize this… so much so that I almost feel I don’t deserve this. And how is reaching this milestone going to change my life? Or in a sense is my life just starting after bidding my perpetual student status adieu? I will need to keep improving myself. Perhaps another PhD (do I ever learn?) and go through the ultimate pondering more than too often: WHY AM I HERE? It may take a lifetime to reach the answers, or longer, after I figure out whether time and space are actually just an illusion. At this moment, I would like to reflect on my long journey and thank all the people who have accompanied me while I take this important step forward.

First of all, I would like to express my most sincere gratitude to my senior supervisor, Dr. John Nesbit, for his selfless and constant support of my research and study. It was his outstanding academic reputation, profound knowledge of the field, and great personality that drew me to his program for a PhD in the first place. I’ll always appreciate this once-in-a-lifetime opportunity to have learned from and worked with him. I would also like to thank Dr. Phil Winne for his brilliant and insightful comments and advice. As a world-renowned scholar and busy executive, Phil always finds time for his students. During my educational career at SFU, both my supervisors supported me with a great amount of knowledge, patience, and compassion. I couldn’t ask for a better committee, it has been a true honour.

This may not be conventional, but I would like to mention that I am thankful to have Dr. Perry Klein as my external examiner, Dr. David Kaufman as my internal examiner, and Dr. Alyssa Wise as the chair for my defense.

I would also like to thank Joan Sharp and her students for their enthusiastic participation in using the instructional tool I developed. Their positive feedback gave me a great deal of confidence in carrying out further research beyond this thesis.
My research wouldn’t have been possible without my wonderful colleagues, especially Liam Doherty, for his exceptional talent, skills, and generosity with his valuable time; Arita Liu, for her prompt, professional assistance and friendship; and Zahia Marzouk, for coordinating the lab schedules and helping me greet research participants.

I’d also like to offer my special thanks to Drs. Carter, Jackson, and Hosken for being my inspiration, role models, and confidants. From them I learned how to be persistent and fearless, and more importantly, how to love and enjoy life while under pressure (butter chicken and lime bars are the key to success). With their bodacious, awesome adventures and quixotic philosophies, they convinced me that being a nerd is actually COOL.

For the time being, I am not entirely certain if I should thank the Muses for always visiting me at the weirdest hours, possibly due to the time differences. I appreciate all the breakthroughs those encounters had enabled. However, as a side effect, I am now afraid of natural light and allergic to fresh air. My gratitude remains arguable.

Last but not least, I am very grateful to have an amazing family and great friends who always supported and believed in me, even when I doubted myself… Looks like they were right.
# Table of Contents

Approval .................................................................................................................. ii
Ethics Statement ........................................................................................................ iii
Abstract ..................................................................................................................... iv
Dedication ................................................................................................................... v
Acknowledgements .................................................................................................... vi
Table of Contents ...................................................................................................... viii
List of Tables ............................................................................................................. xi
List of Figures ............................................................................................................ xiii

## Chapter 1. A Summary of the Research ................................................................. 1
1.1. Background ......................................................................................................... 1
   1.1.1. Argumentation .............................................................................................. 1
   1.1.2. Argumentation and Learning ....................................................................... 3
   1.1.3. Argument Visualization Tools (AVTs) ....................................................... 4
   1.1.4. Theories Germene to Argument Visualization and AVT ......................... 4
   1.1.5. Empirical Findings on AVT Applications ................................................... 6
1.2. Rationale .............................................................................................................. 7
1.3. Development of the Dialectical Map (DM) ......................................................... 8
1.4. Research Questions ............................................................................................ 9
1.5. Method ............................................................................................................... 11
1.6. Results ............................................................................................................... 12
1.7. Classroom Implementations .............................................................................. 13
1.8. Conclusion ......................................................................................................... 13
1.9. Thesis Structure ................................................................................................. 14

## Chapter 2. Literature Review .............................................................................. 15
2.1. What is Argumentation? ..................................................................................... 15
   2.1.1. The Toulmin Model ..................................................................................... 16
   2.1.2. Walton’s Argumentation Schemes .............................................................. 18
   2.1.3. Dialectical Argumentation ......................................................................... 20
2.2. Argumentation and Learning ........................................................................... 22
   2.2.1. Effects of Arguing to Learn ....................................................................... 22
   2.2.2. Teaching and Learning to Argue ................................................................ 28
       Why .................................................................................................................. 28
       What ............................................................................................................... 29
       How ............................................................................................................... 30
2.3. Argument Visualization Tools (AVTs) .............................................................. 36
   2.3.1. Why AVT? ................................................................................................. 36
   2.3.2. Theories Germene to Argument Visualization and AVT ......................... 38
       Cognitive Load Theory .................................................................................... 38
       Schema Theory ............................................................................................... 40
       Cognitive Tools Theory .................................................................................. 43
       Multimedia Learning Theory ......................................................................... 44
   2.3.3. Empirical Findings on AVT Applications ................................................... 45
       Description of Studies .................................................................................... 46
# Findings on Educational Benefits

Summary ........................................................................................................... 56

## Chapter 5. Results................................................................................. 96

5.1. Missing Values, Outliers, and Normality of Distributions ................. 96
5.2. Demographic Information .................................................................. 97
5.3. Self-Reported Information ................................................................. 98
5.4. Coded Short-Answer Responses and Inter-Rater Reliabilities .......... 101
5.5. Free Recall .......................................................................................... 104
5.6. The DM Uses ...................................................................................... 105
5.7. Highlighting and Time-on-Task (ToT) ............................................... 109
  5.7.1. Highlighting .................................................................................. 109
  5.7.2. Time-on-Task .............................................................................. 110
5.8. Critical Thinking Scores ..................................................................... 111
5.9. Summary (Q1) Scores ....................................................................... 111
5.10. Cued-Recall Scores ......................................................................... 116
5.11. Reasoning Scores ............................................................................. 118
5.12. Correlations ...................................................................................... 119
5.13. Multiple Regression Analyses ........................................................... 121

## Chapter 4. An Experiment on Learning with the Dialectical Map............ 79

4.1. Participants ......................................................................................... 79
4.2. Materials and Instruments ................................................................ 81
  4.2.1. The Text for Study ...................................................................... 81
  4.2.2. The Dialectical Map (DM) ........................................................... 81
  4.2.3. The Highlighting Tool and Server Logs in nStudy ...................... 82
  4.2.4. The Critical Thinking Measure ................................................... 83
  4.2.5. The Need for Cognition Measure ............................................... 84
  4.2.6. The Achievement Goal Measure ............................................... 85
  4.2.7. Judgment of Learning ................................................................. 86
4.3. Procedure .......................................................................................... 87
4.4. Piloting .............................................................................................. 89
4.5. Coding Schemes for Short Answers ................................................... 91
  4.5.1. Scoring Free Recall ................................................................... 91
  4.5.2. Scoring the Summary ................................................................. 91
  4.5.3. Scoring Cued-Recall and Reasoning Questions ......................... 94

## Chapter 3. Development and Deployment of the Dialectical Map .......... 64

3.1. Theoretical Basis and Design Principles of the DM ......................... 65
3.2. The DM in Action: A User Manual ..................................................... 69
3.3. Implementing the DM in Authentic Settings ..................................... 73
  3.3.1. The Instructor ............................................................................ 73
  3.3.2. Three Implementations of the DM .............................................. 73
  3.3.3. General Impression .................................................................. 74
  3.3.4. The Instructor Feedback ............................................................. 76
  3.3.5. Student Attitudes ...................................................................... 77

## Chapter 2. The Practicality of the Dialectical Map.............................. 23

2.1. Theoretical Basis and Design Principles of the DM ......................... 24
2.2. The DM in Action: A User Manual ..................................................... 28
2.3. Implementing the DM in Authentic Settings ..................................... 31
  2.3.1. The Instructor ............................................................................ 31
  2.3.2. Three Implementations of the DM .............................................. 31
  2.3.3. General Impression .................................................................. 32
  2.3.4. The Instructor Feedback ............................................................. 34
  2.3.5. Student Attitudes ...................................................................... 35

## Chapter 1. The Use of the Dialectical Map........................................... 1

1.1. Theoretical Basis and Design Principles of the DM ......................... 2
1.2. The DM in Action: A User Manual ..................................................... 7
1.3. Implementing the DM in Authentic Settings ..................................... 11
  1.3.1. The Instructor ............................................................................ 11
  1.3.2. Three Implementations of the DM .............................................. 11
  1.3.3. General Impression .................................................................. 12
  1.3.4. The Instructor Feedback ............................................................. 13
  1.3.5. Student Attitudes ...................................................................... 15

## Multiple Regression Analyses

Correlations

Summary (Q1) Scores

Critical Thinking Scores

Cued

Reasoning Scores

Demographic Information

Highlighting and Time

The DM Use

Rater Reliabilities

The Need for Cognition Measure

The Achievement Goal Measure

The Critical Thinking Measure

The Highlighting Tool and Server Logs in nStudy

The Text for Study

Student Attitudes

Theoretical Basis and Design Principles of the DM

The Critical Thinking Measure
List of Tables

Table 2.1  General Information of Studies ................................................................. 58
Table 2.2  Reliability and Validity Estimates ............................................................ 59
Table 2.3  Conditions of Research ............................................................................ 60
Table 2.4  Conditions and Outcomes of Research ...................................................... 62
Table 3.1  DM Scores ................................................................................................. 75
Table 3.2  Marking Key for DM Assignments ............................................................. 76
Table 4.1  Participants’ Field of Study ......................................................................... 80
Table 4.2  Participant First Language .......................................................................... 80
Table 5.1  Descriptive Statistics for Demographic Variables ...................................... 97
Table 5.2  Test of Homogeneity of Variances for Demographic Variables .................. 97
Table 5.3  Descriptive Statistics for Self-Reported Information ................................... 99
Table 5.4  Test of Homogeneity of Variances for Self-Report Variables ..................... 99
Table 5.5  Correlation Matrix of Self-Report Variables .............................................. 100
Table 5.6  Descriptive Statistics of Rater Agreements ............................................... 102
Table 5.7  Cohen’s Kappa for 16 Main Ideas .............................................................. 102
Table 5.8  Descriptive Statistics for Free Recall Variables ......................................... 104
Table 5.9  Correlations among Free Recall Variables ............................................... 104
Table 5.10 Test of Homogeneity of Variances for Free Recall Variables ..................... 105
Table 5.11 Descriptive Statistics of the DM Activities ............................................... 108
Table 5.12 Other DM Usages .................................................................................. 109
Table 5.13 Correlations Among DM Variables .......................................................... 109
Table 5.14 Descriptive Statistics for Highlighting ...................................................... 109
Table 5.15 Descriptive Statistics for ToT1 and ToT2 (Seconds) .................................... 110
Table 5.16 Descriptive Statistics for Critical Thinking .............................................. 111
Table 5.17 Descriptive Statistics for Q1 ..................................................................... 112
Table 5.18 Test of Homogeneity of Variances for Q1 Variables ................................. 113
Table 5.19 Kruskal-Wallis Tests for Q1 Variables ...................................................... 114
Table 5.20 Correlation Matrix for Q1 Variables ........................................................ 115
Table 5.21 Descriptive Statistics for Sum of Q2–Q6 .................................................... 116
Table 5.22 Test of Homogeneity of Variances ............................................................ 117
Table 5.23 Correlation Matrix for Cued-Recall Question Scores ............................... 118
Table 5.24  Descriptive Statistics for Reasoning Questions ........................................ 118
Table 5.25  Correlation Matrix of Self-Reported Information, Free Recall Idea Units, and Short-Answer Responses ......................................................... 120
Table 5.26  Contrast Coding Weights ........................................................................ 122
Table 5.27  Sequential Regression Table for Predicted Organization Scores .......... 125
Table 5.28  Sequential Regression Table for Predicted Argument Marker Scores .... 126
Table 5.29  Sequential Regression Table for Predicted Q1 Main Idea Scores .......... 127
Table 5.30  Sequential Regression Table for Predicted Cued-Recall Idea Units Scores ........................................................................................................ 128
Table 5.31  Sequential Regression Table for Predicted Cued-Recall Main Idea Scores ........................................................................................................ 129
Table 5.32  Sequential Regression Table for Predicted Total Reasons Scores ........ 130
Table 5.33  Sequential Regression Table for Predicted JOL2 Scores ......................... 131
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>An example of Dialectical Map</td>
<td>9</td>
</tr>
<tr>
<td>3.1</td>
<td>Structure of a Simple Argument Map</td>
<td>66</td>
</tr>
<tr>
<td>3.2</td>
<td>An Argumentation Vee Diagram</td>
<td>67</td>
</tr>
<tr>
<td>3.3</td>
<td>How to Create a DM Assignment</td>
<td>70</td>
</tr>
<tr>
<td>3.4</td>
<td>The Starting Interface of the DM</td>
<td>71</td>
</tr>
<tr>
<td>3.5</td>
<td>A Completed DM</td>
<td>72</td>
</tr>
<tr>
<td>3.6</td>
<td>The DM and the Export Content</td>
<td>73</td>
</tr>
<tr>
<td>4.1</td>
<td>The Dialectical Map Interface</td>
<td>82</td>
</tr>
<tr>
<td>4.2</td>
<td>The Highlighting Interface in FluidSurvey Embedded in nStudy</td>
<td>83</td>
</tr>
<tr>
<td>4.3</td>
<td>Experiment Procedures</td>
<td>89</td>
</tr>
<tr>
<td>4.4</td>
<td>Word Counts for All Short-Answer Responses</td>
<td>95</td>
</tr>
<tr>
<td>5.1</td>
<td>Judgment of Learning Repeated Measures</td>
<td>101</td>
</tr>
<tr>
<td>5.2</td>
<td>Cohen’s Kappa for 20 Idea Units</td>
<td>103</td>
</tr>
<tr>
<td>5.3</td>
<td>Cohen’s Kappa for All 16 Main Ideas</td>
<td>103</td>
</tr>
<tr>
<td>5.4</td>
<td>A Completed DM</td>
<td>107</td>
</tr>
<tr>
<td>5.5</td>
<td>A Zoom-in Image of Figure 5.4</td>
<td>108</td>
</tr>
<tr>
<td>5.6</td>
<td>Word Count for Q1</td>
<td>112</td>
</tr>
<tr>
<td>5.7</td>
<td>Error Repeated Measures</td>
<td>114</td>
</tr>
<tr>
<td>5.8</td>
<td>Word Counts for Q2–Q6</td>
<td>117</td>
</tr>
<tr>
<td>5.9</td>
<td>Word Counts for R1–R5</td>
<td>119</td>
</tr>
</tbody>
</table>
Chapter 1.

A Summary of the Research

This chapter summarizes the thesis with an emphasis on its research questions. The chapter a) presents the theoretical and empirical context, b) formulates the rationale of the research by highlighting theoretical and practical gaps in prior research, c) describes the Dialectical Map I developed to address those gaps, d) outlines the method, results, and conclusions of the research conducted with the Dialectical Map software, and e) summarizes classroom implementations of the Dialectical Map. The chapter concludes with an outline of the thesis structure.

1.1. Background

1.1.1. Argumentation

Argumentation is a verbal and social activity of reason aimed at increasing (or decreasing) the acceptability of a controversial standpoint for the listener or reader, by putting forward a constellation of propositions intended to justify (or refute) the standpoint before a rational judge (van Eemeren, Grootendorst, & Henkemans, 1996, p. 5).

The Toulmin Model

The argumentation field has been heavily influenced by Toulmin’s idea that formal logic is irrelevant when considering the validity of an argument in practice (van Eemeren et al., 1996). Supporting evidence and proper justifications carry more weight. Consisting of six elements, specifically claim, data, warrant, backing, modal qualifier, and rebuttal (these terms are defined in Chapter 2), the Toulmin model lends itself to the construction, analysis, and evaluation of arguments (van Eemeren et al., 1996). Toulmin posited that the basic structure of argumentation is field independent, but components
such as warrant and backing are field variant. For example, the acceptability of a warrant in law is dependent on legal rules and a backing in science is dependent on scientific principles.

**Walton’s Argumentation Schemes**

Walton contended that argumentation is a procedure that contains a starting point, a middle stage, and an end point. The middle stage contains dialogues between agents who take turns asking questions and making assertions. The goal of logical argumentation is not to confirm or deny the veracity of something, but to make decisions based on arguments both pro and contra the claim. Arguments are defeasible and subject to critical questioning until the conclusion of the argumentative discourse. Logical argumentation is manifested via a set of argumentation schemes. Walton (1996) identified 29 commonly used argumentation schemes with corresponding critical questions that respondents of an argument should ask to challenge the proponent of an argument. Critical questions are essential in evaluating the cogency of an argument. The argumentation schemes offer specific models that may be useful in didactic and academic settings.

**Dialectical Argumentation**

Dialectical argumentation involves synthesizing and evaluating different views. During dialectical argumentation, arguers are exposed to a variety of ideas related to the topic either supporting or opposing their own arguments. Such exposure is thought to augment learning (Means & Voss, 1996). Dialectical argumentation would fall in Kuhn’s dialogical category (1991) realized through a discourse between those who advocate a thesis and those who hold an antithesis. Dialectical argumentation, in practice, can be solitary (e.g., weighing pros and cons to make a decision) or a social activity (e.g., a discussion within a social setting; Driver, Newton, & Osborne, 2000). The goal of dialectical argumentation is to resolve contradictions and inconsistencies (Barth & Krabbe, 1982; van Eemeren & Grootendorst, 1992). Resolutions vary. For example, when arguments are contentious, the resolution may involve convincing opponents to accept one’s claim. However, dialectical argumentation need not always be adversarial. Arguers may also seek a compromise between their claims.
1.1.2. Argumentation and Learning

A very high proportion of academic subjects and intellectual endeavours are founded on the ability to identify, construct, and evaluate arguments. Argumentation generally involves stating a claim and providing evidence and reasons to support the claim. For skilled arguers, counterarguments are anticipated, identified, and refuted (Nesbit, Niu, & Mao, 2008). The relationship between argumentation and learning is commonly considered to operate in two directions: a) argumentation has instructional value, which is often discussed as arguing to learn in the literature, and b) argumentation is an important curricular goal in its own right, often abridged as learning to argue.

Effects of Arguing to Learn

Academic subjects such as science (Nussbaum & Sinatra, 2003), mathematics (Schwarz, Neuman, & Biezuner, 2000), and social studies (Wiley & Voss, 1999) can incorporate argumentation into their curricula with the goal of enhancing learning. As an instructional strategy, argumentation has great potential in promoting learning for several types of intended outcomes such as recall of information (Reznitskaya, Anderson, & Kuo, 2007), understanding subject matter (Wiley & Voss, 1999), critical thinking (Sanders, Wiseman, & Gass, 1994), and writing performance (Ferretti, MacArthur, & Dowdy, 2000). Argumentation has also been shown to boost students’ motivation to learn (Chinn, Anderson, & Waggoner, 2001). Increasingly, educators consider argumentation as an effective instructional strategy that should be more broadly adopted within schools (Nesbit et al., 2008).

Teaching and Learning to Argue

Kuhn (1991, 1992) positioned argumentation at the heart of higher-order thinking in the real world. However, students frequently display poorly developed argumentation skills. Researchers have recognized that argumentation skills are not naturally acquired. Similar to other skills, argumentation proficiencies improve with practice and feedback (Kuhn, Goh, Iordanou, & Shaenfield, 2008; Kuhn & Udell, 2007). A few instructional strategies appear to have helped students become sophisticated arguers. For example, group discourse was found to help students include more arguments, counterarguments,
and rebuttals in their essays (Reznitskaya et al., 2001). Argumentation goal instructions were also found to benefit students in producing more counterarguments and rebuttals (Nussbaum & Kardashian, 2005). Another strategy to improve students’ argumentation skills included teaching learners the skills and concepts of argumentation (Dawson & Venville, 2010). Additionally, researchers have been investigating cognitive tools (defined in Chapter 2) and their role in facilitating students’ argumentation skills. These tools are often found beneficial in argument-related learning such as promoting more productive arguments (Lu & Lajoie, 2008) and more in-depth arguments (Teo & Churchill, 2007).

1.1.3. Argument Visualization Tools (AVTs)

AVTs are graphical tools used to model the content and the inferential structure of arguments as an alternative to traditional prose-only representations. Such visualizations contain both graphical and textual elements in order to visualize components of arguments and the links among them (Andriessen, Baker, & Suthers, 2003; Twardy, 2004; van Gelder, 2003; Walton, 2013). An argument map is a typical AVT composed of boxes that contain statements or propositions and arrows that indicate the relations between the propositions.

1.1.4. Theories Germane to Argument Visualization and AVT

Cognitive Load Theory

Cognitive load is the total mental resource required of working memory when an individual performs a cognitive task, such as learning or solving problems (Sweller, 1988; Sweller, Ayres, & Kalyuga, 2011). Cognitive load theory (CLT) surmises that human cognition has limitations when processing information. For an AVT to be effective, the designer must follow principles of how the human cognitive system functions. Argumentation tasks generally demand more cognitive resources than other tasks (Bernardi & Antolini, 1996; Coirier, Andriessen, & Chanquoy, 1999; Kuhn, 2005; Nussbaum, 2008). Instructional designers need to bear in mind that using AVTs benefits
learning so long as the cognitive cost associated with the graphical representation is outweighed by the benefits brought forth by the visualization.

**Schema Theory**

Schemas are hypothetical mental configurations that contain a set of related categories, relationships among the categories, and criteria identifying what's important within the configurations (Anderson, 1984; Rumelhart, 1980; Sweller, 1994). Much of human knowledge is stored in long-term memory as schemas, which can be activated for cognitive tasks such as comprehension and problem solving (Rumelhart, 1980). AVTs are hypothesized to facilitate learning that involves argument schemas. Researchers believe that argument schemas bring many educational benefits (Reznitskaya & Anderson, 2002, p. 321). Since students often lack argumentation skills, one possible explanation is the lack of argument schemas in their knowledge base, particularly advanced schemas (Hidi & Hildyard, 1983; Knudson, 1992a; Scardamalia & Bereiter, 1986).

**Cognitive Tools Theory**

The term cognitive tool is often used to identify an interactive computer program designed for schema construction, activation, or instantiation (Nesbit, Niu, & Liu, in press). Lajoie (1993, p. 261) rationalized four reasons how and why cognitive tools assist learning. Each reason cited applies equally to the use of AVTs in scaffolding argumentation. These include:

- a) Cognitive tools support cognitive processes, such as memory or metacognitive processes;
- b) They share or free cognitive load required by lower-level tasks so that more cognitive resources are available to develop higher-order skills;
- c) They facilitate learners in performing cognitive tasks that would be over-challenging without the tool;
- d) They allow learners to generate and test their hypotheses.
Multimedia Learning Theory

Multimedia learning can be defined as “building mental representations from words and pictures” (Mayer, 2005, p. 2). The integration of graphical and textual elements is an important feature of AVTs. Thus, AVTs are inherently a type of multimedia. Most research-based multimedia learning design principles would consequently be valuable for developing effective AVTs. Multimedia design principles are based on three instructional goals: 1) reducing extraneous processing that is unrelated to learning; 2) managing essential processing in working memory that involves selecting and organizing relevant information; and 3) fostering generative processing that involves reorganizing information and integrate it with prior knowledge. These instructional goals produced 15 design principles (Mayer, 2014). Most of the principles apply to the design of AVTs, such as providing cues that highlight essential material, presenting coherent information with minimal distractions, and presenting associated words and pictures in close proximity.

1.1.5. Empirical Findings on AVT Applications

AVTs are now enjoying increasing research attention in education, where they are being investigated as scaffolding to improve learning and argumentation skills. Learning activities ranged from reading (e.g., Dwyer, Hogan, & Stewart, 2010) or writing tasks (e.g., Nussbaum & Schraw, 2007) to argumentation and critical thinking exercises (e.g., Butchart et al., 2009; Carr, 2003). Many studies reported that argument visualizations improved a measured outcome such as critical thinking (van Gelder, Bissett, & Cumming, 2004), retention (Salminen, Marttunen, & Laurinen, 2010), motivation (Zumbach, 2009), conceptual learning (Lund, Molinari, Séjorné, & Baker, 2007), as well as argumentation and reasoning skills (e.g., Nussbaum, 2008; Harrell, 2011). AVTs were also found to enhance collaborative learning, such as producing more information elaboration in collaborative writing (Suthers & Hundhausen, 2003). Caution is imperative when drawing conclusions about the benefits of argument visualization. The efficacy of argument visualization interacts with the design features of a particular implementation, such as the student characteristics, instructional context, and the technical features of the AVT.
1.2. Rationale

Many students, including those at the university level, continue to lack skills to analyze and construct arguments. There is evidence that university students do not naturally develop or apply advanced argumentation skills (e.g., generating warrants to justify evidence or identifying and rebutting counterarguments) unless explicitly prompted by task instructions and supported by the learning environment (Asterhan & Schwarz, 2007; de Vries, Lund, & Baker, 2002; Harrell, 2011). Researchers discovered that when given scaffolds for an argumentation exercise, such as argument mapping, students no longer had difficulty coordinating theories and evidence and demonstrated deeper engagement with the evidence items. They refrained from giving superficial descriptions and used more causal warrants (Bell, 2001). Students also recognized arguments and their logic structures better (Harrell, 2005). Such argument scaffolding also improved students’ ability to evaluate arguments and counterarguments for an integrated and well-reasoned conclusion (Nussbaum, 2008).

Argument visualization is potentially an effective scaffold for argumentation skills. The AVT literature involves various research methods, learning activities, media, and instructional outcomes. Although the results are not conclusive, there is evidence that AVTs have instructional merits, which are manifested through student improvements in motivation, learning performances, and the ability to argue. Some effects, however, appear to be dependent on their execution conditions. As a fairly new topic, there is a shortage of empirical research to resolve the claims and questions that have motivated researchers. Therefore, more empirical research is needed to verify such claims and develop guidelines for practice, especially for teachers, on how to use the tools in the classroom. Research is also needed to inform instructional design on how to set up learning environments that better engage and motivate students in using the tools for learning activities.

In addition, I found that researchers often did not fully report important features of their research such as treatment implementation, score reliability, and method of data analysis. The research also was rarely able to assess the effects of the particular features of each AVT, such as their interactive capabilities or graphical layout, although the potential impact of these features on results was occasionally discussed by
researchers. Since AVTs bear many design features, and the effects of these features may interact, there may be a need for more design-based research that is coordinated with experimental research. For these reasons, in my thesis research, I led the development of the Dialectical Map and gathered data on its role in instruction and learning.

1.3. Development of the Dialectical Map (DM)

Many AVTs focus on representing the evidential relations among components of an argument and offer little to no guidance on how to present both sides of the argument or how to resolve arguments and counterarguments. Lack of such scaffolding may work for skilled arguers, who tend to habitually synthesize and balance arguments and counterarguments, but for less skilled arguers the lack of scaffolding may lead to confirmation bias (i.e., attending only to information that supports their position). Although the Argumentation Vee Diagram (Nussbaum, 2008; Novak, Gowin, & Johansen, 1983; Novak & Gowin, 1984) supports argument-counterargument integration, it lacks scaffolding for other important aspects of advanced argumentation. For example, it offers no clear representation of a) warrants, b) the hierarchical substructure of arguments whereby evidence can support reasons that in turn support claims, c) the differing strengths of supporting reasons or evidence, and d) the ordering of reasons for rhetorical purposes.

To address these design challenges shared by many AVTs, I and my colleagues developed a new type of argument map called the Dialectical Map (Niu, Sharp, Nesbit, & Doherty, 2015). The dialectical map is an interactive web-based tool written in JavaScript. An instructor or student can enter a primary claim or question at the top (see Figure 1.1 for an example). The reasons supporting the primary claim are organized on the left side, and those contradicting it are entered on the right side. Users enter reasons, evidence, and warrants in web page text boxes, rating strengths of reasons, linking arguments and counterarguments, and annotating whether or not a piece of evidence is supporting or opposing a particular reason. A conclusion that integrates the positions of the pro and con sides can be entered at the bottom. Users can put reasons in sequential order, perhaps to match how they will be presented in an argument essay.
to be written later. To aid integration and resolution of pros and cons, pro reasons can be bound to related con reasons so that they travel together when one of them is pulled higher or lower on the page.

Figure 1.1 An example of Dialectical Map

1.4. Research Questions

This thesis reports the research and development of a special AVT, the Dialectical Map. The following research questions were formulated to guide me in conducting the investigation on the pedagogical efficacy of the DM.

1. How does using the DM affect learners’ studying process?

The reason to investigate this question is that some behaviours (such as highlighting) may indicate engagement and interactions with the DM and how using the
DM may alter the learning process. Understanding how the DM affects the learning process may help to explain any outcome effects I observe.

2. Does using the DM affect judgment of learning (JOL)?

The reason to investigate this question is that no prior study has examined how AVTs affect JOL and learners’ JOL relates to motivational indices such as task persistence. It is reasonable to postulate that since argumentation has the potential to promote deeper cognitive engagement and understanding, using the DM may modify learners’ judgments of learning.

3. Does using the DM affect critical thinking skills?

Research has shown that critical thinking skills are closely connected to argumentation activities (Butchart et al., 2009; Carr, 2003; Carwei, 2009; Davies, 2009; Harrell, 2011; van Gelder et al., 2004). AVTs have also shown positive effects on learners’ critical thinking skills. It is useful to further investigate how using the newly developed DM may affect students’ critical thinking skills.

4. Does using the DM affect recall and comprehension?

Recall and comprehension are important measures to assess learning. Prior research has yielded positive but somewhat inconsistent evidence (see Chapter 2 for more details) that AVTs enhance recall and comprehension.

5. Does using the DM affect the ability and tendency to engage in argumentative writing?

Argumentative writing expresses the argumentative moves learners make as they complete writing tasks. I hypothesize that using the DM develops an argument schema (Reznitskaya & Anderson, 2002), which enhances the learner’s ability and tendency to engage in argumentation. The increased ability and tendency to engage in argumentation may be expressed in completing a writing task that is not explicitly argumentative. Argument schemas may also be activated by the exposure to the
concept of argumentation and the DM and cause learners to pay more attention to information that helps them build an argument.

1.5. Method

Participants were 125 university students randomly assigned to three conditions: a) the DM condition, where participants received training on the concept of argumentation as well as how to use the DM, b) the Argue condition, where participants received only verbal training on argumentation, and c) the Control condition with neither training nor the DM. Participants were then given two texts to study. Text 1 was used to collect pre-test data on free recall (Free Recall) and judgment of learning (JOL1). Text 2 was given to participants to study for a test that involved summarization of the material, cued-recall questions, critical thinking questions, and reasoning questions. I investigated the effects of studying an extended text passage using the DM for representing arguments. In particular, the study examined the effect of using the argument tool on participants’ recall and comprehension, critical thinking, judgment of learning, and argumentative writing. The participants in the research wrote a set of tests after they finished studying the materials.

Data were collected for the following variables:

1) Self-reported demographic variables including gender, age, GPA, field of study, and number of years of schooling with English as the primary language of instruction

2) Self-reported individual differences via two questionnaires:
   • Need for Cognition: inclination toward effortful cognitive activities (18 items)
   • Achievement Goal Orientation: motivation toward academic tasks (9 items)

3) Self-directed studying in an online studying environment

Participants used nStudy (Winne, 2010) to study the materials. The amount and mode of their use of nStudy and the DM was recorded in detail. Specifically, data about the timing, frequency and type of the following operations were collected:

   • Viewing of document
• Information highlighted within learning materials
• The use of the DM, which was administered and logged under nStudy: reasons created, evidence created, warrants created, links created, and repositioning of reasons

4) Responses to a six-item Critical Thinking Test, which was developed for this research

5) Judgment of learning for Text 1 (JOL1) and of Text 2 (JOL2)

6) Measures of learning. Participants were asked to complete a series of tasks:
   • Free recall after studying Text 1
   • Summarization of Text 2
   • Cued-recall of Text 2 (short answers)
   • Reasoning about Text 2 (short answers)

1.6. Results

There were four main outcomes of the research:

A) The DM Group spent more time studying the main text (Text 2) than both other groups.

B) Immediately after studying the main text, the participants were asked to judge how much they had learned while studying. Participants from the DM Group demonstrated higher confidence of their learning than those in the Control Group. After statistically removing the effects of years of English as the primary language for schooling (Years in English), JOL1, Need for Cognition, and Free Recall Idea Units, JOL2 for the DM Group was greater than both the other two groups.

C) The participants wrote a summary of the studied text and responded to several cued-recall questions about the studied text. The DM Group recalled statistically significantly more Main Ideas than the Control Group for the cued-recall questions. After statistically controlling for the following four factors, Years in English, JOL1, NfC, and Free Recall Idea Units, the DM Group outperformed both the other two groups on this measure. The DM Group also used more Main Ideas in their summary than the Argue Group. After statistically controlling for the above four factors, the DM Group used more Main Ideas than both the other two groups.

D) The DM Group wrote summaries whose organization more closely resembled an introduction-body-conclusion format, perhaps reflecting the ordering of the primary claim, reasons, and conclusion in the DM structure. The DM Group also outperformed the Argue Group on the
use of Argument Markers, such as “because,” “however,” “furthermore,” and “evidence,” after controlling for Years in English, JOL1, NfC, and Free Recall Idea Units. Moreover, fewer Neutral Idea Units were included in the DM Group’s summarization compared to the Control Condition. Neutral Idea Units by definition do not contribute to either side of the main argument.

1.7. Classroom Implementations

Currently AVTs are primarily investigated in laboratory settings. Some research has been conducted in the classroom, but there have been very limited semester-length full integrations of AVTs. The DM was used in teaching three undergraduate, writing-intensive, biology courses at Simon Fraser University (SFU; Niu et al., 2015). In three separate assignments, students drew from assigned readings to independently construct a DM. The three DMs were scored by the instructor with emphasis on the students’ ability to identify and apply relevant evidence to support arguments. Students were given detailed feedback on the quality of their evidence and arguments. The biology topics covered in the DMs were assessed later in the courses. In each course, the instructor reported increased quality in argumentation, represented within the students’ DMs, across all three assignments. Student’s comments about the instructional use of the DM tool were generally positive, and many claimed that it helped them to learn or improve their argumentation skills. These results complemented results from my laboratory study and showed that the DM had great potential in facilitating student learning in various settings.

1.8. Conclusion

The thesis offers better understanding of the relationships among argumentation, AVTs, and learning. It makes theoretical and practical contributions in supporting the claims that AVTs facilitate learning. For example, it provides empirical evidence that students can benefit from having a well-developed argument schema and that studying with an AVT can facilitate the acquisition and development of such a schema. Particularly, the DM showed promise in improving learning in several areas: recall, comprehension, and argumentative writing. It addresses several design challenges
many AVTs encounter. The thesis also provides sufficient basis for instructional designers to further explore the applications of the DM in multiple settings. Findings from this work can help instructors, instructional designers, and students to better apply, advance, and learn with AVTs.

1.9. Thesis Structure

This thesis is organized into six chapters. Chapter 1 provides the background and rationale for the research in this thesis to demonstrate the significance of this thesis. It also outlines the research and findings of this thesis.

Chapter 2 provides a review of theories and practices relevant to argumentation, AVTs, and learning. Definitions and the educational significance of each theory are discussed in addition to their connections with argumentation and AVTs. Learning outcomes, as well as conditions for successful use of AVTs, are also evaluated.

Chapter 3 includes the design and development of a new AVT, the Dialectical Map. It describes the design goals and features as well as addresses the design challenges many AVTs face. The chapter also describes three full course-length classroom deployments of the DM.

Chapter 4 reports an experiment that compared the learning outcomes of the DM, the verbal argumentation training only, and the text-only study conditions. Chapter 5 presents the results of this experiment.

Chapter 6 contains a general discussion of the results from the laboratory experiment and feedback from the classroom implementations in the context of the theories presented in Chapter 2. The chapter concludes the thesis with theoretical contributions, practical implications, and limitations of the research conducted in this thesis, as well as how future research may address the limitations and build onto the current research. In sum, this thesis offers new knowledge about argumentation, argument visualization tools, and learning that has direct implications for the design and application of effective AVTs and cognitive tools.
Chapter 2.

Literature Review

It is beyond the scope of this thesis to review the massive literature on argumentation and its origin with ancient Greek philosophers. Rather, the goal of this chapter is to establish a theoretical infrastructure for argumentation visualization and to investigate the effective use of argument visualization tools (AVTs) for teaching and learning. This chapter comprises five sections: a) what is argumentation; b) the relation between argumentation and learning; c) theories pertinent to argument visualization and AVT, including cognitive load theory, schema theory, cognitive tools theory, and multimedia learning theory, as well as d) a critical examination of empirical findings of AVT applications in education.

2.1. What is Argumentation?

Argumentation is defined as a written or dialogical process that involves reasoning for or against a standpoint that is controversial either between the arguer and the interlocutor or within internal discourses of the arguer. During this process a series of propositions are put forth to justify or refute the standpoint in front of a hypothetical or actual “rational judge” (van Eemeren et al., 1996, p. 5). Argumentation can be an individual activity, such as when an arguer weighs the pros and cons of an issue in order to make a decision; or, when reasoning is exchanged between arguers, it can become a group activity.

During the process of argumentation, various kinds of arguments are produced. There are numerous ways to categorize argumentation and arguments. Researchers are yet to reach unanimity on their classifications. For example, Walton (1989) used situations, methods, and goals to classify argumentation into 29 argumentation schemes
(Walton, Reed, & Macagno, 2008); Andriessen, Baker, and Suthers (2003) used five characterizing factors: “object, reasoning, medium, activity, and goal” (p. 6). The object of argumentation refers to what the argument is about. It determines the types of reasoning involved in the argumentation process (such as analogical reasoning and plausible reasoning. Argumentation can be processed in various media (such as speech, writing, and Internet) and representations (such as words, diagrams, and symbols). Media can affect the clarity and effectiveness, which is termed “sharpness” (p. 7), of argumentation, due to the cognitive properties of multiple representations. The activity involved in argumentation refers to the social situation where the argumentation is generated, such as interpersonal and intrapersonal activities. The goal of argumentation can be solving a common problem or convincing others to accept one’s position, or refute other’s positions. Kuhn (1991, 1992) categorized arguments into rhetorical arguments and dialogical arguments based on whether or not the goal is to demonstrate truthfulness of a statement, or a dialog between two agents with opposing standpoints that occurs socially or internally. According to Kuhn, a rhetorical argument refers to “a course of reasoning aimed at demonstrating the truth or falsehood of something” (p. 12), and a dialogic argument refers to “a dialog between two people who hold opposing views. Each offers justification for his or her own view; in addition... each rebuts the other’s view by means of counterargument” (p. 12).

Research on argumentation is still evolving. Occasional confusion and ambiguity may arise due to variations in theoretical foci and inconsistencies in the use of terminologies by different theorists and approaches. Nevertheless, three of the most influential argumentation theories inspired my work in the educational applications of AVTs and are useful in portraying the research described in this thesis. They are: the Toulmin Model, Walton’s argumentation schemes, and Dialectical argumentation. As most argumentation theories share some common roots from ancient philosophy, the theories incorporated in this thesis should not be viewed as mutually exclusive.

2.1.1. The Toulmin Model

Toulmin’s theories have profound impact on the development of the argumentation field. He rejected the view that formal logic is the sole criterion for judging
the cogency of an argument and shifted the focus of argumentation to informal reasoning that values justification of assertions (Toulmin, 1958). According to Toulmin, formal logic is neither sufficient nor necessary in determining the validity of an argument in practice (van Eemeren et al., 1996). The audience may accept a claim if the argumentation is properly justified. More importantly, the audience can also challenge a claim and demand supporting evidence and justifications (van Eemeren et al., 1996).

The Toulmin model has been frequently used in the construction, analysis and evaluation of arguments (van Eemeren et al., 1996). It consists of six parts: claim, data, warrant, backing, modal qualifier, and rebuttal. The argumentation process involves supporting a claim with data (i.e., evidence or grounds), where the relevance of the data to the claim may require justification by a warrant. The warrant may also have backings as its validation. A modal qualifier may be used to condition the claim. And the claim may be subject to challenges and consequently may require refutation of these challenges that are called rebuttals. See Figure 2.1 for an illustration of the Toulmin model. Toulmin recognizes that the basic structure of argumentation is field independent, but components such as warrant and backing are field variant. For example, the acceptability of a warrant in law is dependent on legal rules and a backing in science is dependent on scientific principles.

Despite the significant contributions the Toulmin model has made to the field of argumentation, it is habitually criticized by argumentation researchers. For example, in an argument the warrant is often implicit and difficult to identify. Moreover, arguments without warrants can also be sound (Anderson, Chinn, Chang, Waggoner, & Yi, 1997). Additionally, researchers like Chin and Anderson (1997) believe the Toulmin model is only suitable to analyze simple arguments made by an individual but not appropriate for complex argument networks. Some also criticized that the Toulmin framework fails to address both sides of an issue during the argumentation process (e.g., in Leitão, 2001). This is not necessarily true since “rebuttal” acknowledges counterarguments and calls for their refutations. The components of an argument from the Toulmin model may be transformed to construct counterarguments. That is, the counterargument may also include a claim, its supporting evidence, the necessary warrants, backing, qualifier, and further rebuttal. In this situation, the further rebuttal would be refuting claims from the
opposite (the original argument) side of the issue, or a third voice of an argument.

However, on the methodological level, the Toulmin model does appear to neglect its potential dialogic and dialectic uses. That is, in a practical sense, how the model should map onto real arguments can be somewhat obscure. For example, the potential opponent’s view (the counterarguments) is assigned a secondary role in the model compared to the main claim by the proponent. The Toulmin model also seems to ignore the manifestation of the acceptance or rejection of claims as an outcome of the argumentation, which renders the model problematic for practical use (Leitão, 2001). Therefore, for the model to be useful in practice, explicit add-on explanations may be needed when introducing the structure. Such add-ons could be critical questions integrated in Walton’s argumentation themes (see section 2.1.2).

2.1.2. Walton’s Argumentation Schemes

Another useful of argumentation theory for educational purposes is Walton’s logical argumentation (Walton, 2013), which emphasizes the integration of informal logic and artificial intelligence. Logical argumentation offers inter-disciplinary guidance on argument invention, identification, analysis, and evaluation (p. 3). Walton claims that argumentation is a procedure that contains a starting point, a middle stage with dialogues between agents who take turns to make assertions and ask questions, and an end point. The goal of logical argumentation is not to prove something truth or false, but to make decisions based on arguments both pro and contra the claim. Logical argumentation also views knowledge and intelligence as social instead of only individually located, therefore arguers are committed to their viewpoints instead of having their position as only personal beliefs. A commitment is a statement transparent to others as opposed to an internal belief. This statement is captured by textual, audio, or video means that can be used to identify what an arguer has said, what position he/she has taken during his/her interactions with others. In other words, an arguer publicly declares that he/she is arguing for or against a proposition regardless of his own beliefs on the topic. A proposition is accepted as knowledge after evidence of both sides has been examined, and the justification of proof shows that the supporting evidence is substantially stronger than the counterevidence. Logical argumentation uses a dynamic
method, which means even after the acceptance of an argument based on the procedure of justifications, new proof may be obtained later that refutes the argument. In logical argumentation, knowledge is not considered absolute truth. Conclusions and decisions are generally made “under conditions of uncertainty and lack of knowledge” (p.3). Consequently, arguments are defeasible and subject to critical questioning until the ending point of the argumentative discourse.

Logical argumentation is manifested through a set of argumentation schemes, stereotypical descriptions of reasoning that characterize acceptable arguments in everyday conversation and specific contexts such as politics, law, science, and artificial intelligence (Walton, 2013; Walton et al., 2008; Walton & Reed, 2002). Walton (1996) identified 29 commonly used argumentation schemes with corresponding critical questions that respondents of an argument should ask to challenge the proponent of an argument. Critical questions are essential in evaluating the cogency of an argument. The argumentation schemes offer specific models that may be useful for teaching and instruction. For example, they could help students identify, analyze, and evaluate arguments in day-to-day life, such as processing information from textbooks, newspapers, and cyber space. They could be used to facilitate dialogues in the form of classroom and online discussions. Argumentation schemes could also help teaching critical thinking skills (Walton et al., 2008). Depending on how subtypes of argumentation schemes are classified, Walton and colleagues described a list of 96 argumentation schemes in their later work (Walton et al., 2008). Here is a common example of an argumentation scheme: *argument from expert opinion*. It includes two premises and a conclusion (Walton, 2013, p. 6):

**Premise I:** A is an expert in the subject domain under discussion.

**Premise II:** A states that proposition B is true (false).

**Conclusion:** B is true (false).

Critical questions would include:

1) Expertise question: How credible is A as an expert?
2) Field question: Is A an expert in the field that B is in?
3) Opinion question: What did A assert that implies B?
4) Trustworthiness question: Is A personally reliable as a source?
5) Consistency question: Is B consistent with what other expert assert?
6) Backup evidence question: Is A’s assertion based on evidence?

These schemes relate to one another in a complex way. Walton, Reed and Macagno (2008) classified them into three main categories: 1) reasoning, which includes deductive reasoning, inductive reasoning, practical reasoning, abductive reasoning, and causal reasoning; 2) source-based arguments, which includes arguments from position to know, arguments from commitment, arguments attacking personal credibility, and argument from popular acceptance; and 3) applying rules to cases, which includes arguments based on cases, defeasible rule-based arguments, verbal classification arguments and chained arguments connecting rules and cases. In order to understand these categories and the argument schemes under them, students would need prerequisite knowledge on informal reasoning. Thus, the classification is useful in directing students’ knowledge inquiry on informal logic. These schemes have been used to analyze arguments (Anthony & Kim, 2015) understanding and reconstructing arguments with missing components.

2.1.3. Dialectical Argumentation

Dialectical reasoning can be traced back to Socrates and the Socratic questioning technique that is well known to educators. The most noted dialecticians are Hegel and Marx, who deemed dialectic the driving force of human history. They considered contradictions fundamental in both ontological and epistemological senses. The goal for dialectical reasoning is integration of contradiction, commonly called the thesis-antithesis-synthesis triad (Ravenscroft, Wegerif, & Hartley, 2007). According to Hegel and Marx, the synthesis is an outcome that is not a combination of thesis and antithesis, but is considered different from either the thesis or antithesis and can be used as a thesis for the next triad (Ravenscroft et al., 2007).

Dialectical reasoning is the most sophisticated cognitive activity and is considered the final stage of human cognition (Riegel, 1973). The Piagetian cognitive development model has often been criticized for not accounting for more advanced and creative adult cognitive activities beyond the formal operational stage. For instance,
scientific activities carried out through creative strategies that work with contradictions. Piaget’s adaptation process that enables advancement from one stage to the next may be viewed as a dialectical foundation expressed through equilibrium, assimilation, and accommodation. However, Piaget interprets cognitive development as an on-going separation from this dialectical foundation in the “non-contradictory thinking” of formal operations (Riegel, 1973, p. 1). Riegel updated the Piagetian model by bringing back this foundation and adding dialectical reasoning as the final stage of cognitive development.

As argumentation is generally considered an important tool for reasoning (Andriessen et al., 2003; Coirier et al., 1999; Nussbaum, 2008; Nussbaum, 2002; Voss & Means, 1991; Voss, Wiley, & Sandak, 1999), dialectical argumentation can be seen as an activity or means of dialectical reasoning. It involves evaluating and synthesizing different views. Jonassen and Kim (2010) believed that for educational purposes dialectic arguments are more useful than rhetorical arguments. During dialectical argumentation, arguers are exposed to various ideas on the topic either supporting or opposing their own arguments. Such exposure is thought to benefit learning. For instance, in order to refute counterarguments, the arguer needs to understand the counterarguments as well as strengthen the understanding of their own arguments. Through organizing claims, reasons, and counterarguments, the arguer can better organize their knowledge, which leads to better understanding and recall of the learning materials (Means & Voss, 1996).

Social interactions are closely tied to dialectical reasoning. Therefore, dialectical argumentation would fall in Kuhn’s dialogical category (1991), as discussed earlier in section 2.1. Dialectic and dialogic are distinct yet closely intertwined processes in their relation to learning. According to Ravenscroft, Wegerif, and Hartley (2007), dialectic emphasizes more of the inner processes of reasoning although it may use dialogues as a means. Dialogic emphasizes the interactions among participants and the ability and openness in taking others’ perspectives. Dialectical argumentation is realized through a discourse between holders of a thesis and antithesis. In practice, dialectical argumentation can be a solitary (e.g., weighing pros and cons to make a decision) or group (e.g., a discussion within a social setting) activity (Driver et al., 2000). The goal of dialectical argumentation is to resolve contradictions and inconsistencies (Barth &
Krabbe, 1982; van Eemeren & Grootendorst, 1992). The resolution may take different forms, for example when arguments are contentious, the resolution may involve convincing the opponents to accept one’s claim. However, dialectical argumentation is not always adversarial, that is, arguers may also seek a compromise between their claims.

2.2. Argumentation and Learning

The ability to identify, construct and evaluate arguments underlies a very high proportion of academic subjects and intellectual endeavours. Argumentation generally involves stating a claim and providing evidence and reasons to support the claim. For skilled arguers, counterarguments are anticipated, identified, and refuted (Nesbit et al., 2008). The relationship between argumentation and learning is commonly considered to operate in two directions: a) argumentation has instructional value in various subject areas and multiple aspects of student learning, which is often abridged as arguing to learn in the literature, and b) argumentation is an important curricular goal in its own right, often abridged as learning to argue, which involves the acquisition of argumentation skills (e.g., using reasons and evidence to support claims, raising counterarguments, and refuting counterarguments, etc.; Andriessen et al., 2003; Nesbit et al., 2008). Although in theoretical terms we can discuss these two causal links as distinct, in practice they are inseparable.

2.2.1. Effects of Arguing to Learn

Argumentation has been shown to improve learning in many subjects, such as science (Nussbaum & Sinatra, 2003), mathematics (Schwarz et al., 2000), and social studies (Wiley & Voss, 1999). Argumentation, as an instructional strategy, has great potential in promoting learning for several types of intended outcomes such as recall of information, understanding subject matter, critical thinking, and writing skills. Argumentation can also boost students’ motivation to learn (Chinn et al., 2001). Increasingly, educators consider argumentation as an effective instructional strategy that should be more broadly adopted within schools (Nesbit et al., 2008).
Recall of Information

Researchers have not reached agreement on the effects argumentation has on student recall. Some found that argumentation interventions enhanced recall of information, some found no effects; some even found adverse effects that argumentation had on the recall of text-based information. It has been postulated that when presented with a study material, such as a text, students may have different experiences interacting with the text depending on their schema (see a later section for more details on schema). For example, when they recognized the text as argumentative, their interaction with the text would follow the argument schema theory. Thus, information of the text will be organized in a way that fits into the slots of such a schema, such as claims, supporting evidence, and counterarguments (Reznitskaya et al., 2007). As a consequence, students are expected to encode and recall the text more proficiently. Reznitskaya et al., (2007) investigated the effects of argumentation instruction with elementary school students. They identified enhanced recall instances in students’ written and oral responses, which included most or all elements of an argument schema.

On the other hand, some researchers argue that even though argumentation may better engage students in certain perspectives of learning, higher level of engagement does not necessarily produce better recall (Kintsch & Young, 1984; Mannes & Kintsch, 1987; McDaniel & Donnelly, 1996; McNamara, Kintsch, Songer, & Kintsch, 1996). Since recall does not require extensive inferencing of the study materials, it is reasonable to assume that argumentation interventions may not benefit students’ recall performances. For example, Mao (2005) investigated the effects of three different types of instructions (recall, summarize, and argue) on university students’ and found no difference on free recall across conditions.

Some researchers believe argumentation directs student attention and engagement away from superficial text memorization and focus more on deeper processing and critical thinking. As a consequence, argumentation may hinder recall. For example, Lin, Horng, and Anderson (2014) used texts organized with argumentation schema with five components: claim, explanation, supporting evidence, counterargument, and rebuttal. Subjects were assigned to two conditions: the argument
scaffolding group that was instructed to consider both sides of the argument and components of the argumentation schema, and the control condition without such instructions. The authors found that the argument scaffolding condition recalled less text-based information compared to the control group. They argued that argumentation scaffolding encouraged students’ focus on comprehension and critical thinking instead of trivial details of the text.

Understanding

During argumentative activities, students need to elaboratively process information in order to connect evidence for or against the concepts and principles they are learning (Chinn, 2006). Such deep processing is theoretically beneficial to the understanding and mastery of the study content (Pressley & Woloshyn, 1995). Evidence has supported such conjectures. For example, in an experiment by Zohar and Nemet (2002), the authors found that ninth graders who received argumentation instructions outperformed those who learned the same materials without the argumentation interventions in obtaining knowledge on human genetics in a 20-item multiple-choice test. Wiley and Voss (1999) conducted two experiments to investigate undergraduates’ understanding of historical subject matter. Results showed that students who wrote arguments demonstrated better understanding of the topic than those who wrote non-argumentative essays (narrative, summary, or explanation). Students’ performance on an inference verification task and a principle identification of analogy task was used as evidence for such understanding. In both experiments, students who wrote arguments outperformed the other conditions on inference verification. For the principle identification task, the argumentation group outperformed the narrative condition in the second experiment.

Critical Thinking

Critical thinking is considered central to education (Angeli & Valanides, 2009). It is a complex construct that has been defined in numerous ways. For the purpose of this thesis, critical thinking refers to “reflective and reasonable thinking that is focused on deciding what to believe or do” (Ennis, 1985, p. 45). Critical thinking and argumentation are closely related to each other. As a decision making process, critical thinking involves
applying argumentation skills such as analyzing, evaluating, and weighing the strengths and weaknesses of different viewpoints (Paul, 1990). Argumentation is at the core of critical thinking skills (Beyer, 1995; Paul, 1990) and a means to teach and develop critical thinking skills (Beyer, 1995; Kuhn, 1992; Means & Voss, 1996; Paul, 1990; Sanders et al., 1994). Argumentation is also frequently used to assess critical thinking skills. Various critical thinking assessment instruments consistently measure argumentation skills in their tests. For instance, the California Critical Thinking Skills Test (CCTST), a widely used critical thinking skills measurement tool, requires test takers to analyze information, draw warranted inferences, and evaluate the reasoning provided in various scenarios.

Malamitsa, Kasoutas, and Kokkotas (2009, p. 459) examined how argumentation promoted 6th grader’s critical thinking skills. The argumentation activities were aimed at

1) Stimulating students’ critical reflection on the knowledge and the experience gained inside and outside the classroom;
2) Promoting students’ awareness of subjective and ideological biases; and
3) Developing students’ ability to analyze evidence expressed in rational argumentation.

The Test of Everyday Reasoning (TER) was used to assess students’ critical thinking skills. Students who were engaged in argumentation activities such as dilemmas, debates, and controversies in science outperformed the control condition in their critical thinking.

Sanders et al. (1994) investigated the effects an argumentation course on university students’ critical thinking skills compared to students who did not take the course. The authors found that students who were trained on argumentation showed stronger critical thinking skills through the ability to detect weakness in arguments than those who did not receive argumentation training.

**Writing**

Page-Voth and Graham (1999) investigated the effects training and argumentation goal instructions on writing skills with 7th and 8th grade students with
learning disabilities. Subjects were randomly assigned to three conditions: one was instructed to set goals on supporting claims with reasons and refute counterarguments; another group was instructed to set the same goals and use a strategy that helps sustain the goals; and the control condition with no goal instructions. Compared to the control condition, those who received instruction on goal setting and strategy provided more supporting reasons in their essays, and included more refutation of counterarguments. Students who received goal instructions alone also generated higher quality essays than the control group based on the overall impression on a holistic scale developed by the researchers. Students in the goal instruction conditions also wrote longer essays than the control condition by 31 more words on average per essay. However, the authors did not further investigate whether or not the effects resulted from goal instruction alone or argumentation training or the combined method. They were not clear on whether the control participants were trained on argumentation. From Chinn’s (2006) interpretation, it seems that the control condition did not receive argumentation training. However, according to Ferretti, MacArthur, and Dowdy (2000) the control condition seemed to have received the same training on argumentation concepts.

Ferretti et al. (2000) compared a general goal and an elaborated goal with argumentation elements (e.g., propositions, reasons, counterarguments, and rebuttals) on writing outcomes with 4th and 6th graders. They found that 6th graders in the elaborated goal condition generated more persuasive essays and used more argumentation components than other 6th graders in the general goal condition. Sixth graders from the argumentation goal condition also outperformed all 4th graders in the same measures.

Song and Ferretti (2013) investigated the effects teaching students argumentation schemes and critical questions had on the revision of their own persuasive essays (see a previous section on Walton’s argumentation schemes). Overall essay quality was assessed based on its persuasiveness on a scale the researchers developed, including “a clearly stated opinion,” “supporting reasons,” “elaboration of reasons,” “anticipation and response to counterarguments,” and “organizing arguments with introduction and conclusion” (p. 75). They found that students who were instructed to apply both argumentation schemes and critical
questions in revision generated higher quality essays than those who were instructed to use argumentation schemes as well as the control condition that received neither instructions. This effect held true when the quality of the first draft was controlled for.

**Motivation**

During argumentation activities, students are confronted with inconsistencies and controversies. Contradicting ideas are believed to be naturally more stimulating therefore may enhance students' intrinsic motivation. Conflicts raise uncertainties in students' minds, which motivate them to seek new information in order to resolve the uncertainties (Johnson & Johnson, 1979). In the process of convincing others, students are encouraged to not only articulate their own arguments but also to understand alternative arguments in order to resolve the contradictions (Kruger, 1993; Nussbaum & Sinatra, 2003; Wiley & Voss, 1999). Such exercises are believed to induce a high level of engagement. Students also have more freedom expressing their arguments compared to traditional recitations (Chinn et al., 2001). Such freedom may also increase motivation. Chinn et al. (2001) examined 4th graders and found that argumentative discussions were more engaging compared to the condition with traditional recitations. Students spoke more words per minute in the reasoning condition and were more eager to compete for turns to speak. These results indicated higher engagement in the discussions than recitations.

Smith, Johnson, and Johnson (1981) studied 6th grade students for the effects controversy had on motivation. Subjects were assigned to three conditions: the controversy condition with group debates on pros and cons, the concurrence-seeking condition that studied cooperatively, and the individualistic condition that studied the material with no peer interactions. The researchers found that the controversy group showed higher motivation to continue learning the subject than subjects in the other two conditions.

The learning outcomes brought forth by argumentation discussed in this section are not meant to be exhaustive. Argumentation enhances learning on many more aspects, such as problem solving (Arkes, 1991; Wharton, Cheng, & Wickens, 1993) and conceptual change (e.g., Asterhan & Schwarz, 2007; Nussbaum & Sinatra, 2003;
Salisbury-Glennon & Stevens, 1999). Studies included in this section help endorse the essential role argumentation has in learning and instruction. The next section discusses strategies educators could adopt to help students learn argumentation skills in order to better reap the educational benefits.

2.2.2. Teaching and Learning to Argue

This section discusses three main factors that are key to teaching and learning argumentation skills: a) why argumentation should be a curricular goal in its own right, with reasons in addition to its educational benefits discussed in section 2.2.1; b) what constitutes good argumentation skills, and c) how to teach students to become better arguers, with examples on strategies that are effective in achieving such goals.

Why

In Kuhn’s widely cited work (1991, 1992), argument is positioned at the heart of higher-order thinking in the real world. “Thinking as argument is implicated in all of the beliefs people hold, the judgments they make, and the conclusions they come to; it arises every time a significant decision must be made” (1992, p. 156-157). Examples from section 2.2.1 represent the growing confidence that argumentation has great potential in enhancing student learning. However, students often demonstrate poor argumentation skills. For instance, they often fail to differentiate and coordinate evidence and explanation (Kuhn, 1991); when constructing arguments, students tend to argue in a biased way by searching for evidence that confirms their prior beliefs and ignore counterevidence (Chinn & Brewer, 1998); even philosophy students tend to overlook the logic structure and read arguments as stories (Harrell, 2005). Researchers have recognized that argumentation skills are not naturally acquired. Similar to other skills, argumentation proficiencies improve with practice and feedback. Therefore, teaching argumentation skills is attainable and necessary (Kuhn et al., 2008; Kuhn & Udell, 2007). Students need to learn argumentation skills in school so that they can become better thinkers, problems solvers, and decision makers in society. Many scholars advocate that argumentation should be central in education (e.g., Chinn, 2006; Kuhn, 1991).
What

In order to effectively teach and learn argumentation skills, it is important to understand what constitutes good argumentation skills. Various argumentation theories sustain diverse criteria. They examine argumentation skills from different angles; therefore, one is not necessarily superior to another. For example, Blair and Johnson (1987) examined argumentation from three perspectives: a) whether or not the premises are acceptable, b) whether or not the premises are relevant to the conclusion, and c) whether or not the premises provide sufficient support to the conclusion. On the other hand, according to Walton (1989), effective argumentation involves strengthening one’s own argument and weakening the opponent’s argument. Kuhn also identified specific argumentation skills that are vital to form strong arguments: “a) the skill to generate causal theories, b) the skill to offer evidence to support theories, c) the skill to generate alternative theories, d) the skill to envision conditions that would weaken the alternative theories, and e) the skill to rebut alternative theories” (Jonassen & Kim, 2010, p. 441). Voss and Means (1991) suggested criteria that appeared to incorporate both the Toulmin model and Walton’s argumentation schemes. They indicated that good argumentation contains claims, supportive reasons to the claim, backing, counterarguments, and refutations. Qualifiers that define the condition of an argument are also considered an important element for effective arguments. A good arguer should also be able to use various argument schemes such as “argument from consequence” and “arguments from position to know” (as in Walton, 2013, p. 110). The above criteria may be applied to both individual and group arguments. Chinn (2006, p. 360) provided specific criteria for good argumentation skills within group arguments:

1) Constructing good individual arguments;

2) Collective efforts in integrating multiple arguments into a coherent group product with different viewpoints to examine; and

3) Weighing all evidence on pro and con arguments and drawing a final conclusion.

Regardless of the variation in their presentation and layout, there are basic commonalities among such criteria, manifested in their emphases of at least one of the following factors: 1) both the rhetoric and logic of argumentation, and 2) both individual and collective involvement and significance in argumentation.
How

Jonassen and Kim (2010, p. 442) identified four main barriers students met when learning argumentation skills in school:

1) Teachers lack necessary skills to teach argumentation;
2) Curricular pressure to cover materials other than skill development;
3) Students lack prior knowledge to practice argumentation; and
4) Instructional designers lack knowledge on argumentation.

Researchers have been making on-going endeavours to remove these barriers when teaching students how to argue. It is likely that general approaches such as professional development for teachers focusing on argumentation, curriculum reform making argumentation central in most subjects (if not all), teaching students what good argumentation skills are with abundant exercises, and developing proper cognitive tools to scaffold argumentation may be able to help overcome these obstacles. Consequently, argumentation researchers have been exploring numerous specific instructional strategies to help students become sophisticated arguers. I discuss a few popular techniques in this section with their empirical findings that support these techniques: 1) group discourse; 2) goal instructions; 3) direct instructions on the structure and vocabulary of argumentation; and 4) using cognitive tools as scaffolding.

Group Discourses

Researchers have investigated the effects of engaging students in dialogical settings on their argumentation skills. This method is based on the Vygotskian belief that cognitive skills are developed through social exchanges (Bruning, Schraw, Norby, & Ronning, 2004; Salkind, 2004). During social interactions, internal ideas are externalized and articulated. The diversity of ideas improves individual thoughts because students have the opportunity to compare their own ideas with others' as well as to challenge their peers' reasoning (Kuhn, 1992). Kuhn and Udell (2003) indicated that merely engaging students in extensive argumentative discourse would improve their argumentation.

Reznitskaya et al. (2001) examined the effects of argumentative discussions on argumentation skills reflected in essay writing. In particular they applied a discussion
approach, “collaborative reasoning” (p. 157), which involved small and open group discussions on controversial issues. Students were given the freedom to decide the content and pace of their discussions. The goal of collaborative reasoning was for the students to experience thoughtful deliberation rather than reaching consensus. The authors studied how collaborative reasoning affected students written arguments with subjects from public schools. Students were assigned to the collaborative discussion condition or a control condition for a 5-week investigation. The discussion group met twice a week to discuss controversial topics for generally 15 to 20 minutes. They also held 15-minute on-line discussions twice every week. Students were coached to argue for their own claims and challenge others’ ideas with counterarguments, as well as to counter others’ challenges with rebuttals. The authors found that subjects from the collaborative reasoning condition included in their essays more arguments, counterarguments, and rebuttals. The treatment condition also used more formal argument devices, such as “in the story it said (evidence),” “other people might say (counterargument),” and “some people may disagree because (counterargument)” (p. 161).

Kuhn, Shaw, and Felton (1997) investigated the effects of social discourse through a series of dyadic interactions on the quality of students' written argumentation. Subjects included adolescents and young adults. Each age group was assigned to an experimental condition with the dyadic discussions and a control conditions with no social discourses. During a 5-week period, subjects in the experimental condition had five dyadic discussions with different peers. Each discussion lasted 10 to 15 minutes. Participants were encouraged to reach consensus on the issue, but if disagreement persists after the discussions, they were asked to characterize the discrepancy. The researchers found enhanced argumentation with the discussion condition compared to the control condition. For example, the range of arguments increased, indicating new knowledge acquired from the social discourse. Qualitative measures of reasoning were also identified from both age groups, such as the change from one-sided to two-sided arguments, arguments that contained alternative perspectives, and metacognitive awareness of different views.
Goal Instructions

In a previous section, I discussed the effect of argumentation goal instructions on students’ writing skills (e.g., students generated more persuasive, longer, and better overall quality essays). Argumentation goal instructions were also found to increase students’ use of argumentation components in their writing, indicating improved argumentation skills (Ferretti et al., 2000). Argumentation goals can provide specific guidance by directing student attention to particular aspects of argumentation, such as generating rebuttals. Thus, argumentation goal setting may be an effective instructional strategy to teach students to become better arguers. Nussbaum and Kardash (2005) investigated how specific goals affect college students’ argumentation skills. They reported that students who were given goal instructions to generate counterarguments and rebuttals in writing argumentative essays produced more counterarguments and rebuttals than students in the control condition who were instructed only to express their opinions. The authors also advised caution when using persuasive goals because they tended to diminish the generation of counterarguments and rebuttals. Nussbaum (2005) investigated how goal instructions affected student performance in interactive argumentation. All participants were taking a course designed for pre-service teachers. The author discovered that unlike individual argumentation, interactive argumentation with persuasion goals resulted in greater numbers of claims and more sophisticated arguments. Specific goals to generate counterarguments also resulted in more opposition and debates but the effects were smaller than those from persuasion goal instructions.

Klein and Ehrhardt (2015) investigated how sub-goal distributions affected students’ reasoning performances. They found that sub-goal segmentation was responsible for less my-side bias and more balanced inferences compared to giving students clustered sub-goals, possibly through the reduction of cognitive load. The authors suggested using segmented sub-goals to introduce difficult writing tasks. Caution was also recommended, because sub-goal segmentation may encourage incoherent claims for low achieving writers.
Explicit Instructions

Voss and Means (1991) proposed that in order to effectively teach students argumentation, students would need to learn about the structure and vocabulary of argumentation; understand the purposes of argumentation; and learn what constitutes a good argument. Explicit instructions of these aspects may help students become better arguers. Researchers have been exploring effective ways to help students better understand the notion and process of argumentation in order to apply the skills in learning tasks. A popular approach is using the Toulmin model to teach reasoning skills and argumentation (e.g., Cerbin, 1988; Leeman, 1987; Saunders, 1994). Students are first presented with a clear specification of what skills they need to learn. Such skills are organized in a sequence based on their complexity and difficulty. Direct instructions then clearly define, substantially explain, and demonstrate each element of the model so that students can identify them in a series of exercises. Extensive feedback is also given to guide and shape student learning (Cerbin, 1998). Dawson and Venville (2010) conducted a case study with grade 10 students. The teacher participated in an extensive professional development program called Ideas, Evidence and Argument in Science (IDEAS, see Osborne, Erduran, & Simon, 2004 for more details). The IDEAS program included the Toulmin model to introduce argumentation to the teacher. The teacher then delivered his instruction based on what he learned from IDEAS in two 50-minute lessons. For example, listing components of an argument and explaining their roles in argumentation to students, encouraging counterarguments, and prompting the generation of backings and qualifiers, etc. The authors found that students who were taught argumentation concepts reported improved argumentation skills compared to students who did not receive argumentation instructions.

Models other than the Toulmin structure have also been developed to teach students how to argue. For example, van Lacum, Ossevoort, and Goedhart (2014) developed a scientific argumentation framework that specifies the skills students need to learn as well as how to identify the rhetorical moves when reading research articles. Seven components were specified in this model: “motive,” “objective,” “main conclusion,” “implication,” “support,” “counterargument,” and “refutation” (p. 255). The instruction was realized through 2-hour to 2.5-hour group tutorial meetings that contained tutor
instruction, assignment, feedback, and group discussions. The authors found that students were better able to identify the rhetorical moves they specified after the argumentation instructions. Students also perceived improved skills in reading research articles. No control condition was included in this research.

Sanders et al. (1994) investigated how teaching argumentation turned university students into better arguers. Instruction of argumentation was realized through a general education course on argumentation. The researchers did not give details on how argumentation was taught in this course. Students reported more favourable self-assessment on their argumentation skills after the instruction, compared to the control group who did not take the argumentation course. Students who received argumentation instruction also outperformed the control condition in detecting weak arguments.

Using Cognitive Tools

Cognitive tools are generally considered as technologies that help learners engage in, examine, and accomplish cognitive tasks by sustaining or extending learners’ cognitive capacity (Jonassen & Reeves, 1996; Kim & Reeves, 2007; Lajoie, 1993). They are designed based on learning theories that sustain, enhance, or transform learning for particular educational circumstances (Lajoie & Azevedo, 2006; Lajoie, 2007; Lu & Lajoie, 2008). Researchers have been investigating cognitive tools and their role in facilitating students’ argumentation skills (e.g., Cho & Jonassen, 2002; Lu & Lajoie, 2008). Cognitive tools used for argumentation may take many forms, such as collaborative argumentation tools (Lu & Lajoie, 2008), sentence starters in a discussion forum (Tsai & Tsai, 2014), or constrain-based argumentation scaffolding (Cho & Jonassen, 2002; Oh & Jonassen, 2007). These tools sprout from the close connections among technology, argumentation, and the learners’ cognitive processes at both a collective level and an individual level. At the collective level, there are cognitive tools that may facilitate students’ argumentation through documenting, sharing, and critiquing each other’s arguments in social discourses. The collaborative tools enable a shared mental model that helps students understand problems, define task goals, and make decisions (Lu & Lajoie, 2008). At the individual level, cognitive tools may sustain argumentation skills through schema construction and activation. According to Vygotsky, cognitive tools are
internalized social experiences in the form of abstract mental symbols that form schemas (Nesbit et al., in press). Argumentation is often conceived as both a social and an individual process. Therefore, accounting for both collaborative and individual learning is an attempt to explain more in-depth argumentation skills gained through using cognitive tools (see more details on schema theory in a later section).

Lu and Lajoie (2008) used an interactive whiteboard to support medical decisions with argumentation. Subjects (third year medical students) were divided into two groups. The control group used a traditional whiteboard. Collaborative argumentation tools were built in the interactive whiteboard where students participated, recorded, shared comments and annotated their arguments for decision-making. The authors discovered that the interactive whiteboard group using argumentation tools generated more productive arguments through sharing data and understanding about the patient during collaborative discourses.

Teo and Churchill (2007) investigated how sentence starters affect students’ argumentation in online discussions. Students from a teacher training program used an online forum named Knowledge Community to evaluate and debate about the strengths and weaknesses of educational videos. The authors integrated six types of sentence starters to support student argumentation, such as “taking a stand,” which included two sentence openers “I agree that…” and “I do not agree that…” and the “probe reasons” category, which included “My opinion is...” “An unusual idea is...” “The indicators/facts supporting my opinion are...” and “An improvement to the suggestion...” Student interviews and discussion content analyses showed that sentence starters effectively supported student argumentation in online discourses. Sentence openers were widely used by students especially “taking a stand” and “probe reasons.” A large proportion of online posts were on-task arguments as opposed to off-task, non-arguments. Furthermore, students’ statements were mostly in-depth arguments, which were defined as messages that “reflect organization and critical evaluation of information” (p. 214).

Currently a special type of cognitive tools, argument visualization tools (AVTs), are gaining scholarly attention for the purpose of scaffolding students’ construction,
understanding, and evaluation of arguments. The next sections discuss the definition, theoretical foundations, and empirical findings of using AVTs in educational settings.

2.3. Argument Visualization Tools (AVTs)

In recent years, researchers have investigated instructional uses of argument maps (e.g., Dwyer et al., 2010) and argument visualization software (e.g., Davies, 2009). Currently, more than sixty computer supported argument visualization tools have been developed (Scheuer, Loll, Pinkwart, & McLaren, 2010). We need to ask whether these scaffolds bring educational benefits, and if so, by what means? This section offers theoretical and empirical evidence on pedagogical efficacies of AVTs.

The first argument visualization tool (AVT) is generally believed to be the legal sketching described by John Henry Wigmore in 1913 (Buckingham Shum, 2003). AVTs are graphical tools used to model the content and the inferential structure of arguments as an alternative to traditional prose-only representations. Such visualizations contain both pictorial and textual elements in order to visualize components of arguments and the links among them (Andriessen et al., 2003; Twardy, 2004; van Gelder, 2003; Walton, 2013). These components may include: 1) claims, grounds, warrants, backings, rebuttals, and modal qualifiers based on the Toulmin model; 2) elements of a specific argument scheme according to Walton (2013); or 3) elements of arguments defined by other researchers (such as in Nussbaum, 2008). An argument map is a typical AVT, composed of boxes that contain statements or propositions, and arrows that indicate the relations between the propositions. The primary focus of this thesis is computer-supported AVTs.

2.3.1. Why AVT?

As discussed previously, a rich body of research has documented that argumentation promotes learning, enhances the motivation to learn, and supports higher order thinking skills (Nesbit et al., 2008). However, students are often poor arguers. Researchers discovered that when given scaffolds for an argumentation exercise, such as argument mapping, students no longer have difficulty coordinating theories and
evidence, and demonstrated deeper engagement with the evidence items. They stop giving superficial descriptions and use more causal warrants (Bell, 2001). Students also recognize arguments and their logic structures better (Harrell, 2005). Such argument scaffolding also improves students’ ability to evaluate arguments and counterarguments for an integrated conclusion (Nussbaum, 2008).

Researchers assert that argument visualization tools can improve the ability to understand and present reasoning (Davies, 2009). A traditional text-based argument requires the reader to memorize details and sequences of information, and its structure is subject to individual interpretations. Therefore, it is heavily demanding on the reader’s short-term memory and not reliable in terms of representation accuracy (van Gelder, 2003). Argument visualization simplifies arguments by making the logical structure clear and unambiguous to the reader. As analogized by Davies (2008), argument maps are to arguments as roadmaps are to verbal directions. It is believed that argument visualization reduces cognitive load, promotes clarity and rigor in thinking, and leads to deeper understanding of an issue. When learners are freed from the complex task of extracting the structure of an argument, they have more cognitive resources to dedicate to understanding the abstract structure and content of the argument itself (Ainsworth, 2006; Morgan, 2006). Visual representations can scaffold learners’ information evaluation and integration in order to communicate and solve controversial questions (Nussbaum, 2008). Argument visualization also provides support and helps learners stay more focused throughout the task in comparison to the text-only format (Suthers & Hundhausen, 2003). AVTs sustain cognitive reflection because an argument diagram can be revised, manipulated, recorded, and shared. Argument visualization can also help identify student needs. Students’ graphical representation of their arguments is a good indicator of their thinking and where they may need help in the learning process. When used in a computer-supported collaborative learning environment (CSCL), AVTs function as knowledge representation tools that scaffold collective learning. They initiate negotiation among learners and elaboration on previously presented information (Suthers & Hundhausen, 2003).
2.3.2. Theories Germane to Argument Visualization and AVT

As a new and growing field, many theories underlie argument visualization. Here I discuss several theories that are relevant to this thesis. These theories undertake essential roles in the development of argument visualization and AVTs: 1) cognitive load theory, 2) schema theory, 3) cognitive tools theory, and 4) multimedia learning theory. It is inherently challenging to discuss each theory separately because they intertwine with one another in an intricate fashion. For example, cognitive load theory assumes schemas as the building blocks of knowledge; multimedia-learning theory relies on cognitive load theory and schema theory to guide design and evaluate outcomes; and cognitive tools theory also aligns with schema theory. For presentation purposes, the theories are discussed in separate sections on their definitions, their contribution to education, and their connection with AVT. No parsimony is assumed across these theories.

Cognitive Load Theory

Cognitive load theory (CLT) deals with the capacities and limitations of the human cognitive architecture. It is based on the premise that the human cognition has limitations when processing information. Cognitive load is the total mental resource required of the working memory when an individual performs a cognitive task, such as learning or solving problems (Sweller, 1988; Sweller et al., 2011).

Cognitive load theory has been widely applied to learning and instructional design (Clark, Nguyen, & Sweller, 2006; Kirschner, 2002). It explains how people use their cognitive resources during learning and problem solving (Sweller, Van Merriënboer, & Paas, 1998; Sweller & Chandler, 1994; Sweller, 1988; van Merriënboer & Sweller, 2005). CTL postulates a limited working memory capacity, the representation of information with schemas, and the unlimited capacity of long-term memory (Sweller & Chandler, 1994). A schema is an organized mental structure with related categories of information. Long-term memory is believed to hold a vast amount of schemas as the building blocks of knowledge. The limitation of working memory rationalizes cognitive overload when too many blocks of new information are being processed simultaneously (Kalyuga, Chandler, Tuovinen, & Sweller, 2001; van Merriënboer & Sweller, 2005).
According to CLT, each fully learned schema is usually processed as one item. Thus, information processing in working memory becomes more efficient when smaller items are consolidated into larger ones. Retrieving consolidated schemas from long-term memory, even highly complex ones, does not seem to burden working memory. This makes schema consolidation an important goal for education.

According to CLT, there are three sources of cognitive load: intrinsic, extraneous, and germane cognitive load. Intrinsic cognitive load is inherent to the nature of the instructional tasks and materials. It is dependent on the complexity of the domain as well as the learner’s knowledge level with the topic. Intrinsic cognitive load can be manipulated, for example instructors may reduce learners’ intrinsic cognitive load by decreasing the interconnectedness among components of a task (Clark et al., 2006). Extraneous cognitive load is imposed by the formatting and presentation of instructional tasks and materials that are irrelevant to the instructional goals. This source of cognitive load is considered unfavourable and an uneconomical consumption of mental resources. Extraneous cognitive load should be minimized for better learning efficiency and outcomes (Clark et al., 2006), especially for complex topics and domains (van Merriënboer & Sweller, 2005). Germane cognitive load, on the other hand, is a favourable use of cognitive resources that generates better learning through schema acquisition and automation (Paas & van Merriënboer, 1994). It is generated by instructional design where learners are engaged in active information processing that is directly related to schema acquisition (Kirschner, 2002).

CLT criticizes traditional instructional strategies for neglecting the limitations of the working memory (Schnitz & Kürschner, 2007), thus minimizing extraneous cognitive load and managing intrinsic cognitive load have become a central consideration for contemporary instructional design. Moreover, instructional designs that engage students in activities other than schema acquisition and automation are likely to overdraw students’ cognitive resources (imposing extraneous cognitive load). Therefore, an important focus of instructional design ought to be schema construction and automation within the limits of cognitive load (Sweller, 1994).
The connection between CLT and AVT lies in the conviction that for a visualization tool to be effective, the designer must follow the principles of how the human cognitive system works. Argumentation tasks generally demand more cognitive resources (Bernardi & Antolini, 1996; Coirier et al., 1999; Kuhn, 2005; Nussbaum, 2008). This may help explain why students usually have difficulties engaging in argumentation activities without proper scaffolding.

Some information in a graph can be signalled with more salience than it can be in text. Consequently, graphical representations may demand less cognitive resources because they may require less searching of information compared to text based representations (Cox, 1999). Additionally, a reader may recognize and interpret information easier from a graph than from a text due to the increased saliency provided in the graph. Graphs also limit abstraction and make inferences easier because usually all information needed for inferences locates at the same spot explicitly. Text-based representations, on the other hand, may demand more cognitive resources because they require the recognition of implicit information, such as an implied logical or temporal sequence (Larkin & Simon, 1987).

Despite the widely appreciated proverb "a picture is worth ten thousand words" (Larkin & Simon, 1987), not all applications of graphical representations benefit learning. Some factors may impose extraneous cognitive load. For example, in a collaborative environment, a participant needs to spend more time and effort to make his representation clearer to others. Therefore, sharing representations may increase cognitive cost (van Bruggen, Kirschner, & Jochems, 2002). Instructional designers need to bear in mind that using argument visualization benefits learning as long as the cognitive cost associated with the graphic representation is outweighed by the benefits brought forth by the visualization.

**Schema Theory**

Schemas are hypothetical mental configurations that contain a set of related categories, relationships among the categories, and the criteria of what’s important within the configurations (Anderson, 1984; Rumelhart, 1980; Sweller, 1994). According to cognitive psychology, much of human knowledge is stored in long-term memory as
schemas, which can be activated for cognitive tasks such as comprehension and problem solving (Rumelhart, 1980).

Schema theory has been widely used by educational psychologists to decipher perception, comprehension, remembering, inferencing, problem solving, and learning (e.g., Bigenho, 1992; Sadoski, Paivio, & Goetz, 1991). As described by Reznitskaya and Anderson (2002), schema theory is useful for education for the following reasons: a) schemas determine the selection of information for perception and the allocation of cognitive resources; b) meaning construction is affected by schemas through the integration of new information with prior knowledge; c) a schema contains slots which provide puzzle-fitting types of clues to encapsulate certain information. An individual needs little effort to learn information that fits the slots; d) schemas enable inferencing when perceived information is incomplete. The slots predict expected information and guide interpretation to match the structure, and e) schemas guide the process of remembering. Andersen (1984) also noted that schemas assist tasks like editing and summarization. The reason is that a schema includes conditions of what information is important within the structure. The reader can construct summaries that include significant ideas and exclude negligible ones.

It is important to note that, as discussed earlier in CLT, schema construction and automation should work within a learner’s cognitive limits. The reason is that before new or updated schemas are stored into long-term memory, all crucial information is processed in working memory. If the processing overflows the cognitive capacity, schemas may fail to form. Therefore, instructional designs based on schema theories should also incorporate techniques to manage cognitive load in order to facilitate learning. According the schema theory, as learning occurs, more sophisticated schemas are constructed by aggregating and consolidating preexisting simpler, disconnected schemas. The newly formed, more complex schema is then processed in the working memory as one unit and consequently requires less cognitive resources. This is called schema consolidation, a type of schema construction. Another important role schema plays in learning is that the construction and activation process is automated after abundant practice. This is called schema automation. Automation frees mental resource
consumption and thus reduces cognitive load. Meaningful learning involves both schema construction and automation (van Merriënboer & Kester, 2005).

Schema theory and AVT are connected through the concept of argument schema, defined as a mental structure that characterizes the rhetorical organization of an argument or a network of connected arguments, with standards for plausible reasoning, as well as related social and cognitive understanding and applications of argumentation (Chinn & Anderson, 1998; Reznitskaya & Anderson, 2002; Reznitskaya et al., 2001; Reznitskaya, Kuo, Glina, & Anderson, 2009). Researchers believe that argument schemas bring many educational benefits, including “directing student attention to argument-relevant information, facilitating argumentation construction, comprehension, and refinement, organizing argument-relevant information, providing students the basis for examining one’s own and the opponents’ arguments, and assisting recall of argument-relevant information” (Reznitskaya & Anderson, 2002, p. 321). As mentioned earlier, students often lack argumentation skills. One possible explanation is the lack of argument schemas in their knowledge base, particularly advanced schemas (Hidi & Hildyard, 1983; Knudson, 1992b; Scardamalia & Bereiter, 1986).

Some well-known argumentation theories are often used as prototypes for argument schema, such as the Toulmin model and Walton’s argument schemes (Andriessen & Coirier, 1999; Reznitskaya et al., 2009). For example, Reznitskaya et al. (2009) described a basic argument schema that may include a claim and its supporting reason as well as a more developed and complex argument schema that may include claims, reasons, warrants, backings, qualifiers and rebuttals. Akin to general schema theory, argument schemas also contain the relationships among its components and criteria on what information is important (Anderson, 1984). As mentioned previously, AVTs are tools that display components of arguments and the relationship among the components. The substantial overlapping of the definitions indicates a close tie between argument schema and argument visualization tools. It is reasonable to postulate that AVTs ultimately reflect argument schemas and that AVTs facilitate, inter alia, learning that involves argument schemas. This postulation also aligns well with the external representation theory that a graphical representation integrates with and externalizes the
internal mental model (Zhang, 1997), in this case an argument schema. Similar to
generic schemas, there are numerous variations of argument schemas. As a result,
there is a wide variety of AVTs, and each may incorporate one or more particular
argument schemas. For example, Stegmann, Wecker, Weinberger, and Fischer (2007)
used a Toulmin-type AVT; Walton (2013) visualized various argument schemes;
Araucaria (Reed, 2004) is an AVT that incorporates both the Toulmin model and
Walton’s schemes; LARGO (Pinkwart, Aleven, Ashley, & Lynch, 2007) is based on
domain specific argumentation in law; and the Argumentation Vee Diagram (Nussbaum,
2008) visualizes a dialectical argumentation schema, particularly the thesis-antithesis-
synthesis schema (Basseches, 1979), as well as Walton’s critical questions (Walton et
al., 2008).

Cognitive Tools Theory

Numerous researchers defined cognitive tools and used various terms, such as
mind tools (Jonassen, 1996), cognitive technologies (Pea, 1987), and technologies of
the mind (Salomon, Perkins, & Globerson, 1991). The construct of “cognitive tool” had
shifted from its original theoretical root (anything or technique that facilitates schema
construction) and had come to represent technology-rich, computerized environments
that facilitate, enhance, or transform learning (Lajoie & Azevedo, 2006; Lajoie, 2007;
Nesbit et al., in press). In other words, cognitive tools are essentially interactive
computer programs designed for schema construction, activation, or instantiation (Nesbit
et al., in press).

An effective tool that is developed based on sound theoretical foundations can
be potentially very powerful to learning and instruction. It could potentially lead to
effective instructional interventions with children who a maturation-oriented
developmental theory (e.g., the Piagetian stage theory) might deem as too young and
unready. For example, according to Gal’perin and colleagues, the reason that some
cognitive tools seemed ineffective may not be due to their misfit with children’s
developmental stages, but because of the incomplete support those tools provided
(Arievitch & Stetsenko, 2000). This raises the importance of cognitive tool development
as a research strategy for the field of educational technology.
Lajoie (1993, p. 261) rationalized four reasons how and why cognitive tools assist learning, exemplified through a computerized environment (Sherlock I) that trains airmen on avionics troubleshooting. Each of the reasons she cites apply equally to the use of AVTs in scaffolding argumentation:

a) Cognitive tools support cognitive processes, such as memory or metacognitive processes;

b) They share or free cognitive load required by lower-level tasks so that more cognitive resources are available to develop higher-order skills;

c) They facilitate learners in performing cognitive tasks that would be over-challenging without the tool;

d) They allow learners to generate and test their hypotheses.

**Multimedia Learning Theory**

Multimedia learning can be defined as “building mental representations from words and pictures” and multimedia instruction as “presenting words and pictures that are intended to promote learning” (Mayer, 2005, p. 2). During the past two decades, theories and practices in cognitive psychology have provided a strong basis for instructional designs that involve multimedia tools (Sweller et al., 1998). The multimedia theory is based on three assumptions. The first is that the human mind processes information through two channels: one for verbal information and the other pictorial information. The second assumption is that similar to CLT, these channels have limited capacities. The presentation of both verbal and pictorial information enables the use of both channels. These channels complement each other on the type of information they represent: the verbal for abstract information and the pictorial for more intuitive information. The use of both channels produces better quality information processing that results in deeper understanding. The third assumption is that, during multimedia learning, learners construct knowledge in an active and meaningful way by paying attention to information that is relevant to the learning task, arranging information into a coherent, organized mental structure, and integrating it into their knowledge base (Mayer, 1999, 2014).

An important feature of AVTs is their integration of both graphical and textual elements. Thus, AVTs are inherently multimedia in nature. Most research-based multimedia learning design principles would consequently be valuable for developing
Effective multimedia learning requires designs that respect the human cognitive architecture. Multimedia design principles are based on three instructional goals: 1) reduce extraneous processing that is unrelated to learning; 2) manage essential processing in working memory that involves selecting and organizing relevant information; and 3) foster generative processing that involves reorganizing information and integrate it with prior knowledge (Mayer, 2014). Fifteen design principles were derived from these instructional goals (Mayer, 2014). Most of the principles apply to the design of AVTs, such as providing cues that highlight essential material, presenting coherent information with minimal distractions, and presenting words and pictures near to each other.

2.3.3. Empirical Findings on AVT Applications

Although AVTs have been applied in areas such as law, banking, and the military (van Gelder, 2007), their use in education has been rather recent. They are now enjoying increasing research attention in education, where they are being investigated as scaffolding to improve learning and argumentation skills. This section summarizes the more important findings on the instructional effects of using AVTs. To this end, a search was performed of empirical research in academic journal publications. For a broader review of argumentation systems, including common AVTs and non-educational applications, see Scheuer et al. (2010). The search of English peer-reviewed articles from 1980 to 2016 yielded 489 items from ERIC and PsycINFO combined. In order to catch studies that used parallel terms of argument visualization tools, such as argument map, argument diagram, Argumentation Vee Diagram, causal diagram, argument graphics, and argument graphic organizer, I used the conjunction of the following search terms (* is the wildcard symbol that is recognized by the databases searched):

- Issue: Argument*
- Scope: Map* OR diagram* OR graph* OR visualization
- Outcome: Education* OR learning OR instruction* OR pedagog*

I selected 23 articles to discuss empirical findings using AVTs in education settings based on the following criteria: By reading abstracts of items captured from the database search, studies with learning and education focus were kept; publications not pertinent to learning, such as law practices, were excluded, for example, Walton (2007).
Learning activities by law school students were kept, however, such as Carr (2003). Only empirical studies were kept for analysis.

**Description of Studies**

Due to the substantial rigor demanded by true experimental research, most researchers employed the quasi-experimental approach in their investigations (see Table 1). The lack of randomization and control made research findings less generalizable to the target populations. Design-based research provided valuable insights in the area, especially in improving research designs to generate desirable learning outcomes (such as Nussbaum, 2008). There were few qualitative studies investigating argument visualization in the literature, which offered more in-depth analyses of the issue (e.g., Davies, 2009; Okada & Buckingham Shum, 2008). Most research targeted postsecondary students while only a few studies were found with young subjects (such as Bell, 2001; Okada & Buckingham Shum, 2008; Schwarz, Neuman, Gil, & Merav, 2003). Learning activities ranged from reading (e.g., Dwyer et al., 2010) or writing tasks (e.g., Nussbaum & Schraw, 2007) to argumentation and critical thinking exercises (e.g., Butchart et al., 2009; Carr, 2003). In terms of scaffolding types, some researchers used generic or non-computer graphical tools for argument diagramming, such as Dwyer et al. (2010); others used specially designed tools with more advanced features (such as Pinkwart, Ashley, Lynch, & Aleven, 2009; Schank & Ranney, 1995).

Many studies reported that argument visualizations improved a measured outcome such as critical thinking (van Gelder et al., 2004), retention (Salminen et al., 2010), motivation (Zumbach, 2009), conceptual learning (Lund et al., 2007), as well as argumentation and reasoning skills (e.g., Harrell, 2011; Nussbaum, 2008). AVTs also enhanced collaborative learning, such as producing more information elaboration in collaborative writing (Suthers & Hundhausen, 2003). Some scholars have recommended that argument visualization exercises should be integrated into graduate education or in multi-disciplinary core subjects (Davies, 2009).

However, a number of studies failed to observe the anticipated outcomes from using AVTs (see Table 4). As examples, Carr (2003) failed to find improved reasoning
skills, Dwyer et al. (2010) failed to find better argument comprehension, Pinkwart et al. (2009) failed to find enhanced argumentation skills, Carwie (2009) failed to find pre-test to post-test gains in critical thinking, Suthers & Hundhausen (2003) failed to find knowledge gain, and Munneke, van Amelsvoort, and Andriessen (2003) failed to find more in-depth collaborative learning. Some researchers even asserted that graphical support had no effect on knowledge acquisition or argumentation skills (such as Zumbback, 2009).

**Findings on Educational Benefits**

Caution is warranted when drawing conclusions about the benefits of argument visualization because, as mentioned previously, the relevant research has produced mixed results. If, as seems likely, this is due to the heterogeneous research design and AVTs used in the research, one would expect any conclusions to be conditional. My reading of the 23 selected studies suggested that the efficacy of argument visualization interacts with the design features of a particular implementation, such as a) how the AVTs were used, b) the instructional context, c) student characteristics, d) technical features of the tool, and e) quality of research. This list is not intended to be exhaustive, but rather a starting point for better understanding of how to best use AVTs in improving learning.

**How Argument Visualization Tools are Used**

Research showed that when reading argument maps, in comparison with reading texts, cued information recall was enhanced regardless of the lengths of the materials (Dwyer et al., 2010). Whether the maps were coloured did not have an impact. These effects were obtained when controlling for students’ verbal and spatial reasoning skills. A follow up study consisting of three experiments (Dwyer, Hogan, & Stewart, 2013) confirmed enhanced cued-recall for subjects who read argument maps compared to those who read texts (as in experiment 1 and 2). However, an additional test of delayed cued-recall showed faster decrease for the argument map group, reducing the gap between the map and text reading conditions (as in experiment 2). In experiment 3, the researchers found that subjects who read a text and completed an argument map outperformed those who only read the text and took notes on cued-recall.
A larger number of studies investigated learning effects from the creation or construction of argument diagrams. Critical thinking has been the most often studied outcome in this area, probably because (as described previously) educators generally believe that there is a close relation between critical thinking skills and argumentation. The ability to accomplish cognitive tasks such as reconstructing, understanding, and evaluating arguments is considered an important aspect of critical thinking skills (Ennis, 1987; Fisher & Scriven, 1997; Kuhn, 1991). Studies frequently show that the construction of argument diagrams improves students’ critical thinking skills (Butchart et al., 2009; Carwei, 2009; Davies, 2009; Harrell, 2011; van Gelder et al., 2004). The improvement of a one-semester implementation with undergraduate students can be as large as 0.8 standard deviation (as in van Gelder et al., 2004) on a standardized instrument, the California Critical Thinking Skills Test (CCTST™). This gain is believed to be equivalent to at least three to four years of undergraduate education. Other researchers have reported similar but smaller results, such as Twardy (2004) with a 0.72 standard deviation change, and Butchart et al. (2009) 0.45 standard deviation.

In a design-based research study, Bell (2001) found student arguments continuously improved when constructing argument maps, in terms of argument complexity, customization, and quality. Students could also better evaluate arguments after constructing Argumentation Vee Diagrams (Nussbaum, 2008), an argument visualization tool with critical questions that guide student arguments, such as “Is there a compromise or creative solution?” (p. 553). The construction of argument diagrams also improves retention (Salminen et al., 2010), motivation (Zumback, 2009), conceptual learning (Bell, 2001; Lund et al., 2007), and argumentation skills (Pinkwart et al., 2009). For example, students’ argument coherence and reflective thinking were improved through argument diagram construction (Schank & Ranney, 1995). Such argumentation skills were able to transfer to situations without the AVTs (Pinkwart et al., 2009; Schank & Ranney, 1995; Nussbaum, 2008).

The transfer effects appeared to be stronger for the creation of argument diagrams than for reading existing ones. Students demonstrated deeper understanding of the problem when constructing their own representation of an argument, and when the AVT was removed, argumentation skills learned such as argument comprehension
skills were retained. In other words, creation of argument maps helped students learn more (Easterday, Alevin, & Scheines, 2007). They argued that when students used an AVT, they must first comprehend the source information by forming a mental representation, then externalize the mental representation by constructing an argument diagram (thus reducing cognitive load in comparison to solving the problem internally without the diagram), and then interpret the diagram in order to make inferences to solve the problem. When students read a diagram, they only had the opportunity to interpret, without exercising their comprehension skills. This may help explain why Dwyer et al. (2010, 2013) failed to find effects on argument comprehension when students only read given argument maps.

Even the evaluation and modification of argument diagrams did not yield the same level of learning benefits compared to the creation of argument diagrams. Salminen et al. (2010) discovered that students tended to focus more on the content of the claims and ignore the interconnection among them when given an existing diagram. And students who constructed their own argument diagrams outperformed those who modified existing ones on prior knowledge reactivation measured by the amount of prior knowledge they incorporated in the diagrams created after the intervention (41% vs. 10%).

Instructional Context

When explaining educational outcomes of AVTs the instructional context of the implementation should not be ignored. For example, from Tables 2.1 to 2.4 we see an interconnected body of research generated by Australian researchers focusing on critical thinking with university students, almost uniformly using the same measurement tool, the CCTST. However, the effect sizes vary significantly. The setup of the instructional context may help explain such variation.

First of all, learning how to use an AVT can be time consuming, especially if students are not skilled arguers. They need substantial training and practice, otherwise they may feel rushed (Davies, 2009). However, researchers often do not report how much time was spent on training students how to use the tools. Studies that do provide this information show large disparity. Instruction and practice time vary from 10 and 12
minutes (such as in Suthers & Hundhausen, 2003) to six full tutorials (such as in van Gelder et al., 2004). And the study with the longest training on argument mapping happens to be the one with the largest effect size in the literature. It is curious that van Gelder (2003) also claimed that only a few minutes of such training should be sufficient. If this claim holds true, the large variation in learning outcomes across these studies suggests there may be moderating variables that account for it. However, most authors omitted the details of their training procedures.

What might be a reasonable moderating variable then? Nussbaum (2008) found that argument-diagramming tools could not be effectively used unless students were given instruction and feedback on argumentation strategies. According to Table 3, researchers almost universally included the instruction of argumentation or reasoning in addition to teaching students how to use the tools, probably because it was inevitable to mention constituents of an argument and their relations when explaining how argument mapping works. Fortunately, most researchers gave such instruction to both conditions. However, they often did not report whether the instructions were carefully controlled, so that the experiment group did not receive additional information about argumentation. This may create an unfair situation for the control group, especially when the instruction on the AVT is extensive (such as 6 tutorials). If not precisely controlled, the actual instruction on argumentation or critical thinking could be intensified for the experiment group from extensive training about the tool. This may help explain the particularly large effect size found in van Gelder et al. (2004) on critical thinking gains.

The need for instruction and feedback on argumentation strategies may also help explain the non-significant effects in Carewei (2009). Although the implementation was a semester long, instruction from the teacher did not include giving feedback to students with their completion of critical thinking tasks. However, results from Carrington, Chen, Davies, Kaur, and Neville (2011) told a competing story: significant gain was obtained from allocating only 50 minutes to the argument mapping training plus further classroom practice, but no subsequent follow up or feedback throughout the whole semester. Additional factors may help explain these contradicting stories, such as choice of instruments (see a later section on quality of research).
Time-on-Task was also a significant feature of the instructional context that may have affected learning outcomes. For example, according to Table 2.4, most studies lasted from two weeks to a semester, with a few exceptions, such as Suthers and Hundhausen (2003) and Zumbach (2009). This may help explain the failure to obtain positive effects on knowledge acquisition in these two studies. In their report, Suthers and Hundhausen were aware that the time probably was too short for a difference in the learning outcome to develop.

Collaboration design was another important aspect of instructional context. AVTs were often integrated in CSCL environments as knowledge representation tools to promote individual and collective learning (such as in Janssen, Erkens, Kirschner, & Kanselaar, 2010; Salminen et al., 2010). Bell (2001) argued that such use of argument diagramming tools fostered meaningful collaboration, because the tools made students’ thinking visible to others. The tools then shaped further discussions toward the explanation and interpretation of students’ arguments. Morgan (2006) also states that argument visualization tools support knowledge sharing among team members. It is an efficient way to present the essence of a complex argument to another person. Argument diagrams also help focus group interactions and enable permanent records of group thinking as a community possession. However, there are pitfalls for combining AVTs with CSCL. As noted by Janssen et al. (2010), while AVTs and CSCL may complement each other, working on argument diagrams collaboratively may lead to problems that are inherent to CSCL, such as the disadvantages of communicating electronically. Students may have difficulty interpreting each other’s messages due to reduced vocal and facial cues. Additionally, co-constructing diagrams can make a challenging task more difficult. For example, students need to make sure that their team members understand their arguments. Although it may benefit learning to explain one’s opinion to others, it can also increase extraneous cognitive load (Van Bruggen et al., 2002). If the benefit and burden are not well balanced when setting up the instructional context, students may not learn from collaborative argument visualization (Janssen et al., 2010). This may help explain some unsuccessful results from combining AVTs and CSCL. For example, Munneke et al. (2003) expected the group reasoning to be deeper and broader when students co-constructed argument diagrams. However, they found that the co-constructed diagrams were often chaotic, and the argument-diagramming
tool was often used for taking notes instead of for visualizing arguments. No evidence was found that group reasoning was deepened or broadened.

**Student Characteristics**

As discussed earlier, researchers reported positive learning outcomes with argument visualization. However, some effects of AVTs seem stronger with low-achieving students, such as their beneficial effects on argumentation skills (Pinkwart et al., 2009). Consistent with this are Harrell’s (2011) two quasi-experimental studies on critical thinking, where students with poor argument analysis skills benefited the most from argument diagramming, while the high-achieving students did not seem to improve at all. This may help explain the non-significant effects in Carr (2003), where, as second year law students, subjects from both experimental and control groups might be considered to have high-aptitude for argumentation. In one of Harrell’s studies, medium-achieving students also benefited from using argument diagramming. Although the follow-up study did not repeat the same effects, Harrell was optimistic that average students could benefit from argument diagramming.

Another student factor that may impact the success of the AVT applications is student engagement. When they are not motivated to fully use the features of an argument visualization tool, such as the system feedback on whether their argument map construction is proceeding correctly, they do not do better than students using the traditional text-based tools (Pinkwart et al., 2009). Therefore, instructional design should focus on engaging students in exploring features of the computer scaffolding. Pinkwart et al. suggested that the low engagement in their follow-up study could have been due to the fact that the implementation of the tool only happened in the lab instead of the classroom, and students were not motivated to use the tool when it was not graded. They tended to use the tool minimally as suggested by their data. Therefore, they recommend using the tool in the classroom as the core part, as well as grading students’ argument maps. Another suggestion is to design the computer environment to include features that engage and inspire usage.
Technical Features

Naussbaum (2008) found that the presentation format of the graphics could draw student attention to different components of an argument. For example, when the conclusion box was presented at the top of the diagram, students tended to fill in this part first instead of at the end (as the case in Nussbaum & Schraw, 2007). Such premature commitment to a conclusion may create my-side bias or confirmation bias (Chinn & Brewer, 1998) as described previously. This may have impeded the argument counterargument evaluation and integration, which the researcher failed to obtain (Nussbaum & Schraw, 2007). A redesigned argument visualization tool in Naussbaum (2008) changed this feature into presenting the conclusion at the bottom of the diagram and eliminated such concerns.

As mentioned previously, the critical thinking gain from Butchart et al. (2009) was only half that of Van Gelder et al. (2004), possibly because ReasonAble1 offered more flexibility in constructing argument mapping. Although Butchart et al. claimed their software was flexible, the student feedback suggested that it was overly structured and did not accommodate students’ different learning approaches. Their AVT did not permit dragging or moving of proposition boxes and only allowed constructing an argument from the conclusion, while ReasonAble encouraged students to construct an argument in any way they prefer. On the other hand, as cautioned by Nesbit et al. (in press), node-link type of visual representations such as concept maps and argument maps can grow dauntingly complex for an average user when more than a few links are involved. Since it is not particularly unusual for an argument map to contain many nodes and links, AVT designers need to strike a balance between structure and flexibility. In other words, effective AVTs should provide flexibility that accommodates different student approaches to learning, but should not be overly open that little guidance is offered (see more details in Chapter 3).

Another technical feature Butchard et al. incorporated in their software design was live feedback, which kept students on the right track when constructing argument

1 Computer supported AVT used in Van Gelder et al. (2004).
maps. The idea of live feedback coincides with Pinkwart et al. (2009), Easterday et al. (2007), and Schank & Ranney (1995), who integrated AVTs into intelligent tutoring or simulation systems. Feedback can be flexible, non-directive, and on-demand so that the learner has more control over their learning process while having help available at hand. Students’ argumentation skills were found to correlate with their use of such feedback functions (Pinkwart et al., 2009). The computer interface design is also important in terms of engaging and motivating students. For example, Convince Me features a knowledge-eliciting interface and simulation-driven feedback. Students were more reflective when constructing argument diagrams and made more revisions to their argument structures (Schank & Ranney, 1995).

Additional features were sometimes incorporated into argument mapping software, such as a frame library with a menu of possible conceptual frames (e.g., “light can spread forever,” and “stars are huge light sources”) in Bell (2001, p. 452). Students frequently used these frames to help them elaborate and customize their own arguments. Nussbaum (2008) also added critical questions as cues for various argument evaluation strategies, such as “Which side is stronger and why?” (p. 553). Although no direct learning effects were observed, the researcher was hopeful that such critical questions would promote more reflective thinking and better decision-making.

At this point, an important question arises: how to best design AVTs that maximize learning effects? Currently, not much research has examined the relationship between design features of AVTs and their educational consequences and there are no consensual standards on how such tools should be designed.

Quality of Research

Quality of research affects generalizability and trustworthiness of findings. Research in this area is generally of moderate quality. This is reflected by the occasional omission of a control condition in a quasi-experimental design (see Table 1). Compared to true experiments, most quasi-experimental studies cited here lacked randomization, such as when recruiting the subjects from a course taught by the researchers. The additional lack of control conditions makes results even less generalizable. For example, although exhibiting the largest effect size in the literature, the van Gelder et al. (2004)
study did not compare learning outcomes between subjects who did or did not use the
diagramming tool. Although they made certain comparisons across studies, these
studies had different instructional contexts. Many mediating factors could be at play,
which are difficult to account for.

Absence of reliability and validity estimates is another major problem with studies
in this area. Although some studies demonstrated sound theoretical basis, with custom-
designed tools and questionnaires, few studies involved focus groups, piloting, or field
study, which are crucial for validating an instrument. These approaches help test the
instruments before actual applications and raise unexpected issues, so that the
developers have the opportunity to improve the instruments accordingly. Without these
steps, any uncertainty may contaminate the data collected and thus influence the
accuracy of results. Table 2 provides the validity and reliability estimates of the included
literature, or lack thereof. As mentioned previously, failure of obtaining desirable
outcomes may result from poor measurement tools. For example, Carwei (2009) used
the Health Science Reasoning Test (HSRT) as her critical thinking measure and
obtained non-significant results. While similar research using CCTST has been
consistently successful. Compared to HSRT, CCTST is a much more established and
better tested tool for critical thinking, yet very little validation has been conducted on
HSRT. Similarly, Suthers & Hundhausen (2003) were concerned that their 13-question
post-test was not a good measure for knowledge gain. Consequently, no such effect was
detected in their study.

It is noteworthy that a number of researchers who conducted quantitative studies
obtained feedback from students on their experience using AVTs. This provided valuable
information for explaining how and why argument visualization influenced learning. It
may also help identify factors other than the use of AVTs that contributed to the results.
For example, Butchart et al. (2009) found that some students felt handicapped by the
software when they were not very technology savvy with the computer, especially older
students. Nussbaum (2008) conducted a short survey with open-ended questions and
found that students perceived the tool as helpful in organizing their writing, separating
arguments and counterarguments, and keeping track of the relations among them.
Salminen et al. (2010) also conducted a short survey with one open-ended question on

55
pros and cons of using argument mapping. However, such feedback surveys are usually short and seldom touch on student motivation and engagement. More in-depth methods such as interviews may help collecting such data, which are especially helpful for studies that failed to achieve desirable results. For example, obtaining direct quotes from students about their low engagement of using the tool in the second study in Pinkwart et al. (2009), instead of explaining non-significant effects by intuitive assumptions.

**Summary**

This section reviews research on pedagogical effectiveness of AVTs. Argument visualization is potentially a very useful teaching method and a plausible technique to improve learning, but the learning outcome depends on various factors of the AVT implementations.

As a fairly new topic, there is limited empirical research investigating the pedagogical usefulness of AVTs. More research is needed to develop guidelines for practice, especially for teachers, on how to use such tools in the classroom. Research is also needed to inform instructional design on how to set up learning environments that better engage and motivate students in using these tools for learning activities. Standards are needed for how the instructor should teach students about the tools and about argumentation. In terms of software design, technical standards and evaluation methods for good designs are needed.

Methodologically, better quality research is needed examining learning effects from using AVTs. More rigorous research design and triangulation should be conducted for more trustworthy results. Currently, most research in this area involves quasi-experimental quantitative studies. Given the reasonable number of studies, it is possible to perform a meta-analysis on the use of argument diagrams and critical thinking. Very little qualitative research has been conducted in the area. Good qualitative studies will provide more in-depth examinations and explanations of the relationships between argument visualization and the interacting factors discussed in this chapter.

Finally, although instructional use of argument visualization still presents much uncharted territory, it has already shown significant promise for advancing learning in the
classroom. There is evidence that AVT interventions can be very powerful in terms of learning outcomes, as shown by the large effect size found by van Gelder et al. (2004). AVT researchers should search for ways to consistently produce such benefits.
<table>
<thead>
<tr>
<th>Paper ID</th>
<th>Method</th>
<th>N</th>
<th>Country</th>
<th>Age</th>
<th>Subject</th>
<th>Control Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Davis09</td>
<td>Qualitative</td>
<td>42</td>
<td>Australia</td>
<td>University</td>
<td>Economics</td>
<td>No</td>
</tr>
<tr>
<td>Dwyer10</td>
<td>Quasi-exp</td>
<td>400</td>
<td>Ireland</td>
<td>17–25</td>
<td>General</td>
<td>Yes</td>
</tr>
<tr>
<td>Dwyer13</td>
<td>Quasi-exp</td>
<td>131, 199, 136</td>
<td>Ireland</td>
<td>18–25</td>
<td>General</td>
<td>Yes</td>
</tr>
<tr>
<td>vanGelder04</td>
<td>Quasi-exp</td>
<td>51</td>
<td>Australia</td>
<td>Undergrad</td>
<td>Critical Thinking</td>
<td>No</td>
</tr>
<tr>
<td>Carr03</td>
<td>Quasi-exp</td>
<td>76</td>
<td>U.S.</td>
<td>Second year law school</td>
<td>Law</td>
<td>Yes</td>
</tr>
<tr>
<td>Bell02</td>
<td>Design-based</td>
<td>180</td>
<td>U.S.</td>
<td>Middle school</td>
<td>Physical Science</td>
<td>No</td>
</tr>
<tr>
<td>Butchart09</td>
<td>Quasi-exp</td>
<td>238</td>
<td>Australia</td>
<td>17–55</td>
<td>Critical Thinking</td>
<td>Yes</td>
</tr>
<tr>
<td>Nussbaum08</td>
<td>Design-based</td>
<td>45</td>
<td>U.S.</td>
<td>Undergrad</td>
<td>Ed-psych</td>
<td>Yes</td>
</tr>
<tr>
<td>Nussbaum07</td>
<td>Quasi-exp</td>
<td>84</td>
<td>U.S.</td>
<td>Undergrad</td>
<td>Ed-psych</td>
<td>Yes</td>
</tr>
<tr>
<td>Lund07</td>
<td>Quasi-exp</td>
<td>36</td>
<td>France</td>
<td>Secondary school</td>
<td>Non-specific</td>
<td>No</td>
</tr>
<tr>
<td>Easterday07</td>
<td>Quasi-exp</td>
<td>63</td>
<td>U.S.</td>
<td>Undergrad</td>
<td>Philosophy</td>
<td>Yes</td>
</tr>
<tr>
<td>Pinkwart09</td>
<td>Quasi-exp</td>
<td>28 &amp; 85</td>
<td>U.S.</td>
<td>Beginning law students</td>
<td>Law</td>
<td>Yes</td>
</tr>
<tr>
<td>Schank95</td>
<td>Quasi-exp</td>
<td>20</td>
<td>U.S.</td>
<td>Undergrad</td>
<td>Scientific Reasoning</td>
<td>Yes</td>
</tr>
<tr>
<td>Suthers03</td>
<td>Experimental</td>
<td>60</td>
<td>U.S.</td>
<td>&lt;20</td>
<td>Science &amp; Public Health</td>
<td>Yes</td>
</tr>
<tr>
<td>Janssen10</td>
<td>Quasi-exp</td>
<td>124</td>
<td>Netherland</td>
<td>11th grade</td>
<td>History</td>
<td>Yes</td>
</tr>
<tr>
<td>Carwei09</td>
<td>Quasi-exp</td>
<td>98</td>
<td>U.S.</td>
<td>Undergrad, senior</td>
<td>Community Nursing</td>
<td>Yes</td>
</tr>
<tr>
<td>Carrington11</td>
<td>Comparative</td>
<td>223</td>
<td>Australia</td>
<td>University</td>
<td>Accounting &amp; Marketing</td>
<td>No</td>
</tr>
<tr>
<td>Harrell11</td>
<td>Quasi-exp</td>
<td>269</td>
<td>U.S.</td>
<td>Undergrad</td>
<td>Philosophy</td>
<td>Yes</td>
</tr>
<tr>
<td>Okada08</td>
<td>Qualitative</td>
<td>20</td>
<td>U.K.</td>
<td>12–13</td>
<td>Science</td>
<td>No</td>
</tr>
<tr>
<td>Salminen10</td>
<td>Quasi-exp</td>
<td>16</td>
<td>Finland</td>
<td>16–17</td>
<td>Language</td>
<td>No</td>
</tr>
<tr>
<td>Zumback09</td>
<td>Quasi-exp</td>
<td>60</td>
<td>Austria</td>
<td>University</td>
<td>Environmental Issues</td>
<td>Yes</td>
</tr>
<tr>
<td>Schwarz03</td>
<td>Mixed-method</td>
<td>120</td>
<td>Israel</td>
<td>5th grade</td>
<td>Social Issues</td>
<td>Yes</td>
</tr>
<tr>
<td>Munneke03</td>
<td>Comparative</td>
<td>40</td>
<td>Netherlands</td>
<td>16–17</td>
<td>Science/Biology</td>
<td>No</td>
</tr>
<tr>
<td>Paper ID</td>
<td>Instruments</td>
<td>Reliability</td>
<td>Validity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>------------------------------------------------------------------------------</td>
<td>-------------------------------------------</td>
<td>-------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Davis09</td>
<td>Self-report questionnaire</td>
<td>Not reported</td>
<td>Not reported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dwyer10</td>
<td>Subset of Differential Aptitude Test; cued-recall test; comprehension test</td>
<td>Not reported</td>
<td>Not reported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dwyer13</td>
<td>Subset of Differential Aptitude Test; cued-recall test; delayed recall test; comprehension test</td>
<td>Rater reliability for recall test</td>
<td>Not reported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vanGelder04</td>
<td>CCTST; student questionnaires</td>
<td>Pre-established; not reported</td>
<td>Pre-established; criterion related validity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carr03</td>
<td>Practice final exam</td>
<td>Rater reliability established, not reported</td>
<td>Not reported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bell02</td>
<td>Conceptual understanding pre-post assessments; epistemological ideas about science assessment; self-report questionnaire</td>
<td>Not reported</td>
<td>Not reported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Butchart09</td>
<td>CCTST; LSAT</td>
<td>Pre-established</td>
<td>Pre-established</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nussbaum08</td>
<td>Essay content coding</td>
<td>Rater reliability</td>
<td>Not reported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nussbaum07</td>
<td>Essay content coding</td>
<td>Rater reliability</td>
<td>Not reported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lund07</td>
<td>Argumentation Diagram Analysis Method</td>
<td>Not reported</td>
<td>Not reported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Easterday07</td>
<td>Multiple-choice causal questions</td>
<td>Not reported</td>
<td>Not reported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pinkwart09</td>
<td>CTAT, self-developed aptitude assessment tool</td>
<td>Not done</td>
<td>Face validity;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schank95</td>
<td>Pre-post tests on scientific reasoning</td>
<td>Not reported</td>
<td>Not reported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suthers03</td>
<td>Group interactions coding</td>
<td>Rater agreement</td>
<td>Not reported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Janssen10</td>
<td>Essay and discussion content coding; knowledge acquisition multiple-choice pre-post tests</td>
<td>Rater reliability</td>
<td>Not reported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carwei09</td>
<td>HSRT, health science reasoning test</td>
<td>Not reported</td>
<td>Not reported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carrington11</td>
<td>CCTST; self-report questionnaire</td>
<td>Pre-established; not reported</td>
<td>Pre-established; not reported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harrell11</td>
<td>Argument analysis questions</td>
<td>Rater reliability</td>
<td>Not reported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paper ID</td>
<td>Instruments</td>
<td>Reliability</td>
<td>Validity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------------------</td>
<td>-------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Okada08</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salminen10</td>
<td>Argument rating; questionnaire</td>
<td>Not reported</td>
<td>Not reported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zumback09</td>
<td>Motivation questionnaire; prior knowledge on concept mapping questionnaire</td>
<td>Internal consistency</td>
<td>Not reported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schwarz03</td>
<td>Argument skills questionnaire</td>
<td>Not reported</td>
<td>Not reported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Munneke03</td>
<td>Group interaction coding</td>
<td>Not reported</td>
<td>Not reported</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.3  Conditions of Research

<table>
<thead>
<tr>
<th>Paper ID</th>
<th>Learning activities</th>
<th>Scaffold</th>
<th>Instruction on Argumentation or Reasoning</th>
<th>AVT Training</th>
</tr>
</thead>
<tbody>
<tr>
<td>Davies09</td>
<td>Reading; argument mapping</td>
<td>Rationale</td>
<td>Yes</td>
<td>1 h</td>
</tr>
<tr>
<td>Dwyer10</td>
<td>Reading</td>
<td>Argument maps</td>
<td>Yes, all 6 conditions</td>
<td>&lt;50 mins</td>
</tr>
<tr>
<td>Dwyer13</td>
<td>Reading; argument mapping</td>
<td>Argument maps</td>
<td>Yes</td>
<td>Yes, duration unclear</td>
</tr>
<tr>
<td>vanGelder04</td>
<td>Lectures; tutorial; argument mapping</td>
<td>Reason!Able</td>
<td>Yes</td>
<td>6 tutorials</td>
</tr>
<tr>
<td>Carr03</td>
<td>Legal argument construction</td>
<td>QuestMap</td>
<td>Yes</td>
<td>1 hour</td>
</tr>
<tr>
<td>Bell02</td>
<td>Lab experiments; writing assignments; debates</td>
<td>SenseMaker</td>
<td>Yes</td>
<td>Unclear</td>
</tr>
<tr>
<td>Butchart09</td>
<td>Lectures; tutorials argument mapping homework</td>
<td>Web based tool</td>
<td>Yes</td>
<td>Unclear</td>
</tr>
<tr>
<td>Nussbaum08</td>
<td>Reflective writing</td>
<td>Argumentation Vee Diagram</td>
<td>Yes, both conditions</td>
<td>Yes, duration unclear</td>
</tr>
<tr>
<td>Nussbaum07</td>
<td>Reflective writing</td>
<td>Graphic Organizer</td>
<td>Yes, all conditions</td>
<td>1 page instruction</td>
</tr>
<tr>
<td>Lund07</td>
<td>Debate</td>
<td>JigaDREW</td>
<td>Yes, both conditions</td>
<td>20 mins</td>
</tr>
<tr>
<td>Easterday07</td>
<td>Reading text; reading &amp; constructing argument maps</td>
<td>iLogos</td>
<td>Yes, both conditions</td>
<td>Unclear</td>
</tr>
<tr>
<td>Paper ID</td>
<td>Learning activities</td>
<td>Scaffold</td>
<td>Instruction on Argumentation or Reasoning</td>
<td>AVT Training</td>
</tr>
<tr>
<td>------------</td>
<td>---------------------------------------------------------------</td>
<td>--------------------</td>
<td>------------------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Pinkwart09</td>
<td>Legal arguments; writing assignments</td>
<td>LARGO</td>
<td>Yes, both conditions</td>
<td>Unclear</td>
</tr>
<tr>
<td>Schank95</td>
<td>Writing tasks</td>
<td>Convince Me</td>
<td>Yes, both conditions</td>
<td>5 h</td>
</tr>
<tr>
<td>Suthers03</td>
<td>Group learning; collaborative essay writing</td>
<td>Computer tool</td>
<td>Unclear</td>
<td>10-min instruction &amp; 12-min warm up</td>
</tr>
<tr>
<td>Janssen10</td>
<td>Collaborative inquiry; debate; collaborative essay writing</td>
<td>Diagrammer</td>
<td>Unclear</td>
<td>Unclear</td>
</tr>
<tr>
<td>Carwei09</td>
<td>Critical thinking assignments</td>
<td>Rationale</td>
<td>Unclear</td>
<td>Yes, duration unclear</td>
</tr>
<tr>
<td>Carrington11</td>
<td>Argument mapping; essay writing</td>
<td>Generic</td>
<td>Unclear</td>
<td>Yes, duration unclear</td>
</tr>
<tr>
<td>Harrell11</td>
<td>Argument analysis</td>
<td>Generic</td>
<td>Unclear</td>
<td>Yes, duration unclear</td>
</tr>
<tr>
<td>Okada08</td>
<td>Writing task; group argumentation, dialog mapping</td>
<td>Compendiu</td>
<td>Yes</td>
<td>Yes, duration unclear</td>
</tr>
<tr>
<td>Salminen10</td>
<td>Structured chat; debate; free construction of argument maps, modify and evaluation of maps</td>
<td>Generic</td>
<td>Yes, both conditions</td>
<td>Unclear</td>
</tr>
<tr>
<td>Zumback09</td>
<td>Reading hypermedia materials; argumentative analysis</td>
<td>Generic</td>
<td>Experiment groups</td>
<td>Unclear</td>
</tr>
<tr>
<td>Schwarz03</td>
<td>Debate; essay writing</td>
<td>Argumentative Map</td>
<td>Yes, both conditions</td>
<td>Yes, duration unclear</td>
</tr>
<tr>
<td>Munneke03</td>
<td>Collaborative writing; chatting; diagramming</td>
<td>Diagram tool in TC3</td>
<td>Unclear</td>
<td>Unclear</td>
</tr>
<tr>
<td>Paper ID</td>
<td>Intervention Duration</td>
<td>Collaboration</td>
<td>Outcome</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------</td>
<td>---------------</td>
<td>----------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Davies09</td>
<td>2 weeks</td>
<td>No</td>
<td>Critical thinking</td>
<td></td>
</tr>
<tr>
<td>Dwyer10</td>
<td>3 weeks</td>
<td>No</td>
<td>Comprehension*; cued-recall</td>
<td></td>
</tr>
<tr>
<td>Dwyer13</td>
<td>2–3 weeks</td>
<td>Mixed</td>
<td>Comprehension*; cued-recall</td>
<td></td>
</tr>
<tr>
<td>vanGelder04</td>
<td>12 weeks</td>
<td>Allowed</td>
<td>Critical thinking</td>
<td></td>
</tr>
<tr>
<td>Carr03</td>
<td>1 semester</td>
<td>Allowed</td>
<td>Reasoning skills*</td>
<td></td>
</tr>
<tr>
<td>Bell02</td>
<td>6 weeks</td>
<td>Yes</td>
<td>Individual conceptual progress; argument skills</td>
<td></td>
</tr>
<tr>
<td>Butchart09</td>
<td>12 weeks</td>
<td>Allowed</td>
<td>Critical thinking</td>
<td></td>
</tr>
<tr>
<td>Nussbaum08</td>
<td>4 week</td>
<td>No</td>
<td>Argumentation integration</td>
<td></td>
</tr>
<tr>
<td>Nussbaum07</td>
<td>Unclear</td>
<td>No</td>
<td>Counterargument refutation; argument counterargument Integration*</td>
<td></td>
</tr>
<tr>
<td>Lund07</td>
<td>4 days</td>
<td>Yes</td>
<td>Conceptual learning; better argument skills</td>
<td></td>
</tr>
<tr>
<td>Easterday07</td>
<td>Unclear</td>
<td>No</td>
<td>Reasoning skills</td>
<td></td>
</tr>
<tr>
<td>Pinkwart09</td>
<td>4 weeks; 6 weeks</td>
<td>No</td>
<td>Transferred argument skills: near and far</td>
<td></td>
</tr>
<tr>
<td>Schank95</td>
<td>Unclear</td>
<td>No</td>
<td>Argument coherence and revision</td>
<td></td>
</tr>
<tr>
<td>Suthers03</td>
<td>1 session</td>
<td>Yes</td>
<td>Argument skills; knowledge gain*</td>
<td></td>
</tr>
<tr>
<td>Janssen10</td>
<td>Six 50-min lessons</td>
<td>Yes</td>
<td>Argument skills; knowledge acquisition</td>
<td></td>
</tr>
<tr>
<td>Carwei09</td>
<td>13 weeks</td>
<td>No</td>
<td>Critical thinking skills*</td>
<td></td>
</tr>
<tr>
<td>Carrington11</td>
<td>1 semester</td>
<td>Allowed</td>
<td>Critical thinking improvement</td>
<td></td>
</tr>
<tr>
<td>Harrel11</td>
<td>1 semester</td>
<td>No</td>
<td>Critical thinking; argument analysis skills</td>
<td></td>
</tr>
<tr>
<td>Okada08</td>
<td>1 month</td>
<td>Yes</td>
<td>Scientific reasoning</td>
<td></td>
</tr>
</tbody>
</table>

2 An * marks the outcome that a study failed to obtain.
<table>
<thead>
<tr>
<th>Paper ID</th>
<th>Intervention Duration</th>
<th>Collaboration</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salminen10</td>
<td>Unclear</td>
<td>Yes</td>
<td>Prior knowledge reactivation; deeper understanding; knowledge construction</td>
</tr>
<tr>
<td>Zumback09</td>
<td>1 hour 35 min</td>
<td>No</td>
<td>Motivation; knowledge acquisition*, argument skills*</td>
</tr>
<tr>
<td>Schwarz03</td>
<td>20 hours</td>
<td>Yes</td>
<td>Quality of argument</td>
</tr>
<tr>
<td>Munneke03</td>
<td>Six 50-min sessions</td>
<td>Yes</td>
<td>Interactive argumentation; in depth collaborative learning*</td>
</tr>
</tbody>
</table>
Chapter 3.

Development and Deployment of the Dialectical Map

This chapter describes the development and design rationale of an argument visualization tool, the dialectical map (DM). Furthermore, it describes how the DM was used in three semesters of university-level biology instruction and the attitudes of the instructor and students toward the tool.

From the analytical review of the AVT literature in Chapter 2, I discovered two pivotal challenges that many argument visualization tools impose on their users. First, as a cognitive tool, an AVT should not create excessive extraneous cognitive load (Mayer, 2009; Sweller, 1988). However, many node-and-link types of argument maps represent arguments in a spatially complex and often confusing way due to the freedom allowed in the positioning of boxes (nodes) and arrows (links; see maps created in Belvedere as an example in Suthers, Weiner, Connelly, & Paolucci, 1995). Some AVT developers consider this an advantageous feature that allows authors of argument maps to personalize their representations. However, a reader would need substantial effort to disentangle webs and networks of claims, counterclaims, evidence, and occasionally warrants in order to understand the argument such AVTs intend to depict. The same concerns arise if the author of an intricate argument map decides to review, reappraise, or revise it at a considerably later time.

The second challenge is that many AVTs mainly focus on representing the logical relations among components of an argument, leaving the decision to evaluate and resolve thesis and antithesis largely to the user. Such lack of scaffolding may work for sophisticated arguers, who tend to habitually synthesize and balance arguments and counterarguments. However, for an unskilled arguer, such as a typical university student, this lack of metacognitive scaffolding poses two risks. The first is that there may be a tendency to commit my-side bias or make one-sided arguments without proper scaffolding for counterarguments. This may be a consequence of the extraneous
cognitive load demanded by generating counterarguments (Petrova & Cialdini, 2011). Secondly, an argument may also turn into perpetual iterations of ever more claims and reasons, which may gradually digress from the original arguments or learning goals when there is no explicit prompt signalling the completion of an argument.

The Argumentation Vee Diagram (Novak & Gowin, 1984; Nussbaum, 2008), on the other hand, gives direct support for argument-counterargument integration, for example, through critical questions such as “which side is stronger?” as well as providing a conclusion box. It is also easy to use, as arguments and counterarguments are organized on either side of the V. However, the Argumentation Vee Diagram appears overly textual other than a V-shaped line with horizontal lines on each side to write texts over. A series of critical questions are listed in plain text with no hierarchy underneath the graph (see Figure 3.1). The use of indentations of texts to represent claims and evidence does not offer clear representations of a) warrants, b) the hierarchical substructure of arguments whereby evidence can support reasons, which in turn support claims, c) the differing strengths of supporting reasons or evidence, and d) the ordering of reasons for rhetorical purposes. Some users might also struggle to reference which level represents what information from the indented lines. Moreover, the linking between arguments and counterarguments seems awkward since the lines cut through the V-shaped space at random angles, depending on where the user positioned the arguments and counterarguments they may later decide to link together. It only becomes more intricate and perplexing when the argument gets more complicated with more lines cutting through the V space and each other. According to Tukey (1990), such slanted lines make a graphical display unnecessarily busy and therefore should be avoided.

To address these challenges, a new type of argument map, the dialectical map (DM) was developed.

### 3.1. Theoretical Basis and Design Principles of the DM

Several theories offered the principal basis for our design criteria: Toulmin’s model (1958), Walton’s argumentation schemes (Walton et al., 2008), and dialectical reasoning (Ravenscroft et al., 2007). These theories have been discussed in more detail in Chapter 2.
Furthermore, as described in previous sections, a typical argument map displays relations among claims, reasons or premises, but does not usually provide tools necessary for students to synthesize arguments and counterarguments (see Figure 3.1). An Argumentation Vee Diagram (Nussbaum, 2008), on the other hand, contains critical questions that prompt students to compromise and integrate arguments and counterarguments. However, the Vee diagram lacks graphical illustration of the relations among reasons and claims by listing information over indented lines (see Figure 3.2).

Figure 3.1  Structure of a Simple Argument Map

When sketching the prototype of the DM, I had four design principles in mind. First it must be easy to use. As described in a previous section, complex representation of arguments can be confusing. Therefore, the DM should be straightforward, self-constraining, and free of maze-like complexity. Our design intentionally avoided excessive complexity by emphasizing the main components of the Toulmin hierarchy: claim and evidence. Warrant is built-in as an optional feature that the user can select based on their needs and level of argumentation skills.
Second, our design must achieve a good balance between graphical and textual representations of arguments. According to Mayer, people learn better when processing graphical and textual information than text or graphs alone (Mayer, 2005, 2009). Graphical displays make the logic and inferences in an argument more explicit than text only (Vekiri, 2002). Graphics also organize and group related information and consequently accelerate information retrieval (Larkin & Simon, 1987). To maximize comprehension and communication of arguments, a combination of graphics and textual components is often employed in practice rather than text or graphics only (Birdsell & Groarke, 2007). There has been no standard for the ultimate ratio between graphic and textual elements that are used to represent arguments in the literature. Researchers
have theorized that concept maps reduce extraneous cognitive load by balancing graphical and textual elements, and it may be that AVTs can offer similar benefits.

Third, our design must visually prompt the user to generate both arguments and counterarguments. Researchers have discovered that people are less likely to produce counterarguments when they argue (Kuhn, 1991, 2005). Typically they provide more evidence to support their own beliefs on the topic (Chinn & Anderson, 1998; Perkins, 1985; Perkins, 1989; Stein & Bernas, 1999; Wolfe & Britt, 2008). Identifying and refuting counterarguments are considered advanced argumentation skills that need explicit prompts and supports by the learning environment even for university students (Asterhan & Schwarz, 2007; de Vries et al., 2002; Harrell, 2011). Therefore, our design should provide scaffolding that assists and prompts counterargument generation, for example incorporating slots in the AVT to fill in counterarguments.

Fourth, our design must prompt the synthesis and resolution of claims and counterclaims. Effective argumentation requires not only the identification of counterarguments (Andriessen et al., 2003; Kuhn & Udell, 2007; Kuhn, 1991; Voss, Wolfe, Lawrence, & Engle, 1991), but also the process that Nussbaum and Schraw (2007) labeled as “argument-counterargument integration.” This integration includes the evaluation, weighing, and combination of arguments and counterarguments in order to reach a well-rounded solution. Thus, our design should also provide supports to reach a balanced conclusion, weigh the strength of arguments, rebut counterclaims, and form conclusions.

Based on these design principles, I drafted the prototype of the DM and finalized it with my supervisor. A computer programmer at Simon Fraser University wrote the codes. The DM is embedded in nStudy (see Chapter 4), a multi-facet computer-supported learning tool, which also allows the collection of log data, such as Time-on-Task. Furthermore, it has the potential to collect micro-level process data, such as time spent on creating each claim, to further examine students’ interaction with the DM.

The DM organizes arguments on the left side and counterarguments on the right side in coloured themes (Niu et al., 2015). It affords students to a) consider claims on both sides of an argument, b) judge and weigh the strength of a claim, c) support claims with evidence, d) support evidence with warrants, e) link claims and counterclaims that
directly oppose each other to signify rebuttals, f) reorder the points for linking and composing a rhetorical narrative, and eventually g) reach a balanced conclusion based on previous steps and write in the conclusion box positioned at the bottom of the DM.

3.2. The DM in Action: A User Manual

Users create a dialectical map by identifying and composing claims, evidence, and warrants in a computer-supported environment. They then draw a synthesized conclusion by evaluating arguments and counterarguments in a visually hierarchical structure. Nussbaum’s (2008) textual prompts, such as “Which side is stronger?” and “Is any of the arguments not as important as others?” are replaced by colour themes and weight numbers. These features function the same way as critical questions in prompting the arguer to evaluate and compare strengths of claims and counterclaims. The users may add as many claims, evidence statements, and warrants as they desire. However, a reversed pyramid-shaped margin pointing at the conclusion box encourages users to wrap up. A text box with a greyed out cue “write your conclusion here” is positioned at the bottom to visually prompt the arguer to reach a balanced conclusion.

Figure 3.3 shows how an instructor creates a DM assignment. First, the instructor logs in with a special link created for him or her. The link takes the instructor to a page with a textbox in which the assignment topic can be typed, usually in the form of a question. The DM system then generates a link that the instructor deploys to each student. Since the Toulmin structure often invites criticism on the subtlety of warrants (Anderson et al., 1997), instructors are given the option to exclude the warrant boxes if they consider their students unready for the concept. Some instructors may prefer using a full version of the DM with the warrant boxes when setting up their DM templates.
Figure 3.4 shows the starting interface a student sees from the link deployed by the instructor. It has two color themes for pro reasons and con reasons respectively. As students fill out a DM, the window may take up a full page. Users may adjust the screen size according to their own preferences. Occasionally, the DM window is open side by side with another window displaying the study material that the user will be reading in order to construct a dialectical map. The font size is between 12 and 16 on a full screen.

On top of the DM, there is the fixed question created by the instructor prior to deploying the URL. On the left side, students write down their arguments in the claim boxes, and counterarguments on the right side claim boxes. Each DM starts with one claim box on each side, which the instructor may choose to term as “argument” and “counterargument” or “pro reasons” and “con reasons” depending on the topic under study. The user may add more claims on each side of the map by clicking the button with a cue “add pro reason” or “add con reason.” These features visually prompt the consideration of claims on both sides of the issue under discussion.
Evidence boxes are visually supporting their corresponding claim box with an arrow pointing from the evidence to the claim box, marked by the word “supports,” which upon clicking switches into “opposes” depending on whether the user decides to use a specific piece of evidence to support an argument, or oppose a counterargument. The text boxes offer users sufficient room to write as much as needed.

Prompts indicating where to input claims, evidence and warrants are written in dimmed fonts, “Enter reason,” “Enter evidence,” or “Enter warrants.” This text is overwritten by user input. A double-headed arrow can be placed between arguments and counterarguments that are directly opposing each other, making the relationship between the claims and counterclaims more visual. This affords direct juxtaposing of argument and counterargument in order to prepare for rebuttals.

Intuitively, the “undo” and “redo” buttons allow users to cancel or repeat changes they made in the DM. One can undo as many steps as originally taken back to the starting point. A pull down menu at the end of each claim box contains numbers from 1 to 5 for the user to weigh and judge each claim based on its contribution to the overall argument, 1 being the weakest and 5 the strongest. The user can also delete the whole claim with associated evidence and warrants with the delete button included in this pull down menu. Deletion of an evidence or warrant box is done by a red “–” sign at the end of each text box.

At the bottom of the DM, there is a blank text box with a dimmed text that reads, “Enter conclusion.” Compared to a conclusion box at top of an argument diagram (see
Nussbaum, Winsor, Aqui, & Poliquin, 2007) positioning the conclusion box at the bottom potentially helps avoid prompting students to make up a conclusion first and then look for arguments and reasons to confirm this conclusion, which runs the risk of being prematurely committed to a conclusion or reinforcing the my-side bias.

Figure 3.5  A Completed DM

The DM is automatically saved upon closing the window. The instructor is able to view, modify, and download all student maps instantly from his or her login page. Figure 3.5 shows an example of a completed DM. It appears to be an overall balanced argument with similar numbers of arguments and counterarguments. The DM contains two pro reasons, two con reasons, seven evidence statements, one warrant, one link, and four weights of the strength for claims. The conclusion integrated arguments from both sides of the issue and compromises were made to resolve the issue.

An export button converts the content of a completed graphical DM into a text-based summary, that is, transcribing the entire argumentation process that occurred in...
the DM. Figure 3.6 shows the DM after it is exported into a text editor. It potentially gives the full structure of an argument essay including a list of arguments with their corresponding evidence and warrants, a list of counterarguments with their evidence and warrants, and a conclusion paragraph. This function potentially helps the user to write an argumentative essay using this summary as a guiding framework that contains all the essential components as well as an integrated final conclusion.

Figure 3.6 The DM and the Export Content

3.3. Implementing the DM in Authentic Settings

3.3.1. The Instructor

The instructor had taught university level biology for over 30 years before implementing the DM. She enjoys facilitating argumentation, guiding students to regulate their own learning or challenge their thinking, instead of merely lecturing. In the following three courses, she implemented the DM in order to teach students how to argue effectively.

3.3.2. Three Implementations of the DM

The DM was used to teach students argumentation skills in three biology courses. In Vertebrate Biology BISC 316 A (Implementation 1), Introduction to Biology
BISC 100 (Implementation 2), and BISC 316 B (Implementation 3), 92 non-majors and major’s Biology students used the online DM to complete three assignments in which they weighed evidence and made arguments in support and opposition to hypotheses drawn from three current biological controversies. For example, “British Columbia government should kill grey wolves to protect endangered caribou herds.”

In Implementation 1, BISC 316 (A) students were assigned to complete three dialectical maps over the semester. They were assessed on the quality of their reasoning and their use of relevant evidence to support arguments for specific hypothesis for each assignment. They were also provided with detailed feedback on the quality of their maps. Students were later asked to present arguments and evidence relevant to these controversies on later assignments and lecture examinations.

In Implementation 2, BISC 100, in addition to three similar DM assignments, students conducted a writing task before and after completing the three DMs. In the initial writing task, students wrote a summary of a brief text that presents arguments and counterarguments on a biological topic. Later in the course, students were again presented with the same brief text and asked to write a short essay on the topic addressed in the text. Implementation 3 was another BISC 316 (B) course. Course requirements and assignments were identical to the previous BISC 316 course (A). The instructor provided brief training on how to use the DM and opted to include the warrant boxes for all three assignments.

3.3.3. General Impression

The instructor expressed high interest and appreciation of the DM implementations in her courses. In a conference presentation (Sharp, Niu, & Nesbit, 2015) she commented that her involvement with the DM was “timely” because of a recent curriculum remapping in the science departments at Simon Fraser University. For Biological Science, at the end of the BSc program, students should be able to develop coherent arguments, supported by relevant and credible evidence, as recognized among other essential learning outcomes.

Student data were obtained from the university prior to the research. The mean GPA of the students in BISC 316 (A) was 2.74. They were on average 23 years old. Of
the participants, 53.1% were male. In BISC 100, mean student GPA was 2.88, mean age was 21 years, and 40% of the class was male.\(^3\) In an informal interview, the instructor commented that the DM was able to help both the high and the low achieving students improve their argument skills. The instructor was surprised when her biology major students were not able to argue well at the beginning of the course. For example, they tended to confuse evidence with explanations, and did not support their claims with evidence properly. Their argumentation skills improved substantially over the semester for all three implementations. Table 3.1 gives student marks on the three dialectical map assignments for the implementations. Improvement in the DM scores appeared to be consistent and steady. In addition, the instructor observed that students were better able to summarize arguments and present relevant evidence in support of these arguments in written assignments and the final exam following their training and practice with the DM argumentation tool. See Table 3.2 for the marking key for the DM assignments.

### Table 3.1 DM Scores

<table>
<thead>
<tr>
<th>Courses</th>
<th>Map 1</th>
<th>Map 2</th>
<th>Map 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>BISC 100</td>
<td>74%</td>
<td>83%</td>
<td>95%</td>
</tr>
<tr>
<td>BISC 316 A</td>
<td>67%</td>
<td>73%</td>
<td>83%</td>
</tr>
<tr>
<td>BISC 316 B</td>
<td>51%</td>
<td>79%</td>
<td>87%</td>
</tr>
</tbody>
</table>

\(^3\) Some student data for BISC 316 (B) is currently unavailable due to an ongoing TA strike at the university. Such data may not become accessible by the time I publish this thesis.
### Table 3.2  Marking Key for DM Assignments

<table>
<thead>
<tr>
<th>Marks</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Your reasons for supporting or opposing the statement are clear and well explained. The supporting evidence you provide is relevant to each reason. You clearly explain the evidence and explain how it supports the statement. You state a clear conclusion and briefly explain the evidence that supports your conclusion.</td>
</tr>
<tr>
<td>3</td>
<td>Your reasons for supporting or opposing the statement are not fully clear. Your supporting evidence is not fully relevant, is not fully clear, or is not fully understood. You fail to state a clear conclusion or do not clearly explain the evidence that supports your conclusion.</td>
</tr>
<tr>
<td>2</td>
<td>Your reasons for supporting or opposing the statement are not relevant or are not clearly explained. Your supporting evidence is not relevant, is not clear, or is poorly understood. You don’t reach a conclusion or you don’t explain why you reached your conclusion.</td>
</tr>
<tr>
<td>1</td>
<td>Your map is confusing and inaccurate and it is not clear you understand the statement or the evidence that supports or opposes it.</td>
</tr>
<tr>
<td>0</td>
<td>You failed to complete the assignment on time.</td>
</tr>
</tbody>
</table>

### 3.3.4. The Instructor Feedback

The instructor provided detailed feedback to each DM every student created. Most feedback contained an overall impression of a particular DM as being “weak,” “fine,” “good,” or “excellent” based on the marks received out of 4. According to the feedback, the students generally appeared to have a basic understanding of the content knowledge, as well as the concept of argumentation. When particular students appeared to be struggle with the assignments, the instructor provided both written feedback, as well as suggestions to meet face-to-face.

The feedback provided by the instructor generally covered a number of areas. First, it assessed the accuracy of domain knowledge the student exhibited, for example: “Your second con argument and the supporting evidence is a pair of correct statements,” and “You are using the word ‘vertebrate’ when you mean to say ‘tetrapod’ in this sentence. You must be more careful about your use of terminology.” Second, the feedback evaluated the presentation of the assignment. For example, comments included “I would prefer that you write in your own words. However, you do use quotation marks and cite your sources, so that is acceptable.” Third, it commented on the quality of an argument and suggested ways to improve, for example: “The fact that ‘Tiktaalik and early tetrapods were primarily aquatic’ does not provide evidence about the timing of the divergence of elpistostegid fishes and tetrapods,” and “You need to
expand on your final argument that ‘Ground dwelling early birds also developed wings to help them escape from predators. You provide no supporting evidence for this key argument, which is based on the ubiquity of WAIR in young and cursorial birds.” The instructor also challenged students with questions to clarify terminologies, statements, and arguments. For instance, clarifying questions included “What is the Fram formation,” “In the supporting evidence, you refer to a ‘strong anteroposterior component, which is different from the transverse stroke during flight.’ What does this mean?” and “As supporting evidence, you say, ‘In birds [,] hind limbs were used for running and jumping. Forelimbs are used for gliding. Indicating that the birds must have developed as tree-dwellers.’ What does this have to do with ridge gliding? Why is this relevant to the debate on the origin of bird flight?”

Such feedback directed student attention and encouraged improvement on specific aspects without providing an easy solution. Students would need to raise their level of attention on these issues in order to make improvement in later assignments.

3.3.5. Student Attitudes

According to the instructor, students responded very positively to their DM experiences. Here are some favourite quotes the instructor provided. One student said, “It challenged us to learn how to argue effectively.” Another said, “The DM challenged my ability to argue a topic. I thought I had skills [in argumentation] before, but I don't think I was actually very skilled.”

In general, the following themes emerged from synthesizing students’ feedback. First, a few students commented on the organizational assistance the DM offered. Examples included statements such as “I like the fact that it’s an organized way to present the various arguments, very visual and easy to follow.” as well as “[the DM] helps you organize your thoughts in a concise manner.” Second, several students mentioned how the DM helped their understanding, learning, and recall of the study materials. Examples of this included “It helps a lot to understand the course material,” “I really had to thoughtfully think about the materials [,] read and then separate those thoughts into pros and cons, which really helped to integrate the knowledge I learned,” and “Finding and providing arguments helped me learn the concept. The act of finding material and rewording it in my own way made it easier to recall on tests.” Third, the
majority mentioned how it helps to learn how to argue well. Selected statements included “I think the dialectical maps are quite interesting as they allow you to think beyond just the reasons behind [a] certain fact/event. These maps allow us to think of the evidence to that reasoning and make a connection between them.” The fourth theme was clear representation of logic, for example, “The format can help me to answer [a] question with much more logic.” Fifth, the DM prompted thinking on both sides and involved critical thinking. An example included statements such as “It wasn't just marked on an ability to memorize material like the exams were. I felt the dialectical map was more fair because it showed more critical thinking.” Sixth, the DM helped students focus on the study topic, for example, “[The] map allowed me to focus more on reading articles.” And last but not least, the DM was easy to use, as cited in the comments “The simplicity is nice,” and “it was pretty self-explanatory.”

When prompted to give feedback on barriers/issues and improvements of the DM, most students made no negative comments. Several students mentioned a glitch of the software. The instructor later clarified that the glitch was caused by the deployment of a wrong URL, which as was resolved promptly. Nevertheless, we value constructive feedback. For instance, the lack of instruction on how to use the tool was mentioned several times. We will make improvements accordingly for future implementations.
Chapter 4.

An Experiment on Learning with the Dialectical Map

This chapter describes the method used to investigate the outcomes of studying a text using the dialectical map (DM). The chapter describes the following components of the study: a) participants, b) the study material, c) the Dialectical Map and instruments used to collect data, d) the experiment procedure, and e) the coding system developed to score non-numerical data.

4.1. Participants

Participants were 125 (31 male; age range: 17–43 yrs.) post-secondary level students (GPA ranging from 1.18 to 4.54). The majority were undergraduate students from a medium sized Canadian university. Twenty participants were enrolled in a graduate program. Among all participants, 74 participants were non-native English speakers. Tables 4.1 and 4.2 show the distributions of participants’ field of study and first languages.
### Table 4.1  Participants' Field of Study

<table>
<thead>
<tr>
<th>Field of Study</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health Sciences</td>
<td>14</td>
<td>11.2%</td>
</tr>
<tr>
<td>Business</td>
<td>28</td>
<td>22.4%</td>
</tr>
<tr>
<td>Communication</td>
<td>7</td>
<td>5.6%</td>
</tr>
<tr>
<td>Criminology</td>
<td>8</td>
<td>6.4%</td>
</tr>
<tr>
<td>Economics</td>
<td>6</td>
<td>4.8%</td>
</tr>
<tr>
<td>Education</td>
<td>5</td>
<td>4.0%</td>
</tr>
<tr>
<td>Engineering</td>
<td>10</td>
<td>8.0%</td>
</tr>
<tr>
<td>Humanities</td>
<td>6</td>
<td>4.8%</td>
</tr>
<tr>
<td>Mathematics</td>
<td>4</td>
<td>3.2%</td>
</tr>
<tr>
<td>Natural Sciences</td>
<td>5</td>
<td>4.0%</td>
</tr>
<tr>
<td>Psychology</td>
<td>17</td>
<td>13.6%</td>
</tr>
<tr>
<td>Other</td>
<td>15</td>
<td>12.0%</td>
</tr>
</tbody>
</table>

### Table 4.2  Participant First Language

<table>
<thead>
<tr>
<th>First Language</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing</td>
<td>7</td>
<td>5.6%</td>
</tr>
<tr>
<td>Arabic</td>
<td>1</td>
<td>0.8%</td>
</tr>
<tr>
<td>Chinese</td>
<td>39</td>
<td>31.2%</td>
</tr>
<tr>
<td>English</td>
<td>43</td>
<td>34.4%</td>
</tr>
<tr>
<td>Farsi</td>
<td>2</td>
<td>1.6%</td>
</tr>
<tr>
<td>French</td>
<td>8</td>
<td>6.4%</td>
</tr>
<tr>
<td>Korean</td>
<td>5</td>
<td>4.0%</td>
</tr>
<tr>
<td>Persian</td>
<td>1</td>
<td>0.8%</td>
</tr>
<tr>
<td>Portuguese</td>
<td>1</td>
<td>0.8%</td>
</tr>
<tr>
<td>Punjabi</td>
<td>1</td>
<td>0.8%</td>
</tr>
<tr>
<td>Russian</td>
<td>2</td>
<td>1.6%</td>
</tr>
<tr>
<td>Spanish</td>
<td>2</td>
<td>1.6%</td>
</tr>
<tr>
<td>Vietnamese</td>
<td>2</td>
<td>1.6%</td>
</tr>
<tr>
<td>Other</td>
<td>11</td>
<td>8.8%</td>
</tr>
</tbody>
</table>
4.2. Materials and Instruments

4.2.1. The Text for Study

The main reading material was a narrative on shale gas extraction and hydraulic fracturing (aka: fracking, see Appendix A and B for full texts of the reading materials). There were 16 paragraphs and 73 sentences in this text, which were divided into 97 Idea Units. An Idea Unit is a sentence or a component of a sentence that conveys a meaningful notion. A sentence may contain one Idea Unit, or it may contain multiple Idea Units (see Appendix C for a list of all Idea Units). The identification of an Idea Unit also depends on its relationship with other Idea Units, for example, in sentence #37, health and environmental issues were discussed in more detail respectively in later sections, therefore they were treated as separate Idea Units even though there was no verbal markers separating them in the same sentence, which is a segmentation approach used by some researchers (e.g. Mayer, 1985).

Thirty-two of these Idea Units potentially supported the pro reasons for fracking (32.99%), 32 con reasons (32.99%), and 33 Idea Units were intended to provide neutral information (34.02%). The text had an expository layout rather than being manifestly argumentative. The rationale was that readers were given the opportunity to demonstrate their argumentation skills by identifying supporting or opposing claims and evidence. In addition, there were intentional uncertainties integrated into the text that required the readers to judge the weight and trustworthiness of the information in deciding how to use such information in forming their arguments. For instance, some information may be used to support either pro or con claims, and some seemingly pro information could potentially be used as con evidence.

4.2.2. The Dialectical Map (DM)

The DM combines and builds on argument maps (Dwyer et al., 2001) and the Argumentation Vee Diagram (Nussbaum, 2008). Users create DMs by identifying and composing claims, evidence, and warrants in a computer-supported environment. They then draw an integrated conclusion by evaluating arguments and counterarguments in a visually hierarchical structure. Figure 4.1 is the DM starting interface used for this
experiment. See Chapter 3 for more details on the design, theoretical foundations, and piloting of the DM in authentic classrooms settings.

![Dialectical Map Interface](image)

**Figure 4.1** The Dialectical Map Interface

### 4.2.3. The Highlighting Tool and Server Logs in nStudy

nStudy is an online learning tool that stores, tags, annotates, organizes, and analyzes information for learners (Winne, Doherty, Nesbit, & Hadwin, 2014; Winne & Hadwin, 2013). Besides numerous useful functions such as bookmarking, chatting, and archiving, nStudy also enables users to highlight web contents and documents in order to facilitate their learning process. In this study, only the highlighting feature was introduced to help participants mark text information that they considered important for the test later. Highlighting is an easy feature even for novice nStudy users. Instructions were given on site with a Mac computer in the lab. Since functions other than highlighting were not part of this experiment, nStudy was set to run in the background while participants took the survey so that they were not tempted to spend time exploring links and buttons of this multi-facet learning tool. Figure 4.2 shows how highlighting was performed in nStudy running in the background of FluidSurvey™ (see Section 4.3 for more details). nStudy also records user activity log data that reflects learning patterns and the use of learning strategies.
Shale gas extraction: Horizontal drilling and hydraulic fracturing

Natural gas is widely used for heating homes, generating electricity, and as fuel for manufacturing. Conventional natural gas is generally free gas trapped in relatively small, porous locations beneath rock formations and flows freely to the surface once the well is drilled. Unconventional natural gas is trapped in rock formations with very low permeability and does not flow freely to the surface after the well is drilled. Most of the growth in supply from today's recoverable gas resources is found in unconventional forms such as shale gas. Shale gas is usually located further underground than conventional gas and is more difficult to extract, but the potential supply is thought to be huge. Nations like China, Argentina, Canada and US each hold hundreds of trillions of cubic feet of technically recoverable shale gas.

4.2.4. The Critical Thinking Measure

Research has shown that critical thinking skills are closely connected to argumentation activities (e.g. Butchard et al., 2009; Carr, 2003; Carwei, 2009; Davies, 2009; Harrell, 2011; van Gelder et al., 2004). The ability to accomplish cognitive tasks such as reconstructing, understanding, and evaluating arguments are considered important aspects of critical thinking skills (Ennis, 1987; Fisher & Scriven, 1997; Kuhn, 1991). Studies in the literature (such as Twardy, 2004, Butchard et al., 2009, and van Gelder et al., 2004) often reference a standardized instrument with multiple-choice items, the California Critical Thinking Skills Test (CCTST), a critical thinking measurement tool.
that purportedly has outstanding psychometric properties. However, the developers of CCTST do not allow a portion of the full test to be administered and the whole test is overwhelmingly time consuming for the current experiment (32 scenario items in total and 45 minutes minimum to complete). In addition, the developers place restrictions on the data gathered by this instrument. Such policies prevent researchers from obtaining full access to the data. Therefore, we developed our own critical thinking measures with six multiple-choice items. These items were designed to resemble the CCTST in its use of challenging scenarios and answer options that evaluate the quality of reasoning on various topics. For example, question 1 assesses moral reasoning with the scenario on whether or not a lawyer should break client confidentiality when the client may pose harm to other people. See Appendix D for a list of these items.

4.2.5. The Need for Cognition Measure

The Need for Cognition (NfC) scale used in this study contains 18 self-report items that measure an individual’s tendency to engage in and enjoy complex thinking. The scale includes items such as “I really enjoy a task that involves coming up with new solutions to problems” and “I would rather do something that requires little thought than something that is sure to challenge my thinking abilities” (Cacioppo, Petty, & Kao, 1984, p. 307). It is a subset of a 34-item scale originally developed by Cacioppo and Petty in 1982.

NfC has been extensively studied in various contexts, such as religion, law, education, and social sciences. This instrument has been shown to be mature in terms of its validity and reliability measures. Both the original and the scale used in this study have been proven to possess good psychometric properties (Cacioppo et al., 1984; Cacioppo & Petty, 1982; Cacioppo, Petty, Feinstein, & Jarvis, 1996). In educational contexts, studies show that those who score lower on the NfC scale have a tendency towards less mindful learning behaviours, while those score higher display more diligence and openness to learning and new experiences. Higher NfC scorers are also more inclined to sort out irrelevant information from essential information and efficiently process and systematize that information (Cacioppo et al., 1984; Cacioppo & Petty, 1982). In a meta-analysis by Liu and Nesbit (2014), NfC was found to moderately correlate with academic achievement such as retention, comprehension, and transfer.
In terms of argumentation, Nussbaum (2005) connected learners’ NfC scores to their online argumentation patterns. Higher NfC scores positively correlated with the quality, complexity, and total number of argument claims made. Higher scores also indicated greater satisfaction from thinking, deeper engagement in the thinking process, and better motivation in applying their thinking skills.

4.2.6. The Achievement Goal Measure

Prior to the 1990s, it was commonly thought that performance goals and mastery goals were two parallel factors coexisting in the process of learners’ acquisition of knowledge (e.g., Dweck & Leggett, 1988). This model was the Achievement Goal Theory. If a learner reported an intrinsic interest in gaining knowledge, he or she was considered to have a mastery goal. If a learner reported a desire to satisfy a requirement or demonstrate competence, then he or she had a performance goal.

Theorists later revised the above scheme by incorporating an approach-avoidance dimension (Elliot & Church, 1997; Elliot, 1999; Elliott & Harackiewicz, 1996; Pintrich, 2000), thus creating two sub-factors for each of the previously mentioned factors. For example, within the performance goal there is a performance approach goal (PAP) and a performance avoidance goal (PAV). In the former, a learner strives to demonstrate competence and satisfy requirements as a primary consideration. In the latter, a learner intends to avoid failure. The approach-avoidance dichotomy was later applied to the mastery goal, resulting in a mastery approach goal (MAP) and a mastery avoidance goal (MAV), bringing the model to four factors (Elliot & McGregor, 2001).

Mastery approach is theorized to promote deeper cognitive engagement, while performance approach is theorized to focus on competing with others (Linnenbrink & Pintrich, 2002). Positive correlations have been noted between mastery goal applications and self-reports about deep learning or elaboration strategies (Elliot, McGregor, & Gable, 1999; Harackiewicz, Barron, Tauer, Carter, & Elliot, 2000; Ravindran, Greene, & DeBacker, 2000; Schraw, Horn, Thorndike-Christ, & Bruning, 1995). Studies have also found positive correlations between self-reported surface learning processes and performance approach goals (e.g., Harackiewicz et al., 2000; Ravindran et al., 2000).
The achievement goal questionnaire used in the current study was adapted from Elliot and Murayama (2008), with modifications to make the questions more general for the experiment rather than for a specific course/class the way the original measure was designed. The original version consists of 12 items. The current study used a subset of 9 items of this measure due to the low internal consistency of performance avoidance items in previous studies.

There has been limited research on how argumentation relates to achievement goal adoption. It is reasonable to postulate that argumentation encourages students’ adoption of a mastery goal approach since they both promote a high level of engagement (Hatano & Inagaki, 1991; Herrenkohl & Guerra, 1998; Mason & Samti, 1994).

4.2.7. Judgment of Learning

Judgment of learning (JOL) is a meta-cognitive judgment in which people assess how well they have learned specified information. It has important implications in learning. The pedagogical significance of JOL is two-fold. First, learners are monitoring and predicting their learning. Second, JOL helps learners manage their time and effort allocated to different tasks, that is, learners are inclined to assign more time to contents that are not yet grasped and less time for items already learned (Benjamin, Bjork, & Schwartz, 1998; Dunlosky & Hertzog, 1998; Dunlosky & Nelson, 1994). There are generally four categories of JOL based on different perspectives of learning and types of tasks learners execute. These include free-recall, ease-of-learning, ease-of-recognition, and paired-associate JOLs (Schwartz, 1994).

In studies involving judgment of learning, individuals typically give an overestimate of their JOL if the judgment is made immediately after studying. The more the JOL is delayed relative to the study period, the narrower the gap between JOL and reality (Nelson & Dunlosky, 1991). Other than delaying the JOL there are additional interventions that can improve the accuracy of JOL, for example asking learners to recall something they studied and then asking for their recall JOL.

No study to date has examined how argumentation affects JOL. It is reasonable to hypothesize that since argumentation has the potential to promote deeper cognitive
engagement and understanding, using the DM may modify a learner’s judgment of their learning. The current study investigated argumentation and JOL for recall. Participants were given two reading tasks, Text 1 and Text 2 (See Section 4.3 for more details). After each text had been presented for study, participants were asked to judge how well they would be able to recall later in the experiment the content of the text that they had just studied. For Text 1, participants were prompted to make the judgment as follows: “Now you will be asked to recall information from the page you just studied. How confident are you that you will be able to recall the information? Enter your estimate below, using any value from 0 to 100, where 0 means you definitely won’t be able to recall the information and 100 means you definitely will be able to recall it.” For Text 2, participants were prompted to make the judgment as follows: “At the end of the experiment, you will be asked to recall information from the text you just studied. You will be asked short-answer recall questions about the text. How confident are you that you will be able to answer such questions? Enter your estimate below, using any value from 0 to 100, where 0 means you definitely won’t be able to answer the questions and 100 means you definitely will be able answer the questions.” The wording of these prompts aligns well with what has been researched in the literature.

4.3. Procedure

I recruited all the participants on the SFU Burnaby campus based on the following criteria: a) being enrolled in an undergraduate or graduate program; and b) being at least 18 years old. Upon arrival at the lab and completion of informed consent forms, participants were given a brief introduction to the project, including how to complete the experiment on a computer and how to highlight in nStudy. They were also informed that there would be a test at the end of the experiment to examine how well they learned. Participants were not given much detail on what kind of test they would receive. Based on their academic career, it is reasonable to assume that participants expected a test that was similar to what they had experienced in their schooling, typically multiple-choice questions and short-answer questions, of which the test in this experiment was composed. Participants were then randomly assigned to the following three conditions: A) the DM Group (n = 44), which received training on both argumentation concepts and the Dialectical Map, and also used the DM for studying the source materials; B) The Argue Group (n = 40), which received training on
argumentation concepts; and C) the Control Group \((n = 41)\), which received no training on argumentation or the DM.

Before starting the reading tasks, participants were given questionnaires to collect data on prior differences regarding: a) Demographic information such as age, gender, experience with English, and field of study; b) Prior knowledge on the topic of fracking; c) Need for Cognition (18 items); and d) Achievement Goal Questionnaire (9 items).

The main goal of this design was to test the effects of the DM on participants’ learning outcomes. However, I was also interested in examining the effects of argumentation training on learning. This research design provided a means to separate the effects of a) the argumentation training plus the use of the DM and b) the argument training alone.

The study tasks were divided into three consecutive phases (See Figure 4.3). During Phase I, all three groups studied Text 1, which was the first page of the longer Text 2. Then all participants were tested for free recall of Text 1. Participants were then asked to estimate their learning from reading Text 1, that is, Information on judgment of learning (JOL1) was recorded for all groups. During Phase II, The DM Group (Condition A) received a 5-minute training on the concept of argumentation and a 5-minute training on how to use the DM. The Argue Group (Condition B) received a 5-minute training on argumentation and watched a 5-minute YouTube video on black holes. The Control Group (Condition C) watched a 10-minute YouTube video clip on black holes. The purpose of these videos was to remove differences that may be caused by uneven time allocated to different conditions. During Phase III, all groups studied Text 2. Each participant in the DM Group was asked to construct a Dialectical Map while studying the text. All participants were given the option to highlight the contents of the material that they considered important for their studying. A timer for 30 minutes was set commencing their reading of Text 2. Instruction to all participants was identical regarding how they were encouraged to study the material for the full 30 minutes. It was acceptable to spend a little more or less than 30 minutes on the main reading task, but participants were told that they would be nudged by the researcher to adjust their speed if they spent too much or too little time. Participants were then asked to describe on a scale of 0–9 (a) how well
they understood the material and (b) how interested they were in reading it. Participants were also required to make an estimate of their learning on Text 2 (JOL2).

We developed the DM for participants to present and evaluate arguments and counterarguments, supporting reasons, warrants, and a final conclusion in preparation for understanding the reading material in order to complete a test. The final test included 11 short-answer questions and a 6-item critical thinking measure. The 11 short-answer questions included a summary, five cued-recall questions, and five reasoning questions. Participants were not given extensive instructions on how to write the summary.

The experiment was administered through FluidSurvey™, a survey tool where participants input their responses and researchers later download the responses to a spreadsheet for further analyses. The experiment was embedded in nStudy, therefore log data could be collected through the nStudy server.

![Diagram of Experiment Procedures](image)

**Figure 4.3  Experiment Procedures**

### 4.4. Piloting

Prior to recruiting participants for piloting, the experiment was tested with two graduate students and one faculty member from the Education Department at SFU. The process provided valuable information on the improvement of the flow of the procedure, technical features, cultural and gender sensitivity, difficulty level, and the semantic clarity of the questions.
A pilot study involving 11 participants was conducted prior to the main experiment in order to evaluate the design and correct possible errors. Since most components of the three conditions (DM, Argue, and Control) were identical except that the DM condition included the DM in addition, I decided to run all 11 participants under the DM condition because it seemed that the DM might require extensive preparation and adjustments during administration. The DM Condition was piloted from beginning to end. At the end of the experiment, the researcher sought verbal feedback from participants on their thoughts of the DM and the experimental procedure. Similar to findings in Chapter 3, the majority considered the DM a very helpful tool in terms of understanding the reading material. One participant commented that the DM might not work very well for people who were already advanced at argumentation, but he could see how it would help those who were less advanced, especially with argumentative writing. According to the literature, most undergraduate students and even some graduate students lack argumentation skills (Nesbit, Niu, Mao, 2008). Therefore, I was optimistic that the DM would help most of the participants with their learning in the main experiment.

Improvements were made according to the feedback and observations obtained in the pilot study. I noticed that pilot participants were copying and pasting from the source text into the DM text boxes even though the option of paraphrasing was also given at the same time. Merely copying and pasting may diminish deep thinking and understanding of the source material if learners mechanically record text content without reflecting on it. Other researchers expressed the same concerns; For example, Igo, Bruning, and McCrudden (2005) believed that even when the number of words students were allowed to copy and paste was intentionally limited in order to promote cognitive engagement, typing the words into a text box may generate more intensive decision-making and deeper processing. The research design was modified accordingly for the main experiment: Participants were asked to paraphrase and summarize what they put in the DM in their own words. They were instructed to write two to three sentences maximum in each text box, and no copying and pasting was accepted.

The piloting procedure also detected difficulties participants encountered when having the DM and the reading material on different webpages, that is, participants needed to click back and forth to switch the main viewing window, which interfered with
their thinking and learning processes. In the actual experiment, I assisted each participant in opening the two pages within the same window and participants no longer complained about distractions.

The piloting procedure also informed the layout of survey questions and responses. For example, in the NfC questionnaire, answer options of each question were repeated several times on the same webpage, so that respondents did not need to scroll back to the top to consult what each numbered option meant.

4.5. Coding Schemes for Short Answers

I intended to capture the structure of participants’ memory and their understanding of the source material by using a series of coding schemes to evaluate their short-answer responses.

4.5.1. Scoring Free Recall

All participants wrote a free recall test on Text 1 before the interventions commenced. This test collected data on participants’ recall performance on: 1) number of Idea Units, 2) number of Main Ideas, and 3) number of errors made. Idea Unit was defined in a previous section. For ease of representation and analysis, a Main Idea was defined in this study as the central or principal theme of a paragraph (see Appendix E for a list of Main Ideas). An error was defined as erroneous information provided, either by misquoting the source text or making statements generally considered wrong relative to the text.

4.5.2. Scoring the Summary

In this study, all participants were asked to write a summary based on the main reading material (Text 2) they studied. Summarization has been theorized to demand deeper cognitive processing than rote recall. While rote recall tends to focus more on the details of a text, summarization involves understanding and expressing the macrostructure of a text (van Dijk & Kintsch, 1983). Readers often interpret the macrostructure based on their prior background knowledge through generalizing,
synthesizing, and condensing the microstructure into a coherent manifestation. Similar to how the Argumentation Vee Diagram (Nussbaum, 2008) draws student attention to the macrostructure of an argument, the DM construction process may also direct participants’ attention to the macrostructure of the text.

Participants in the DM condition did not have access to their DMs while responding to the test items. This eliminated the additional cues from the DM contents while they tried to recall information and organize their thinking. The coding system for the summary included: 1) the organization of information or the rhetorical structure, 2) number of Argument Markers used, 3) number of Neutral Idea Units used, 4) the coverage of study content measured by number of Main Ideas and Idea Units used, and 5) errors included in the summary. A partial mark was given for an incomplete Idea Unit or Main Idea used in the summary.

1) Organization

Regardless of the quality of the content and types of composition, a well-organized essay usually contains the following components: an introduction, a body, and a conclusion. An introduction is an overview statement that is most relevant to the rest of the response. A body is an expansion and description of Main Ideas that are related to the introduction. In some participants’ responses, an introduction was implied, instead of clearly stated, but the body clearly followed a list of premises or how the text material flowed. A full score was awarded to body in such responses. A conclusion is a concise statement that arises from the body that is also related to the introduction. One point was awarded for each of these components so that a summary could obtain a maximum score of 3 if all three components were present and a minimum score of zero if none were present.

2) Number of Argument Markers Used in The Summary

These markers are lexical indicators that signal the attempt to make an argument. For example, the connective “because” connects one or more claims to one or more reasons, “however” connects one or more claims to one or more counter-claims, and “furthermore” connects one or more claims to other claims. Keywords such as “argument,” “evidence,” and “reasons” were also included in the scoring of
argumentative structure of the summary. Each marker was given a score of 1. See Appendix F for a full list of argumentative markers used for this study.

3) Number of Neutral Idea Units Used in The Summary.

All Idea Units were tagged as Pro Idea Units, Con Idea Units, or Neutral Idea Units based on a latent theme in the reading material, which was whether or not fracking should be promoted. Neutral Idea Units provided information that contributed to neither side of the argument. Each Neutral Idea Unit was given a score of 1, and the sum was used for further statistical analyses.

The above three measures (in addition to Number of Reasons described in a later section) expressed participants’ tendency to argue when given a written task. Even though the summary task did not explicitly require argumentation, it is possible that the argumentation interventions, including the argumentation training and the DM, changed the way participants study a text. Argumentation schemas may be activated through the exposure to the concept of argumentation and the DM. According to Reznitskaya and Anderson (2002), such schemas influence information processing in a way that people will pay more attention to information that helps them build an argument.

4) Coverage

A good summary should be succinct but thorough and should therefore contain most of the important information from the source text. Participants were given points for the number of Main Ideas and Idea Units covered in their summary. Main ideas were not directly given in the source text. Additional effort was required for paraphrasing, integrating, and synthesizing Idea Units in order to extract Main Ideas. Participants were given one point for each Idea Unit they expressed, and one point for each Main Idea they expressed. These measures may be useful to characterise participants’ recall and comprehension of the source text.

5) Errors

Errors in the short-answer responses were defined the same way as for the free recall test. These errors may be useful in characterizing participant’s memory and their
understanding of the reading material. Therefore, it is worth examining such errors and how the DM may affect participant’s learning. Each error was given a score of 1 and the sums were used for further statistical analyses.

4.5.3. Scoring Cued-Recall and Reasoning Questions

Due to participants’ focusing heavily on Q1, as evidenced by the word count of responses compared to any other question (see Figure 4.4), the scoring of the remaining questions was simplified. This was necessitated by the lack of information provided in these responses.

Questions Q2 to Q6 were designed to collect data on cued-recall. For example, Q2 asked, “How does the combination of horizontal drilling and hydraulic fracturing make unconventional gas supplies commercially feasible?” This question cued participants to recall information on the process of fracking. Q2 to Q6 were scored on these variables: 1) Number of Idea Units and 2) Number of Main Ideas. Participants were given one mark for each Idea Unit and each Main Idea used in the answer. However, to measure whether or not participants truly understood the reading material, only answers that contained relevant information were desirable; therefore two additional variables were examined: 3) Number of Idea Units used correctly according to a scoring rubric developed for each question; and 4) Number of Main Ideas used correctly according to a scoring rubric developed for each question (See Appendix G for the questions and corresponding rubric).
Questions R1 to R5 (see Appendix H) were designed to assess participants’ reasoning, which reflected their understanding of the material and ability to apply argumentation skills. While the Toulmin model presents a complex argument structure that includes claims, evidence, warrants, backings, modal qualifiers, and rebuttals, researchers have investigated arguments using a basic model that contains a claim and its supporting reason (e.g., Angell, 1964). Due to the complexity of the Toulmin model and the limited words participants used in answering these questions, it was viable to use the basic model to evaluate argumentation and reasoning in the current study. For example, question R3 asked, “How do you think fracking could affect YOUR life and WHY?” which required participants to list their claims and reasons to support their claims in order to answer the question. R5 in particular presented a novel situation hoping to assess whether or not such reasoning skills were context-specific or participants were able to transfer them to different settings. These questions were scored by Number of Reasons. The term Reason refers to any supporting evidence or justification to a particular claim. For example, “Natural gas is cleaner energy compared to coal. Natural gas burning emits 50% less greenhouse gas than coal fire generation,” where the first sentence is a claim and the second is the supporting evidence of this claim. The Sum Reasons of all five questions was used for statistical analyses.
Chapter 5.

Results

To evaluate the effect on learning using the Dialectical Map, this experiment collected data on: 1) demographics including age, gender, first language, English experience, education level, field of study, and self-reported GPA; 2) cognitive dispositions including Need for Cognition and Achievement Goal Orientation; 3) nStudy server log records including Time-on-Tasks and highlighting activities; 4) perceptions of prior knowledge, interest in and understanding of the topic, and learning; 5) performance on a researcher-constructed critical thinking instrument; 6) free recall scores measuring retention of the first reading material, Text 1; 7) variables measuring argumentative writing in a summary of Text 2; 8) variables measuring cued-recall of Text 2; and 9) the variable measuring reasoning activities.

5.1. Missing Values, Outliers, and Normality of Distributions

Prior to data analysis, all variables were inspected for data entry accuracy, missing values, outliers, and normality of distributions. Missing values were detected through the SPSS command MVA. Most of the variables had a small percentage of missing values, ranging from 0 to 2.4%. Missing values of these variables were imputed using the EM algorithm, which obtains a maximum likelihood estimate. The variable for GPA had 17.6% missing values. Any imputation technique is likely to create error that could skew the analytical results. Therefore, subsequent analyses involving GPA excluded participants with missing GPA. All variables were examined for outliers by calculating z-scores and inspecting Mahalanobis distances (Tabachnick & Fidell, 2006). One univariate outlier was discovered for Age (z = 5.59). Through a winsorization process (Field, 2013), the value of this outlier was adjusted to lower its impact on the normality of the variable (Tabachnick & Fidell, 2007). The next highest z-score was z =
2.94, with a corresponding raw value of 33. The original value of the outlier was lowered to 34, which was one unit higher than the second highest value. No multivariate outliers were detected. Normality of distributions was inspected through SPSS Frequencies under Descriptive Statistics with histograms and normal curves. Distributions were mostly normal with acceptable skewness and kurtosis (with standard error values between -3 and +3, Tabachnick & Fidell, 2006). Data for 125 participants were retained for subsequent analyses.

5.2. Demographic Information

Table 5.1 contains the means and standard deviations (in parentheses) of participants’ age, years of English as the primary language for schooling (Years in English), GPA, and Year in Program. It also shows the gender makeup of each group.

<table>
<thead>
<tr>
<th></th>
<th>DM Group</th>
<th>Argue Group</th>
<th>Control Group</th>
<th>Pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>27.3%</td>
<td>25.0%</td>
<td>22.0%</td>
<td>24.8%</td>
</tr>
<tr>
<td>Age</td>
<td>21.77 (3.30)</td>
<td>22.66 (5.05)</td>
<td>21.02 (2.51)</td>
<td>21.81 (2.78)</td>
</tr>
<tr>
<td>Year in Program</td>
<td>3.17 (1.54)</td>
<td>3.44 (1.81)</td>
<td>2.81 (1.42)</td>
<td>3.14 (1.60)</td>
</tr>
<tr>
<td>Years in English</td>
<td>9.06 (6.01)</td>
<td>10.80 (5.55)</td>
<td>11.20 (5.64)</td>
<td>10.32 (5.78)</td>
</tr>
<tr>
<td>GPA</td>
<td>3.33 (.55)</td>
<td>3.04 (.59)</td>
<td>3.19 (.58)</td>
<td>3.20 (.58)</td>
</tr>
</tbody>
</table>

Table 5.2 Test of Homogeneity of Variances for Demographic Variables

<table>
<thead>
<tr>
<th></th>
<th>Levene Statistic</th>
<th>df1</th>
<th>df2</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>4.53</td>
<td>2</td>
<td>122</td>
<td>.01</td>
</tr>
<tr>
<td>Year in Program</td>
<td>1.90</td>
<td>2</td>
<td>122</td>
<td>.15</td>
</tr>
<tr>
<td>Years in English</td>
<td>1.00</td>
<td>2</td>
<td>122</td>
<td>.37</td>
</tr>
<tr>
<td>GPA</td>
<td>.32</td>
<td>2</td>
<td>100</td>
<td>.72</td>
</tr>
</tbody>
</table>

All the variables met the assumption of homogeneity of variance, except for Age, Levene(2,122) = 4.53, p = .01 (see Table 5.2). An independent-samples Kruskal-Wallis test showed no difference across groups on their distributions of Age, H(2) = 1.74, p = .42. A MANOVA test was run for the rest of the variables on the list except GPA, which had a large percentage of missing values. The MANOVA found no statistically detectable differences among the three conditions for Year in Program and Years in
English, Wilk’s $\lambda = .95$, $F(4,242) = 1.59$, $p = .18$. An ANOVA showed no statistically significant group differences for GPA, $F(2,100) = 2.10$, $p = .13$.

5.3. Self-Reported Information

As mentioned in Chapter 4, the AGO measure used in this study contained three dimensions: mastery approach (MAP), performance avoidance (PAV), and performance approach (PAP). According to the literature on the AGO measure, items 3, 6, and 9, represented the performance approach subscale (PAP), Cronbach’s $\alpha = .88$. Items 2, 5, and 8, represented the mastery approach subscale (MAP), Cronbach’s $\alpha = .78$. No improvement of the $\alpha$-values occurred by deleting any items for these two subscales. Items 1, 4, and 7, represented the performance avoidance subscale (PAV), Cronbach’s $\alpha = .76$. Deleting item 1 would make a small difference, bringing Cronbach’s $\alpha$ to .77. I decided to keep all items for each subscale. Items loaded on the same scale were aggregated with unit weights to form three separate variables for further analyses. The internal consistency of the 18 Need for Cognition items was acceptable (Cronbach’s $\alpha = .84$), therefore another scale (NfC) was created by aggregating all items with unit weights. These scales were used in subsequent analyses.

Table 5.3 shows the group means and standard deviations (in parentheses) of the Need for Cognition scale, the three Achievement Goal Orientation scales, participants’ subjective judgments of their own learning before and after the interventions (JOL1, JOL2), and participants’ perceptions on their prior knowledge, interest level and understanding of the subject they studied.
### Table 5.3  Descriptive Statistics for Self-Reported Information

<table>
<thead>
<tr>
<th></th>
<th>DM Group</th>
<th>Argue Group</th>
<th>Control Group</th>
<th>Pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior Knowledge</td>
<td>.61 (1.37)</td>
<td>.95 (1.63)</td>
<td>.61 (1.61)</td>
<td>.72 (1.53)</td>
</tr>
<tr>
<td>JOL1</td>
<td>54.41 (19.87)</td>
<td>63.17 (20.25)</td>
<td>58.24 (19.92)</td>
<td>58.47 (20.17)</td>
</tr>
<tr>
<td>Understanding</td>
<td>6.89 (1.77)</td>
<td>6.63 (1.98)</td>
<td>6.17 (1.78)</td>
<td>6.57 (1.85)</td>
</tr>
<tr>
<td>Interest</td>
<td>4.98 (2.36)</td>
<td>4.93 (2.53)</td>
<td>4.29 (2.61)</td>
<td>4.74 (2.50)</td>
</tr>
<tr>
<td>JOL2</td>
<td>66.93 (16.53)</td>
<td>60.80 (22.56)</td>
<td>56.31 (20.98)</td>
<td>61.49 (20.40)</td>
</tr>
<tr>
<td>NfC</td>
<td>63.52 (9.29)</td>
<td>64.05 (8.44)</td>
<td>59.88 (10.66)</td>
<td>62.50 (9.61)</td>
</tr>
<tr>
<td>PAP</td>
<td>10.67 (2.85)</td>
<td>10.65 (3.25)</td>
<td>10.83 (3.27)</td>
<td>10.72 (3.10)</td>
</tr>
<tr>
<td>MAP</td>
<td>11.98 (2.17)</td>
<td>11.82 (2.51)</td>
<td>12.00 (2.31)</td>
<td>11.93 (2.31)</td>
</tr>
<tr>
<td>MAV</td>
<td>10.54 (3.13)</td>
<td>10.46 (2.63)</td>
<td>11.34 (3.28)</td>
<td>10.78 (3.03)</td>
</tr>
</tbody>
</table>

All self-report variables passed the Levene test of homogeneity of variance (see Table 5.4).

### Table 5.4  Test of Homogeneity of Variances for Self-Report Variables

<table>
<thead>
<tr>
<th></th>
<th>Levene Statistic</th>
<th>df1</th>
<th>df2</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior Knowledge</td>
<td>.90</td>
<td>2</td>
<td>122</td>
<td>.41</td>
</tr>
<tr>
<td>JOL1</td>
<td>.07</td>
<td>2</td>
<td>122</td>
<td>.93</td>
</tr>
<tr>
<td>Understanding</td>
<td>.54</td>
<td>2</td>
<td>122</td>
<td>.58</td>
</tr>
<tr>
<td>Interest</td>
<td>1.11</td>
<td>2</td>
<td>122</td>
<td>.33</td>
</tr>
<tr>
<td>JOL2</td>
<td>2.72</td>
<td>2</td>
<td>122</td>
<td>.07</td>
</tr>
<tr>
<td>NfC</td>
<td>.98</td>
<td>2</td>
<td>122</td>
<td>.38</td>
</tr>
<tr>
<td>PAP</td>
<td>.98</td>
<td>2</td>
<td>122</td>
<td>.38</td>
</tr>
<tr>
<td>MAP</td>
<td>.84</td>
<td>2</td>
<td>122</td>
<td>.43</td>
</tr>
<tr>
<td>MAV</td>
<td>.62</td>
<td>2</td>
<td>122</td>
<td>.54</td>
</tr>
</tbody>
</table>
Table 5.5 contains the correlation matrix of self-reported variables. Values of \( r \) coefficient range from -0.33 to 0.69. There are a few moderate correlations, for example, between JOL1 and JOL2, between students’ perceived understanding of the study material (Understanding) and JOL2, and between Need for Cognition and JOL2.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Prior Knowledge</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Understanding</td>
<td>0.28</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Interest</td>
<td>0.28</td>
<td>0.39</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. JOL1</td>
<td>0.31</td>
<td>0.44</td>
<td>0.35</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. JOL2</td>
<td>0.27</td>
<td>0.65</td>
<td>0.45</td>
<td>0.69</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Need for Cognition</td>
<td>0.20</td>
<td>0.21</td>
<td>0.21</td>
<td>0.25</td>
<td>0.34</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. PAP</td>
<td>0.07</td>
<td>-0.01</td>
<td>0.02</td>
<td>0.18</td>
<td>0.09</td>
<td>-0.06</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. MAP</td>
<td>0.18</td>
<td>-0.03</td>
<td>0.24</td>
<td>0.18</td>
<td>0.16</td>
<td>0.36</td>
<td>0.23</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>9. PAV</td>
<td>0.01</td>
<td>-0.08</td>
<td>-0.09</td>
<td>-0.06</td>
<td>-0.16</td>
<td>-0.33</td>
<td>0.19</td>
<td>-0.10</td>
<td>1</td>
</tr>
</tbody>
</table>

A MANOVA showed no statistically significant overall group effects on these variables (excluding JOL2, which unlike other variables, was measured post intervention), Wilk’s \( \lambda = 0.86, p = 0.32 \). When asked to write down everything they knew about fracking in another question before the study process commenced, 80.8% of the participants exhibited no knowledge on the topic. Among the 19.2% participants who provided non-zero responses, their level of acquaintance with the topic was equivalent to 2.85 Idea Units on average, indicating very limited prior knowledge on the study topic in general. This minimized the effect of past experience on their performances of the main task.

An ANOVA detected a statistical difference across conditions on Judgment of Learning after the interventions, JOL2, \( F(2,122) = 3.00, p = 0.05 \). Tukey HSD multiple comparisons showed that the DM Group participants were more confident with their learning than the Control Group for Text 2, \( d = 0.56 \). A mixed ANOVA using the SPSS repeated measure analysis was then conducted, with Judgment of Learning (JOL) as the within-subject factor that had two levels: JOL1 and JOL2 collected at two different time points, and the group conditions as the between-subject factor with three levels: conditions DM, Argue, and Control. The assumption of sphericity was met since there
were two levels for the repeated measure variable JOL. The interventions had a statistically significant effect on participants’ Judgment of Learning overall, $F(1,122) = 4.43, p = .04$, with a fairly small effect size, partial $\eta^2 = .04$. There was also a statistically significant interaction between JOL and Group, $F(2,122) = 14.50, p < .001$, indicating that changes of participants’ confidence on learning varied across groups, with a medium effect size, partial $\eta^2 = .19$. Figure 5.1 shows the estimated marginal means of JOL1 and JOL2 for the three groups.

![Figure 5.1 Judgment of Learning Repeated Measures](image)

5.4. Coded Short-Answer Responses and Inter-Rater Reliabilities

As described in Chapter 4, short-answer responses were coded with a coding scheme that included number of Idea Units, number of Main Ideas, and number of Reasons, etc. To establish inter-rater reliability, a second coder who was blind to the treatment conditions was trained. Inter-rater reliabilities were established for the following variables, which were subject to discrepancies across coders: Idea Units, Main Ideas, and Reasons. Other measures such as Word Count and Argumentative Markers could be objectively counted, and were therefore not examined for reliability across raters.

Operational definitions of terms (see Chapter 4 for more details) were provided to the second rater with examples. Ten cases were randomly selected from Q1 for the
training purposes. The raters then met periodically to discuss the assignment of codes and resolve disagreements. The training process dealt extensively with interpretive nuances due to the overlapping of Main Ideas and Idea Units. Specific cases were discussed and agreement reached between the raters. The second rater then independently coded a subset of 30 participants for the summary question responses (10 participants randomly selected from each of the three groups) on number of Idea Units and number of Main Ideas. She also coded 40 answers randomly selected from questions R1 to R5 for the number of Reasons. Table 5.6 shows the descriptive statistics of agreements averaged across a sample of 30 participants (k stands for the total number of Idea Units or Main Ideas). All disagreements on coding instances were discussed and resolved prior to the independent coding.

Reliability for Idea Units was assessed by first calculating percent agreements for all 97 Idea Units for each participant, and then using the mean percentage across 30 participants as the reliability statistic. Agreement for Main Ideas was calculated in the same fashion.

Table 5.6  Descriptive Statistics of Rater Agreements

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>IU (k = 97)</td>
<td>97.1%</td>
<td>2.7%</td>
<td>90.7%</td>
<td>100.0%</td>
</tr>
<tr>
<td>MI (k = 16)</td>
<td>88.3%</td>
<td>8.6%</td>
<td>75.0%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Rather than calculating Kappa-like coefficients for each of the 97 Idea Units, I randomly selected 20 Idea Units and calculated 20 Cohen’s Kappa values (see Figure 5.2).

I also calculated 16 Cohen’s Kappa values for all the 16 Main Ideas. Table 5.7 contains the means and standard deviations of the Kappa statistics for 20 Idea Units randomly selected from 97 Idea Units and all 16 Main Ideas. Figure 5.3 maps the Kappa values for the 16 Idea Units.

Table 5.7  Cohen’s Kappa for 16 Main Ideas

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Idea Units (k = 20)</td>
<td>.89</td>
<td>.11</td>
<td>.69</td>
<td>1.00</td>
</tr>
<tr>
<td>Main Ideas (k = 16)</td>
<td>.73</td>
<td>.18</td>
<td>.49</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Cohen’s Kappa values indicated that inter-rater reliability for Idea Units and Main Ideas was moderate to high on average. The minimum value .49 was within the conventional moderate range (Landis & Koch, 1977).

**Figure 5.2**  Cohen’s Kappa for 20 Idea Units

**Figure 5.3**  Cohen’s Kappa for All 16 Main Ideas

These levels of inter-rater reliability suggested that the coding scheme was stable and robust. Since questions Q2 to Q6 were scored in the same manner using the same coding criteria on Idea Units and Main Ideas as the summary (Q1), it was reasonable to assume the same inter-rater reliability established for the summary question.
The second rater was also trained to score responses to the reasoning questions R1 to R5. She independently coded a subset of the reasoning responses randomly selected from all five reasoning questions \((k = 40)\), Cohen’s Kappa = .48, which is conventionally considered as a moderate level of rater agreement. All disagreements for the independent coding were then discussed and resolved. After establishing inter-rater reliability, I coded the remaining responses for all the variables (See Chapter 4 for a full list) for further analyses.

### 5.5. Free Recall

As mentioned in Chapter 4, participants were tested for their free recall of Text 1 before I implemented the interventions. Performance on this test was assessed by total Idea Units recalled (Free Recall Idea Units), total Main Ideas recalled (Free Recall Main Ideas), and total errors made (Free Recall Error).

Table 5.8 contains the means and standard deviations (in parentheses) of free recall scores on Idea Units, Main Ideas, and errors. Table 5.9 shows the correlations among the free recall variables.

#### Table 5.8  Descriptive Statistics for Free Recall Variables

<table>
<thead>
<tr>
<th></th>
<th>DM Group</th>
<th>Argue Group</th>
<th>Control Group</th>
<th>Pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free Recall Word Count</td>
<td>84.52 (49.34)</td>
<td>98.65 (56.19)</td>
<td>94.44 (52.78)</td>
<td>92.30 (52.65)</td>
</tr>
<tr>
<td>Free Recall Idea Units</td>
<td>3.23 (3.05)</td>
<td>4.26 (3.31)</td>
<td>3.99 (3.56)</td>
<td>3.81 (3.31)</td>
</tr>
<tr>
<td>Free Recall Main Ideas</td>
<td>.22 (.25)</td>
<td>.25 (.24)</td>
<td>.15 (.21)</td>
<td>.20 (.24)</td>
</tr>
<tr>
<td>Free Recall Error</td>
<td>2.00 (1.88)</td>
<td>1.38 (1.21)</td>
<td>1.02 (1.33)</td>
<td>1.48 (1.56)</td>
</tr>
</tbody>
</table>

#### Table 5.9  Correlations among Free Recall Variables

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Free Recall Word Count</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Free Recall Idea Units</td>
<td>.80</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Free Recall Main Ideas</td>
<td>.39</td>
<td>.28</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>4. Free Recall Error</td>
<td>.32</td>
<td>.04</td>
<td>.14</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 5.10  Test of Homogeneity of Variances for Free Recall Variables

<table>
<thead>
<tr>
<th></th>
<th>Levene Statistic</th>
<th>df1</th>
<th>df2</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free Recall Word Count</td>
<td>.96</td>
<td>2</td>
<td>122</td>
<td>.39</td>
</tr>
<tr>
<td>Free Recall Idea Units</td>
<td>2.13</td>
<td>2</td>
<td>122</td>
<td>.12</td>
</tr>
<tr>
<td>Free Recall Main Ideas</td>
<td>1.07</td>
<td>2</td>
<td>122</td>
<td>.35</td>
</tr>
<tr>
<td>Free Recall Error</td>
<td>4.54</td>
<td>2</td>
<td>122</td>
<td>.01</td>
</tr>
</tbody>
</table>

According to Table 5.10, heterogeneity of variance was not detected for Free Recall Word Count, Free Recall Idea Units, and Free Recall Main Ideas. A MANOVA test showed no overall group differences in these variables, Wilk’s $\lambda = .94$, $p = .31$.

The homogeneity of variance assumption for the total number of errors made was violated, $Levene(2,122) = 4.54$, $p = .01$. Several attempts at transforming the variable did not produce homogeneity. According to an independent-samples Kruskal-Wallis test, errors were significantly different across groups, $H(2) = 7.59$, $p = .02$. Pairwise comparisons with adjusted $p$-values showed that the DM Group made statistically significantly more errors when recalling the study material than the Control Group, $p = .02$, $r = .30$.

5.6. The DM Uses

I examined how participants used the DM by reviewing every Dialectical Map and counting the number of claims, evidence statements, and warrants. The means, standard deviations, and minimum and maximum values of these variables are presented in Table 5.11. Every claim, evidence, and warrant thread was a deliberate attempt of representing the text in an argumentative form (see Figure 5.4 and 5.5 as an example). This participant rated the strength of every claim, supported all claims with evidence, and provided relevant warrants. Arguments and counterarguments were clearly stated; statistics were accurately cited in the evidence boxes; claims that directly opposed each other were linked to demonstrate an effort of counterargument and refutation. Moreover, an integrated conclusion was reached to complete the Dialectical Map.
The DM shown in Figure 5.4 has a rather short (33 words) conclusion compared to the mean word count (53.07 words) for all DM conclusions. Nevertheless, it contained the evaluation of both sides of the argument instead of making a one-sided argument. A more typical conclusion obtained from this experiment usually included a brief explanation of the evidence, as well as the weighing and comparisons that derived the conclusion. For example:

The number of advantages associated with fracking is nearly equal to the number of disadvantages. Fracking has been known to negatively affect societies and the environment in terms of releasing methane (a greenhouse gas), radioactive elements, and other health-affecting substances into the air as well as both contaminating aquifers and depleting water sources in the process of extracting the shale gas. On the other hand, fracking is estimated to be highly beneficial in terms of capital; it may be both inexpensive to obtain while simultaneously lowering income taxes and creating jobs, and making economies more independent. It is possible that it may reduce the carbon footprint, although it is debateable [sic] whether it is productive to reduce carbon outputs but increase methane outputs. Canada’s decision must take into account both sides, but in the end the question really asks what Canada cares about more: the economy or environmental and social wellbeing.

The majority of participants used evidence correctly, reflecting a good understanding of its function in an argument. Although researchers criticized the complexity, abstractness, and non-universality of a warrant in the Toulmin model, participants in this study appeared to have a good understanding of its purpose and used it relatively accurately. For example, in Figure 5.5, the statement “Chemical added and natural salts, metals, radioactive elements could leak into the environment” was used as a warrant to “Storing, transporting, treating, disposing of fracking wastewater is dangerous,” which was used as supporting evidence to claim “Environmental harm.”
Canada should engage in large-scale fracking because of its substantial economic benefits, but environmental and health hazards should be considered and regulations preventing leakage should be enforced in order to reduce this harm.

Figure 5.4   A Completed DM
Table 5.11 contains the descriptive statistics of participants’ DM activities. All participants but one reached a conclusion (i.e., an answer to the question on top of the DM) after comparing and integrating pro reasons and con reasons with an average of 53.75 words. Participants were also using DM functions in addition to creating the basic elements of an argument (see Table 5.12). For instance, more than half of the participants were linking and weighing their arguments. Table 5.13 gives the correlations among DM variables. Evidence was correlated with reasons and warrants, \( r = .56 \), and \( .42 \) respectively.

**Table 5.11  Descriptive Statistics of the DM Activities**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reasons</td>
<td>9.04</td>
<td>3.57</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>Evidence</td>
<td>11.62</td>
<td>4.38</td>
<td>3</td>
<td>22</td>
</tr>
<tr>
<td>Warrants</td>
<td>4.77</td>
<td>3.85</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>Conclusion Word Count</td>
<td>53.75</td>
<td>52.36</td>
<td>0</td>
<td>258</td>
</tr>
</tbody>
</table>
### Table 5.12 Other DM Usages

<table>
<thead>
<tr>
<th>DM functions</th>
<th>Counts</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linking</td>
<td>26</td>
<td>59.1 %</td>
</tr>
<tr>
<td>Weights</td>
<td>28</td>
<td>63.6 %</td>
</tr>
<tr>
<td>Conclusions</td>
<td>43</td>
<td>97.7 %</td>
</tr>
</tbody>
</table>

### Table 5.13 Correlations Among DM Variables

<table>
<thead>
<tr>
<th></th>
<th>Reasons</th>
<th>Evidence</th>
<th>Warrants</th>
<th>Link</th>
<th>Conclusion</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reasons</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evidence</td>
<td>.56</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Warrants</td>
<td>-.07</td>
<td>.42</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Link</td>
<td>.07</td>
<td>.05</td>
<td>-.10</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conclusion</td>
<td>.02</td>
<td>-.02</td>
<td>.06</td>
<td>.22</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Weight</td>
<td>.20</td>
<td>.29</td>
<td>.14</td>
<td>.01</td>
<td>.17</td>
<td>1</td>
</tr>
</tbody>
</table>

### 5.7. Highlighting and Time-on-Task (ToT)

The nStudy log server recorded participant activities including highlighting in both Text 1 and Text 2, as well as time spent studying the texts. Highlighted contents were segmented into Idea Units and the total number of Idea Units highlighted was calculated for each participant.

#### 5.7.1. Highlighting

Participants were given the option to highlight for study tasks with both Text 1 and Text 2. Table 5.14 shows the means and standard deviations of numbers of Idea Units highlighted for Text 1 (HL1) and Text 2 (HL2).

### Table 5.14 Descriptive Statistics for Highlighting

<table>
<thead>
<tr>
<th></th>
<th>DM Group</th>
<th>Argue Group</th>
<th>Control Group</th>
<th>Pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td>HL1</td>
<td>3.77 (4.12)</td>
<td>3.63 (3.97)</td>
<td>2.22 (2.77)</td>
<td>3.22 (3.72)</td>
</tr>
<tr>
<td>HL2</td>
<td>2.43 (6.36)</td>
<td>23.38 (21.12)</td>
<td>10.8 (12.97)</td>
<td>11.88 (16.83)</td>
</tr>
</tbody>
</table>

109
Both HL1 and HL2 failed the Levene test for homogeneity of variance, \( Levene(2, 122) = 4.10, p = .02, \) and \( Levene(2, 122) = 28.51, p < .001, \) indicating the violation of the homogeneity of variance assumption. An independent-samples Kruskal-Wallis test indicated that there was no group differences on the highlighting count with Text 1, \( H(2) = 3.05, p = .22. \) There were, however, statistically significant group differences on HL2, \( H(2) = 33.95, p < .001. \) Pairwise comparisons showed that the DM Group highlighted significantly fewer Idea Units than both the Argue Group, adjusted \( p < .001, r = .64, \) and the Control Group, adjusted \( p = .008, r = .32. \) The Argue Group also highlighted more Idea Units than the Control Group, adjusted \( p = .015, r = .31. \)

5.7.2. Time-on-Task

Time-on-Tasks (ToTs) reflected how participants interacted with the studying materials. The time durations (seconds) for each participant studying Text 1 (ToT1) and Text 2 (ToT2) were recorded by the nStudy server (See Table 5.15 for descriptive statistics).

<table>
<thead>
<tr>
<th>Table 5.15</th>
<th>Descriptive Statistics for ToT1 and ToT2 (Seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DM Group</td>
</tr>
<tr>
<td>TOT1</td>
<td>274.20 (125.96)</td>
</tr>
<tr>
<td>TOT2</td>
<td>2284.93 (659.00)</td>
</tr>
</tbody>
</table>

Both variables passed the Levene test for homogeneity of variance, \( Levene(2,105) = .08, p = .92, \) and \( Levene(2,105) = 1.18, p = .31. \) An analysis of variance showed no statistically detectable differences across groups on time spent studying Text 1, \( F(2,105) = .57, p = .57. \) The DM Group spent more time studying Text 2, \( F(2,105) = 38.05, p < .001. \) Tukey HSD multiple comparisons showed that participants in the DM Group spent more time studying Text 2 than the Argue Group and the Control Group, \( d = 1.62 \) and \( 1.71 \) respectively.
5.8. Critical Thinking Scores

Results for the critical thinking test is reported in this section. Scores for all 6 critical thinking items were totalled, generating a sum score (Critical Thinking) for each participant. Table 5.16 gives the means and standard deviations of this sum score for the three conditions.

Table 5.16 Descriptive Statistics for Critical Thinking

<table>
<thead>
<tr>
<th></th>
<th>DM Group</th>
<th>Argue Group</th>
<th>Control Group</th>
<th>Pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical Thinking</td>
<td>3.05 (1.41)</td>
<td>3.26 (1.23)</td>
<td>2.90 (1.13)</td>
<td>3.06 (1.27)</td>
</tr>
</tbody>
</table>

Heterogeneity of variance was not detected for Critical Thinking, \( Levene(2,122) = 2.46, p = .09 \). An ANOVA showed no statistically detectable differences among the means of three groups on the critical thinking test, \( F(2,122) = .83, p = .44 \). It was informative to consider descriptive statistics and group differences for the sum score of the critical thinking test. However, the six items had low internal consistency, Cronbach’s \( \alpha = .15 \). Deleting any item only slightly improved their reliability (Cronbach’s \( \alpha = .18 \)). Therefore, no further analysis was conducted using these items as a scale.

5.9. Summary (Q1) Scores

As described in Chapter 4, each summary was scored for five types of data: its rhetorical organization (Organization), number of Argument Markers, use of Neutral Idea Units, coverage, and number of errors. Coverage incorporated measures on the total number of Idea Units and total number of Main Ideas, which were obtained by coding the summaries with 97 Idea Units and 16 Main Ideas. Table 5.17 includes the means and standard deviations (in parentheses) of participants’ scores on their summary performance and the word count of the summaries (Q1 Word Count). See Figure 5.6 for the means and standard errors of Q1 Word Count.
### Table 5.17 Descriptive Statistics for Q1

<table>
<thead>
<tr>
<th></th>
<th>DM Group</th>
<th>Argue Group</th>
<th>Control Group</th>
<th>Pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 Word Count</td>
<td>171.25 (145.30)</td>
<td>217.25 (165.53)</td>
<td>232.85 (165.89)</td>
<td>206.18 (159.76)</td>
</tr>
<tr>
<td>Organization</td>
<td>1.84 (.81)</td>
<td>1.30 (.79)</td>
<td>1.22 (.76)</td>
<td>1.46 (.83)</td>
</tr>
<tr>
<td>Argument Marker</td>
<td>5.25 (4.49)</td>
<td>3.65 (3.61)</td>
<td>4.85 (3.6)</td>
<td>4.61 (3.97)</td>
</tr>
<tr>
<td>Pro Idea Units</td>
<td>1.87 (2.75)</td>
<td>5.25 (4.49)</td>
<td>3.65 (3.61)</td>
<td>4.85 (3.6)</td>
</tr>
<tr>
<td>Con Idea Units</td>
<td>2.81 (2.74)</td>
<td>1.96 (2.20)</td>
<td>2.85 (3.03)</td>
<td>2.55 (2.69)</td>
</tr>
<tr>
<td>Neutral Idea Units</td>
<td>1.17 (2.39)</td>
<td>2.80 (3.77)</td>
<td>3.83 (3.88)</td>
<td>2.57 (3.54)</td>
</tr>
<tr>
<td>Total Idea Units</td>
<td>5.85 (6.54)</td>
<td>7.24 (7.96)</td>
<td>9.06 (8.54)</td>
<td>7.35 (7.75)</td>
</tr>
<tr>
<td>Main Ideas</td>
<td>3.39 (2.22)</td>
<td>2.1 (.22)</td>
<td>2.77 (1.7)</td>
<td>2.17 (2.13)</td>
</tr>
<tr>
<td>Q1 Error</td>
<td>.52 (1.17)</td>
<td>1.45 (2.23)</td>
<td>1.85 (2.62)</td>
<td>1.26 (2.14)</td>
</tr>
</tbody>
</table>

**Figure 5.6** Word Count for Q1
According to Table 5.18, the following variables passed the Levene test for homogeneity of variance: Q1 Word Count, Organization, Argument Marker, Pro Idea Units, Con Idea units, and Total Idea Units. A MANOVA test showed an overall group difference on these variables, Wilk’s $\lambda = .64$, $p < .001$, partial $\eta^2 = .20$. Subsequent univariate analysis showed no statistically detectable differences across the three conditions on the length of the summaries participants generated measured by word counts and most of the other variables for Q1. An effect for Organization was detected, $F(2,122) = 7.92$, $p = .001$. Tukey HSD multiple comparisons showed that the DM Group participants were more likely to organize their summaries in an introduction-body-conclusion structure than both the Argue Group, $d = .67$, and the Control Group, $d = .79$. For variables that did not meet the assumption of homogeneity of variances, an independent-samples Kruskal-Wallis test was conducted. Table 5.19 presents the main effects and multiple comparisons with adjusted pairwise $p$-values. The DM Group (A) statistically outperformed the Argue Group (B) on the number of Main Ideas used in their summaries ($r = .70$). The Control Group (C) also used more Main Ideas than the Argue Group ($r = .65$). The DM Group used statistically significantly fewer Neutral Idea Units when composing a summary than the Control Group ($r = .45$). The DM Group also made statistically significantly fewer errors when summarizing the study material than the Control Group ($r = .35$).
Recall that the DM Group participants made significantly more errors than the Control Group in their free recall test (see section 5.5, adjusted $p = .02$, $r = .30$). After the interventions, the DM Group participants remarkably reduced their number of errors compared to both the Argue Group and the Control Group. A mixed ANOVA using the SPSS repeated measure analysis was conducted, with Error as the within-subject factor that had two levels: Free Recall Error and Q1 Error collected at two different time points, and the group conditions as the between-subject factor with 3 levels: conditions DM, Argue, and Control. The assumption of sphericity was met. The interventions had a statistically significant effect on number of errors made overall (Error), Wilk’s $\lambda = .73$, $F(1,122) = 44.06$, $p < .001$, with a medium effect size, partial $\eta^2 = .27$. There was also a significant interaction between Error and Group, Wilk’s $\lambda = .85$, $F(2,122) = 11.05$, $p < .001$, indicating the changes of participants’ accuracy in recalling information varied across groups, with a small to moderate effect size, partial $\eta^2 = .15$. Figure 5.5 shows the estimated marginal means of Error for the three groups.

![Figure 5.7 Error Repeated Measures](image)
Table 5.20 contains the correlations among Q1 variables. The majority of the correlations are moderate to large in size.

Table 5.20  Correlation Matrix for Q1 Variables

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Q1 Word Count</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Organization</td>
<td>.23</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Argument Marker</td>
<td>.79</td>
<td>.37</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Pro Idea Units</td>
<td>.77</td>
<td>.07</td>
<td>.56</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Con Idea Units</td>
<td>.68</td>
<td>.25</td>
<td>.67</td>
<td>.53</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Neutral Idea Units</td>
<td>.78</td>
<td>-.02</td>
<td>.49</td>
<td>.66</td>
<td>.54</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Total Idea Units</td>
<td>.88</td>
<td>.10</td>
<td>.66</td>
<td>.86</td>
<td>.79</td>
<td>.89</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Main Ideas</td>
<td>.39</td>
<td>.23</td>
<td>.51</td>
<td>.28</td>
<td>.55</td>
<td>.26</td>
<td>.42</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>9. Q1 Error</td>
<td>.58</td>
<td>.03</td>
<td>.40</td>
<td>.47</td>
<td>.29</td>
<td>.42</td>
<td>.47</td>
<td>.09</td>
<td>1</td>
</tr>
</tbody>
</table>
5.10. **Cued-Recall Scores**

The remaining short-answer tasks were categorized as cued-recall questions (Q2 to Q6) and reasoning questions (R1 to R5) based on the types of information the questions were designed to obtain. As described in Chapter 4, cued-recall responses were scored using 4 criteria: a) number of Idea Units used, b) number of Idea Units used correctly according to the scoring rubric for each question, c) number of Main Ideas used, and d) number of Main Ideas used correctly. Due to the similarity in the nature of questions Q2 to Q6, it would be cumbersome and repetitive to report statistics for each question. Therefore, I summed all the four types of scores for these questions. Since this approach was not aggregating variables to create a scale that measures a latent construct, but rather a sum of counts of limited elements of an answer, no computation of Cronbach’s α was necessary to examine the internal consistency of the items. Table 5.21 gives the descriptive statistics of the total word counts (Qs Word Count) and the 4 types of scores for the sum of Q2 to Q6: Cued-Recall Total Idea Units, Cued-Recall Correct Idea Units, Cued-Recall Total Main Ideas, and Cued-Recall Correct Main Ideas.

<table>
<thead>
<tr>
<th></th>
<th>DM Group</th>
<th>Argue Group</th>
<th>Control Group</th>
<th>Pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qs Word Count</td>
<td>117.82 (86.20)</td>
<td>186.53 (124.62)</td>
<td>164.90 (99.74)</td>
<td>155.25 (107.30)</td>
</tr>
<tr>
<td>Cued-Recall Total Idea Units</td>
<td>3.80 (3.14)</td>
<td>4.87 (3.89)</td>
<td>4.19 (3.15)</td>
<td>4.27 (3.40)</td>
</tr>
<tr>
<td>Cued-Recall Correct Idea Units</td>
<td>3.38 (3.06)</td>
<td>4.39 (3.44)</td>
<td>3.60 (2.99)</td>
<td>3.78 (3.17)</td>
</tr>
<tr>
<td>Cued-Recall Total Main Ideas</td>
<td>2.27 (1.13)</td>
<td>1.88 (1.11)</td>
<td>1.73 (.80)</td>
<td>1.97 (1.05)</td>
</tr>
<tr>
<td>Cued-Recall Correct Main Ideas</td>
<td>2.17 (1.09)</td>
<td>1.67 (1.05)</td>
<td>1.46 (.77)</td>
<td>1.78 (1.02)</td>
</tr>
</tbody>
</table>

Table 5.22 contains the Levene test results. Heterogeneity of variance was detected for Qs Word Count, Levene(2,122) = 3.51, p = .03. Attempts at transforming this variable did not produce homogeneity. An independent-samples Kruskal-Wallis Test showed statistically significant differences across three conditions, H(2) = 9.11, p = .01. Pairwise comparisons with an adjusted p-value showed that the Argue Group wrote more words than the DM Group, adjusted p = .01, r = .31.
Table 5.22  Test of Homogeneity of Variances

<table>
<thead>
<tr>
<th></th>
<th>Levene Statistic</th>
<th>df1</th>
<th>df2</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qs Word Count</td>
<td>3.51</td>
<td>2</td>
<td>122</td>
<td>.03</td>
</tr>
<tr>
<td>Cued-Recall Total Idea Units</td>
<td>1.44</td>
<td>2</td>
<td>122</td>
<td>.24</td>
</tr>
<tr>
<td>Cued-Recall Correct Idea Units</td>
<td>.94</td>
<td>2</td>
<td>122</td>
<td>.39</td>
</tr>
<tr>
<td>Cued-Recall Total Main Ideas</td>
<td>2.91</td>
<td>2</td>
<td>122</td>
<td>.06</td>
</tr>
<tr>
<td>Cued-Recall Correct Main Ideas</td>
<td>2.81</td>
<td>2</td>
<td>122</td>
<td>.06</td>
</tr>
</tbody>
</table>

Homogeneity of variances were obtained for the remaining four scores. A MANOVA test showed an overall group effect on these four variables, Wilk’s $\lambda = .85$, $p = .01$, partial $\eta^2 = .08$. Subsequent univariate analyses (tests of between-subjects effects) revealed statistically significant group effects for the total number of Main Ideas used, $F(2,122) = 3.23$, $p = .04$, $\eta^2 = .05$, and the number of Main Ideas used correctly, $F(2,122) = 5.92$, $p = .004$, $\eta^2 = .09$. Tukey HSD multiple comparisons showed that the DM Group outperformed the Control Group on total Main Ideas used, $d = .55$, and number of Main Ideas used correctly, $d = .75$.

![Figure 5.8  Word Counts for Q2–Q6](image)

$^4$ Error Bars: Standard Error.
Table 5.23 gives the correlations among the cued-recall question variables. Most of the correlations are moderate to large in size.

**Table 5.23 Correlation Matrix for Cued-Recall Question Scores**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Qs Word Count</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Cued-Recall Total Idea Units</td>
<td>.83</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Cued-Recall Correct Idea Units</td>
<td>.80</td>
<td>.97</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Cued-Recall Total Main Ideas</td>
<td>.36</td>
<td>.22</td>
<td>.23</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>5. Cued-Recall Correct Main Ideas</td>
<td>.28</td>
<td>.15</td>
<td>.18</td>
<td>.95</td>
<td>1</td>
</tr>
</tbody>
</table>

### 5.11. Reasoning Scores

Questions R1 to R5 were designed to test participants' reasoning activities. Answers were scored on number of Reasons used to support a claim. As mentioned in Chapter 2, considering both arguments and counterarguments is deemed good reasoning (Santos & Santos, 1999). However, responses to the reasoning questions in this study tended to be fairly short (46.27 words per question on average, approximately three sentences). Since establishing a plot that contains both arguments and counterarguments needs a certain level of elaboration, which was lacking here in the participant responses, only the number of Reasons was used to score these questions. The Reasons included any kind of supporting evidence to a particular claim. I summed the numbers of Reasons across questions R1 to R5. Consistent with aggregating the scores across all cued-recall questions in a previous section, no Cronbach’s α was calculated for totalling the numbers of Reasons across all the reasoning questions. Table 5.24 demonstrates the descriptive statistics of the sum of R1 to R5, including all the Reasons (Total Reasons) as well as the total word count across the five responses (Rs Word Count).

**Table 5.24 Descriptive Statistics for Reasoning Questions**

<table>
<thead>
<tr>
<th></th>
<th>DM Group</th>
<th>Argue Group</th>
<th>Control Group</th>
<th>Pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rs Word Count</td>
<td>206.30 (119.24)</td>
<td>243.42 (127.75)</td>
<td>246.49 (134.57)</td>
<td>231.36 (127.48)</td>
</tr>
<tr>
<td>Total Reasons</td>
<td>10.33 (3.63)</td>
<td>9.48 (3.31)</td>
<td>8.83 (3.58)</td>
<td>9.56 (3.54)</td>
</tr>
</tbody>
</table>
The homogeneity of variance assumption was met for total word counts (Rs Word Count), $Levene(2,122) = .92, p = .40$, and the sum of Reasons (Total Reasons), $Levene(2,122) = .05, p = .95$. A MANOVA identified a significant overall difference on these variables, Wilk’s $\lambda = .83, p < .001$, $\eta^2 = .09$. However, subsequent univariate analyses showed no group effects on total words participants wrote for the answers, $F(2,122) = 1.33, p = .27$, or the total number of Reasons used, $F(2,122) = 1.96, p = .15$.

![Figure 5.9 Word Counts for R1–R5](image)

5.12. Correlations

Table 5.25 contains the correlation matrix of the variables investigated in this experiment. There are both positive and negative correlations among self-reported information, free recall information, and short-answer responses. Size of correlations ranges from $r = -.04$ to $r = .69$. 
Table 5.25  Correlation Matrix of Self-Reported Information, Free Recall Idea Units, and Short-Answer Responses

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Years in English</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. JOL1</td>
<td>.02</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. NfC</td>
<td>-.04</td>
<td>.25</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Free Recall Idea Units</td>
<td>.24</td>
<td>.41</td>
<td>.15</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Organization</td>
<td>.03</td>
<td>.07</td>
<td>.15</td>
<td>.14</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Q1 Main Ideas</td>
<td>.05</td>
<td>.02</td>
<td>-.03</td>
<td>.25</td>
<td>.23</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Argument Marker</td>
<td>.17</td>
<td>.16</td>
<td>-.04</td>
<td>.27</td>
<td>.37</td>
<td>.51</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Cued-Recall Idea Units</td>
<td>.17</td>
<td>.30</td>
<td>.15</td>
<td>.53</td>
<td>.11</td>
<td>.04</td>
<td>.17</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Cued-Recall Main Ideas</td>
<td>.12</td>
<td>.17</td>
<td>.21</td>
<td>.24</td>
<td>.16</td>
<td>.32</td>
<td>.20</td>
<td>.22</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Total Reasons</td>
<td>.25</td>
<td>.11</td>
<td>.21</td>
<td>.31</td>
<td>.16</td>
<td>.29</td>
<td>.34</td>
<td>.32</td>
<td>.29</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>11. JOL2</td>
<td>-.03</td>
<td>.69</td>
<td>.34</td>
<td>.40</td>
<td>.25</td>
<td>.18</td>
<td>.16</td>
<td>.36</td>
<td>.29</td>
<td>.23</td>
<td>1</td>
</tr>
</tbody>
</table>
5.13. Multiple Regression Analyses

According to Table 5.25, participant performance on the summary, the cued-recall, and the reasoning questions had weak to moderate correlations with the following factors: Years in English, Judgment of Learning (JOL1), Need for Cognition, and Free Recall Idea Units. Regression analyses were conducted to further investigate the extent to which the DM treatment affected the outcome variables after controlling for these factors. Theoretically, English proficiency is a feasible predictor for some learning outcomes. For example, low English proficiency for students learning English as a second language has been found to account for one-sided arguments in their writing (Rusfandi, 2015). Additionally, a large proportion of participants in the current research were non-native English speakers. It is useful to investigate whether their English proficiency is a predictor of their performance, such as reasoning activities. Need for Cognition (NfC) is another plausible predictor for performance, such as the quality of writing. Additionally, research has shown that Need for Cognition is positively correlated with argumentativeness (Kardash & Scholes, 1996; Nussbaum, 2005). I am also interested in finding out whether pre-treatment Judgment of Learning (JOL1) is a predictor of performance, such as recall of information as well as post-treatment Judgment of Learning (JOL2). Performance on the pre-test free recall may also forecast how well subjects would answer the post-test questions.

Sequential regressions enable step-by-step examinations of relationships between the dependent variable and the predictor variables. Predictor variables are entered into the equation in a sequence to determine whether or not introducing an additional predictor would enhance the prediction of the dependent variable (Tabachnick & Fidell, 2006). Tables 5.27 to 5.33 contain the sequential multiple regressions that used the following predictor variables 1) Years in English, 2) Judgment of Learning (JOL1), 3) Need for Cognition (NfC), and 4) Free Recall Idea Units. The categorical variable “Group,” indicating the three treatment conditions (i.e., the DM, Argue, and Control), was recoded into two dummy variables for the regression analyses. Dummy variables C1 and C2 were generated using contrast coding (Cohen, Cohen, West, & Aiken, 2003). These dummy variables fulfilled two purposes: 1) contrasting the DM Group to the Argue Group for their performances on the summary, cued-recall, and reasoning scores, and 2)
contrasting the Control Condition to the DM and Argue Groups combined for their performances on the same outcome variables. These goals align well with my research questions investigating the effects the DM had on the dependent variables. Based on these goals, I coded C1 and C2 with the weights shown in Table 5.26. In C1, the DM was coded +1, the Argue Group was coded -1, and Control Group was coded 0. In C2, the DM Group and Argue Group were coded as +1 and the Control Group was coded -2. This coding technique generated two orthogonal variables. JOL2 was also examined using the same sets of predictor variables.

**Table 5.26  Contrast Coding Weights**

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>DM Group</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Argue Group</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>Control Group</td>
<td>0</td>
<td>-2</td>
</tr>
</tbody>
</table>

As shown in Table 5.27 to 5.33, outcome variables were analyzed using sequential multiple regression. At Step 1, Years in English, JOL1, Need for Cognition (NcF), and Free Recall Idea Units were entered in the equation; at Step 2, C1 and C2 were added to the model to examine their effect after statistically controlling for Years in English, JOL1, NcF, and Free Recall Idea Units.

For Organization scores (Table 5.27), effects from Years in English, JOL1, Need for Cognition, and Free Recall Idea Units were not statistically detectable predictors ($p = .31$). When C1 and C2 were entered to the regression equation, 12% more of the variance for the dependent variable was explained by the model ($p < .001$). Standardized regression coefficients showed that both C1 and C2 were significant predictors of organization scores ($\beta = .30, p < .001$ and $\beta = .19, p = .03$ respectively), indicating that the DM Group outperformed both the Argue Group and the Control Group on this measure. The adjusted $R^2$ for the final model including all the predictors was .12.

In Table 5.28, the regression model with Years in English, JOL1, Need for Cognition, and Free Recall Idea Units were statistically significant predictors of Argument Marker scores ($p = .02$), explaining 7% of the variance. With C1 and C2 added to the model, 5% additional variance was explained ($p = .03$). The adjusted $R^2$ for the final model including all the predictors was .11. Standardized regression coefficients showed that C1 was a significant predictor of participants’ use of Argument Markers ($\beta =$
.23, \( p = .01 \)), indicating that the DM Group outperformed the Argue Group on this measure. C2 was not a statistically detectable predictor of Argument Marker scores (\( p = .92 \)). This created uncertainty in whether the DM Group outperformed the Control Group on Argument Marker. As noted by Cohen et al. (2003) a plausible dummy-variable-like coding system requires \( g – 1 \) dummy variables, where \( g \) represents the number of levels in the categorical variable being recoded. In this case \( g \) equals 3 because the Group variable has three levels: the DM Group, the Argue Group, and the Control Group. Inherently, the contrast coding described previously did not allow direct comparisons between the DM Group and the Control Group, and a different coding system is not able to achieve all three pairwise comparisons. Therefore, in order to test for the group differences between the DM Group and the Control Group with the same predictor variables in the first block in Table 5.28, a post hoc regression was conducted using SPSS filter to exclude cases from the Argue Group. A new dummy variable C3 was created, where the DM cases were coded 1 and the Control cases were coded 0. The follow-up regression showed that C3 was not a significant predictor of Argument Marker when the predictor variables in the first block in Table 5.28 were statistically controlled (\( p = .20 \)), indicating the DM Group did not outperform the Control Group on this measure.

In Table 5.29, at Step 1 with variables Years in English, JOL1, Need for Cognition, and Free Recall Idea Units in the equation, the regression model predicting Q1 Main Ideas scores was statistically detectable (\( p = .05 \)), explaining 5% of the variance. At Step 2, with C1 and C2 added to the equation, the model accounted for an additional 4% of the variance (\( p < .001 \)). Standardized regression coefficients showed that both C1 and C2 were significant predictors of Q1 Main Ideas scores (\( \beta = .66, p < .001 \) and \( \beta = -.20, p < .001 \) respectively), indicating that the DM Group outperformed both the Argue Group and the Control Group on Q1 Main Ideas. The adjusted \( R^2 \) for the final model including all the predictors was .50.

In Table 5.30, at Step 1 with variables Years in English, JOL1, Need for Cognition, and Free Recall Idea Units in the equation, the regression model predicting Cued-Recall Idea Units scores was statistically detectable, adjusted \( R^2 = .29, p < .001 \). At Step 2, adding C1 and C2 to the equation did not produce a statistically detectable change in \( R^2 \) (\( p = .79 \)). C1 and C2 were not significant predictors of participants'
performance on Cued-Recall Idea Units. In other words, the DM treatment did not have an effect on this measure.

In Table 5.31, at Step 1 with variables Years in English, JOL1, Need for Cognition, and Free Recall Idea Units in the equation, the regression model predicting Cued-Recall Main Idea scores was statistically detectable ($p = .02$), explaining 7% of the variance. At Step 2, when C1 and C2 were added to the model, an additional 7% of the variance was explained ($p = .01$). Standardized regression coefficients showed that C1 was a significant predictor of Cued-Recall Idea Units scores ($\beta = .22, p = .01$), indicating that the DM Group outperformed the Argue Group on this measure. The adjusted $R^2$ for the final model including all the predictors was .12. As discussed earlier, because C1 was a significant predictor while C2 was not ($p = .09$), a post hoc regression excluding the Argue Group cases was needed. The goal was to compare the DM and Control groups directly without the influence from the Argue Group. The same dummy variable C3 was used for the follow-up regression. C3 was a significant predictor of Cued-Recall Main Ideas ($\beta = .31, p = .004$), indicating that the DM Group outperformed the Control Group on this outcome variable.

In Table 5.32, at Step 1 with variables Years in English, JOL1, Need for Cognition, and Free Recall Idea Units in the equation, the regression model predicting Total Reasons scores was statistically detectable, explaining 14% of the variance ($p < .001$). At Step 2, with C1 and C2 added to the equation, the model accounted for an additional 5% of the variance ($p = .04$). However, C1 and C2 were not significant predictors of Total Reasons ($p = .06$ and .08 respectively).

In Table 5.33, at Step 1 with variables Years in English, JOL1, Need for Cognition, and Free Recall Idea Units in the equation, the regression model predicting JOL2 scores was statistically detectable, explaining 51% of the variance ($p < .001$). At Step 2, with C1 and C2 added to the equation, an additional 8% of the variance was explained by the model. Standardized regression coefficients showed that both C1 and C2 were significant predictors of JOL2 scores ($\beta = .26, p < .001$ and $\beta = .14, p = .02$ respectively), indicating that participants in the DM Group were more confident with their learning than those in both the Argue Group and the Control Group. The adjusted $R^2$ for the final model including all the predictors was .59.
Table 5.27  Sequential Regression Table for Predicted Organization Scores

<table>
<thead>
<tr>
<th>Variable</th>
<th>$R^2$</th>
<th>Adjusted $R^2$</th>
<th>$\Delta R^2$</th>
<th>Sig $\Delta F$</th>
<th>$B$</th>
<th>SEB</th>
<th>$\beta$</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years in English</td>
<td>.04</td>
<td>.01</td>
<td>.04</td>
<td>.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JOL1</td>
<td>.00</td>
<td>.01</td>
<td>-.02</td>
<td>- .22</td>
<td>.83</td>
<td></td>
<td>.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>NfC</td>
<td>.01</td>
<td>.01</td>
<td>.14</td>
<td>1.49</td>
<td>.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Free Recall Idea Units</td>
<td>.03</td>
<td>.03</td>
<td>.13</td>
<td>1.3</td>
<td>.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Step 2</strong></td>
<td>.16</td>
<td>.12</td>
<td>.12</td>
<td>&lt; .001*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years in English</td>
<td>.01</td>
<td>.01</td>
<td>.05</td>
<td>.58</td>
<td>.56</td>
<td></td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JOL1</td>
<td>.00</td>
<td>.00</td>
<td>.03</td>
<td>.34</td>
<td>.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NfC</td>
<td>.01</td>
<td>.01</td>
<td>.09</td>
<td>1.04</td>
<td>.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Free Recall Idea Units</td>
<td>.04</td>
<td>.02</td>
<td>.15</td>
<td>1.58</td>
<td>.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>.30</td>
<td>.09</td>
<td>.30</td>
<td>3.5</td>
<td>&lt; .001*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>.11</td>
<td>.05</td>
<td>.19</td>
<td>2.22</td>
<td>.03*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 5.28  Sequential Regression Table for Predicted Argument Marker Scores

<table>
<thead>
<tr>
<th>Variable</th>
<th>$R^2$</th>
<th>Adjusted $R^2$</th>
<th>$\Delta R^2$</th>
<th>Sig $\Delta F$</th>
<th>$B$</th>
<th>SEB</th>
<th>$\beta$</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years in English</td>
<td>.10</td>
<td>.07</td>
<td>.10</td>
<td>.02*</td>
<td>.08</td>
<td>.06</td>
<td>.11</td>
<td>1.25</td>
<td>.21</td>
</tr>
<tr>
<td>JOL1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.02</td>
<td>.02</td>
<td>.09</td>
<td>.94</td>
<td>.35</td>
</tr>
<tr>
<td>NfC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.04</td>
<td>.04</td>
<td>-.09</td>
<td>-1.03</td>
<td>.30</td>
</tr>
<tr>
<td>Free Recall Idea Units</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.27</td>
<td>.12</td>
<td>.22</td>
<td>2.24</td>
<td>.03*</td>
</tr>
<tr>
<td>Step 2</td>
<td>.15</td>
<td>.11</td>
<td>.05</td>
<td>.03*</td>
<td>.10</td>
<td>.06</td>
<td>.14</td>
<td>1.55</td>
<td>.12</td>
</tr>
<tr>
<td>Years in English</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.03</td>
<td>.02</td>
<td>.13</td>
<td>1.35</td>
<td>.18</td>
</tr>
<tr>
<td>JOL1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.04</td>
<td>.04</td>
<td>-.10</td>
<td>-1.08</td>
<td>.28</td>
</tr>
<tr>
<td>NfC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.28</td>
<td>.12</td>
<td>.23</td>
<td>2.38</td>
<td>.02*</td>
</tr>
<tr>
<td>Free Recall Idea Units</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.13</td>
<td>.42</td>
<td>.23</td>
<td>2.68</td>
<td>.01*</td>
</tr>
<tr>
<td>C1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.03</td>
<td>.24</td>
<td>-.01</td>
<td>-.10</td>
<td>.92</td>
</tr>
<tr>
<td>C2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5.29  Sequential Regression Table for Predicted Q1 Main Idea Scores

<table>
<thead>
<tr>
<th>Variable</th>
<th>$R^2$</th>
<th>Adjusted $R^2$</th>
<th>$\Delta R^2$</th>
<th>Sig $\Delta F$</th>
<th>$B$</th>
<th>SEB</th>
<th>$\beta$</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years in English</td>
<td>.08</td>
<td>.05</td>
<td>.08</td>
<td>.05</td>
<td>-.01</td>
<td>.03</td>
<td>-.02</td>
<td>-.25</td>
<td>.81</td>
</tr>
<tr>
<td>JOL1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.01</td>
<td>.01</td>
<td>-.09</td>
<td>-.95</td>
<td>.35</td>
</tr>
<tr>
<td>NfC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.01</td>
<td>.02</td>
<td>-.05</td>
<td>-.56</td>
<td>.57</td>
</tr>
<tr>
<td>Free Recall Idea Units</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.20</td>
<td>.06</td>
<td>.31</td>
<td>3.07</td>
<td>&lt; .001*</td>
</tr>
<tr>
<td><strong>Step 2</strong></td>
<td>.53</td>
<td>.50</td>
<td>.04</td>
<td>&lt; .001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years in English</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.01</td>
<td>.02</td>
<td>.03</td>
<td>.52</td>
<td>.60</td>
</tr>
<tr>
<td>JOL1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.00</td>
<td>.01</td>
<td>.01</td>
<td>.17</td>
<td>.87</td>
</tr>
<tr>
<td>NfC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.01</td>
<td>.01</td>
<td>-.03</td>
<td>-.43</td>
<td>.67</td>
</tr>
<tr>
<td>Free Recall Idea Units</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.21</td>
<td>.05</td>
<td>.32</td>
<td>4.49</td>
<td>&lt; .001*</td>
</tr>
<tr>
<td>C1</td>
<td>1.71</td>
<td>.17</td>
<td>.66</td>
<td>10.15</td>
<td>&lt; .001*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>-.29</td>
<td>.10</td>
<td>-.20</td>
<td>-3.00</td>
<td>&lt; .001*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5.30  Sequential Regression Table for Predicted Cued-Recall Idea Units Scores

<table>
<thead>
<tr>
<th>Step 1</th>
<th>Variable</th>
<th>(R^2)</th>
<th>Adjusted (R^2)</th>
<th>(\Delta R^2)</th>
<th>(\text{Sig } \Delta F)</th>
<th>(B)</th>
<th>(SEB)</th>
<th>(\beta)</th>
<th>(t)</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Years in English</td>
<td>.29</td>
<td>.27</td>
<td>.29</td>
<td>&lt; .001*</td>
<td>.03</td>
<td>.05</td>
<td>.06</td>
<td>.74</td>
<td>.46</td>
</tr>
<tr>
<td></td>
<td>JOL1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.02</td>
<td>.02</td>
<td>.10</td>
<td>1.15</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>NfC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.02</td>
<td>.03</td>
<td>.05</td>
<td>.68</td>
<td>.50</td>
</tr>
<tr>
<td></td>
<td>Free Recall Idea Units</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.48</td>
<td>.09</td>
<td>.47</td>
<td>5.35</td>
<td>.00*</td>
</tr>
<tr>
<td></td>
<td>C1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.18</td>
<td>.33</td>
<td>-.04</td>
<td>-.56</td>
<td>.58</td>
</tr>
<tr>
<td></td>
<td>C2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.08</td>
<td>.19</td>
<td>.03</td>
<td>.40</td>
<td>.69</td>
</tr>
<tr>
<td></td>
<td>Free Recall Idea Units</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.48</td>
<td>.09</td>
<td>.47</td>
<td>5.31</td>
<td>.00*</td>
</tr>
<tr>
<td></td>
<td>Years in English</td>
<td>.30</td>
<td>.26</td>
<td>.00</td>
<td>.79</td>
<td>.03</td>
<td>.05</td>
<td>.06</td>
<td>.70</td>
<td>.49</td>
</tr>
<tr>
<td></td>
<td>JOL1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.02</td>
<td>.02</td>
<td>.09</td>
<td>1.05</td>
<td>.30</td>
</tr>
<tr>
<td></td>
<td>NfC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.02</td>
<td>.03</td>
<td>.05</td>
<td>.60</td>
<td>.55</td>
</tr>
<tr>
<td></td>
<td>Free Recall Idea Units</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.48</td>
<td>.09</td>
<td>.47</td>
<td>5.31</td>
<td>.00*</td>
</tr>
</tbody>
</table>

* Significant at the .05 level.
Table 5.31  Sequential Regression Table for Predicted Cued-Recall Main Idea Scores

<table>
<thead>
<tr>
<th>Variable</th>
<th>$R^2$</th>
<th>Adjusted $R^2$</th>
<th>$\Delta R^2$</th>
<th>Sig $\Delta F$</th>
<th>$B$</th>
<th>SEB</th>
<th>$\beta$</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years in English</td>
<td>.10</td>
<td>.07</td>
<td>.10</td>
<td>.02*</td>
<td>.02</td>
<td>.02</td>
<td>.09</td>
<td>.96</td>
<td>.34</td>
</tr>
<tr>
<td>JOL1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.00</td>
<td>.01</td>
<td>.06</td>
<td>.65</td>
<td>.52</td>
</tr>
<tr>
<td>NFC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.02</td>
<td>.01</td>
<td>.17</td>
<td>1.88</td>
<td>.06</td>
</tr>
<tr>
<td>Free Recall Idea Units</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.05</td>
<td>.03</td>
<td>.16</td>
<td>1.67</td>
<td>.10</td>
</tr>
<tr>
<td><strong>Step 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years in English</td>
<td>.16</td>
<td>.12</td>
<td>.07</td>
<td>.01*</td>
<td>.02</td>
<td>.02</td>
<td>.12</td>
<td>1.41</td>
<td>.16</td>
</tr>
<tr>
<td>JOL1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.01</td>
<td>.01</td>
<td>.10</td>
<td>1.07</td>
<td>.29</td>
</tr>
<tr>
<td>NFC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.02</td>
<td>.01</td>
<td>.13</td>
<td>1.50</td>
<td>.14</td>
</tr>
<tr>
<td>Free Recall Idea Units</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.06</td>
<td>.03</td>
<td>.18</td>
<td>1.86</td>
<td>.07</td>
</tr>
<tr>
<td>C1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.27</td>
<td>.11</td>
<td>.22</td>
<td>2.49</td>
<td>.01*</td>
</tr>
<tr>
<td>C2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.11</td>
<td>.06</td>
<td>.15</td>
<td>1.72</td>
<td>.09</td>
</tr>
</tbody>
</table>
Table 5.32  Sequential Regression Table for Predicted Total Reasons Scores

<table>
<thead>
<tr>
<th>Variable</th>
<th>$R^2$</th>
<th>Adjusted $R^2$</th>
<th>$\Delta R^2$</th>
<th>Sig $\Delta F$</th>
<th>$B$</th>
<th>SEB</th>
<th>$\beta$</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years in English</td>
<td>.16</td>
<td>.14</td>
<td>.16</td>
<td>&lt; .001*</td>
<td>.12</td>
<td>.05</td>
<td>.19</td>
<td>2.25</td>
<td>.03*</td>
</tr>
<tr>
<td>JOL1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.01</td>
<td>.02</td>
<td>-.05</td>
<td>-.52</td>
<td>.61</td>
</tr>
<tr>
<td>NfC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.07</td>
<td>.03</td>
<td>.19</td>
<td>2.19</td>
<td>.03*</td>
</tr>
<tr>
<td>Free Recall Idea Units</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.28</td>
<td>.10</td>
<td>.26</td>
<td>2.74</td>
<td>.01*</td>
</tr>
<tr>
<td><strong>Step 2</strong></td>
<td>.21</td>
<td>.17</td>
<td>.05</td>
<td>.04*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years in English</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.14</td>
<td>.05</td>
<td>.23</td>
<td>2.64</td>
<td>.01*</td>
</tr>
<tr>
<td>JOL1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.00</td>
<td>.02</td>
<td>-.02</td>
<td>-.19</td>
<td>.85</td>
</tr>
<tr>
<td>NfC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.06</td>
<td>.03</td>
<td>-.02</td>
<td>1.61</td>
<td>.07</td>
</tr>
<tr>
<td>Free Recall Idea Units</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.29</td>
<td>.10</td>
<td>.27</td>
<td>2.91</td>
<td>.00*</td>
</tr>
<tr>
<td>C1</td>
<td>.70</td>
<td>.36</td>
<td>.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>.37</td>
<td>.21</td>
<td>.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5.33  Sequential Regression Table for Predicted JOL2 Scores

<table>
<thead>
<tr>
<th>Variable</th>
<th>Variable Type</th>
<th>R^2</th>
<th>Adjusted R^2</th>
<th>Δ R^2</th>
<th>Sig Δ F</th>
<th>B</th>
<th>SEB</th>
<th>β</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td>.52</td>
<td>.51</td>
<td>.52</td>
<td>&lt; .001*</td>
<td>-26</td>
<td>.23</td>
<td>-.07</td>
<td>-1.12</td>
<td>.26</td>
</tr>
<tr>
<td>Years in English</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JOL1</td>
<td>JOL1</td>
<td>.59</td>
<td>.07</td>
<td>.58</td>
<td>8.26</td>
<td>.64</td>
<td>.07</td>
<td>.63</td>
<td>9.60</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>NfC</td>
<td>NfC</td>
<td>.36</td>
<td>.14</td>
<td>.17</td>
<td>2.61</td>
<td>.29</td>
<td>.13</td>
<td>.14</td>
<td>2.22</td>
<td>.03*</td>
</tr>
<tr>
<td>Free Recall Idea Units</td>
<td>Free Recall</td>
<td>.94</td>
<td>.44</td>
<td>.15</td>
<td>2.13</td>
<td>1.03</td>
<td>.41</td>
<td>.17</td>
<td>2.55</td>
<td>.01*</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td>.61</td>
<td>.59</td>
<td>.08</td>
<td>&lt; .001*</td>
<td>-.11</td>
<td>.21</td>
<td>-.03</td>
<td>-.52</td>
<td>.60</td>
</tr>
<tr>
<td>Years in English</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JOL1</td>
<td>JOL1</td>
<td>.64</td>
<td>.07</td>
<td>.63</td>
<td>9.60</td>
<td>.29</td>
<td>.13</td>
<td>.14</td>
<td>2.22</td>
<td>.03*</td>
</tr>
<tr>
<td>NfC</td>
<td>NfC</td>
<td>.29</td>
<td>.13</td>
<td>.14</td>
<td>2.22</td>
<td>1.03</td>
<td>.41</td>
<td>.17</td>
<td>2.55</td>
<td>.01*</td>
</tr>
<tr>
<td>Free Recall Idea Units</td>
<td>Free Recall</td>
<td>6.37</td>
<td>1.47</td>
<td>.26</td>
<td>4.33</td>
<td>2.06</td>
<td>.86</td>
<td>.14</td>
<td>2.41</td>
<td>.02*</td>
</tr>
<tr>
<td>C1</td>
<td>C1</td>
<td>6.37</td>
<td>1.47</td>
<td>.26</td>
<td>4.33</td>
<td>2.06</td>
<td>.86</td>
<td>.14</td>
<td>2.41</td>
<td>.02*</td>
</tr>
<tr>
<td>C2</td>
<td>C2</td>
<td>2.06</td>
<td>.86</td>
<td>.14</td>
<td>2.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 6.

General Discussion and Conclusion

This thesis set out to investigate the pedagogical efficacy of the DM, primarily related to information recall and comprehension, judgment of learning, critical thinking and argumentative writing. Prior research on AVTs indicated that they could be instructionally effective depending on the context in which they were applied. My research showed that students were engaged to use the tool and benefited from it in laboratory and likely classroom settings. However, students might have been constrained with the limited time allocated to finish the tasks. Higher quality argumentative writing might be generated if students were given more time and graded for the test. The laboratory study also sought to know whether training on the concept of argumentation alone could facilitate student learning without the AVT scaffolding. It was found that such training did not improve learning performances measured in this study. Furthermore, it was noted that giving the argumentation training alone might confuse students and contribute to some negative outcomes.


Research Question 1: Effects on the Studying Process

The first research question asked how using the DM affected learners’ studying process. This question is answered by group differences on highlighting activities and time spent on studying the main text (Time-on-Task). The results showed that although there was no association between group and highlighting activities prior to treatment when participants were studying Text 1, participants in the DM Group highlighted significantly less content when studying Text 2 than either of the other two groups. It may indicate that the DM had drawn participants’ attention away from highlighting. In a way, entering text in the DM boxes may have replaced highlighting as a means for
selecting and marking key information. The Time-on-Task measures showed that the three groups did not differ on time spent reading Text 1. However, the DM Group spent more time studying Text 2 than the other two groups. One plausible explanation is that the DM students needed to complete the DM in addition to reading the text, while the other two groups did not.

**Research Question 2: Effects on Judgment of Learning**

The second research question asked whether using the DM had an effect on Judgment of Learning (JOL2). As described in Chapter 4, JOL is a type of metacognitive judgment that learners use to predict their learning performances. It provides useful information on how learners monitor and judge their own learning and helps learners allocate time and effort to various tasks for better learning outcomes (Benjamin et al., 1998; Dunlosky & Hertzog, 1998; Dunlosky & Nelson, 1994). I found that the DM Group had higher JOL after the intervention (JOL2) compared to the Control Group, but the Argue Group was not statistically significantly different from the other groups. Also, the DM participants’ gain from pre-intervention JOL1 to post-intervention JOL2 was larger than that of the Control Group. A regression analysis found that, after statistically controlling for Years in English, JOL1, Need for Cognition, and Free Recall Idea Units, the participants in the DM Group were more confident in their learning than both the other two groups.

There has been no prior research on how AVTs affect individuals’ judgment of learning. Researchers have suggested that confidence increases are unlikely to be caused by the novelty of multimedia representations. For example, Chen and Fu (2003) found that representation type, word only vs. word plus picture, did not affect JOL. It is possible that certain functions of the AVT provided students an opportunity to reflect on their own learning process and consequently they felt they experienced improved learning. For example, the DM provided a visual record of what information had been studied and used in the argumentation process, so that students could constantly review their progress on argument analysis and evaluation. The DM also signalled an ending point for the argument, so that students would conclude the process when they felt they had studied all of the material.
Another possible explanation is that the DM spatially laid out information including salient interrelations among components of arguments. This might have increased students’ perceived comprehension of the study material, and if the readers were unclear with or even confused by the meaning of the material they were less likely to feel confident with their learning.

There is abundant research in the literature on the reliability of JOL measures as predictors of actual learning outcomes. Results from the present study showed moderate to weak correlations between judgment of learning and their actual performances on the recall measures. Although overestimating JOL can be detrimental (Thiede & Dunlosky, 1994), my research did not measure accuracy of JOL and the data gave no reason to believe that using the DM led participants to overestimate their learning. The goal of this study was not to investigate how accurately students predicted their learning, but to establish the association between argumentation, AVTs, and JOL. The overall pattern of the data suggested that the DM could effectively enhance students’ confidence in their learning.

Research Question 3: Effects on Critical Thinking

The third research question asked whether there was an effect of using the DM on critical thinking skills. An analysis using scores from the critical thinking scale developed as part of this thesis research yielded inconclusive results. Although definitions of critical thinking have been diverse, one of the core elements most of them share is the ability to evaluate both sides of an argument in order to solve problems (Angeli & Valanides, 2009; Chance, 1986; Willingham, 2008). The DM requires students to weigh the pro and con arguments of the main issue by providing supporting or opposing evidence, warrants to the evidence, weights to each reason, and links across arguments and counterarguments. As a result, students reach their conclusions with sound justification from the above argumentation process. These processes align well with critical thinking skills. In addition, previous research showed close connection between the use of AVTs and critical thinking skill gains. For example, van Gelder et al. (2004) reported a 0.8 standard deviation gain in just one semester of AVT implementation. However, data analyses in the present study did not detect a difference
on students’ performance on the six critical thinking questions across the treatment conditions.

One plausible explanation is that the scale lacked validity as a measure of critical thinking skill. Due to limitations in time and resources I chose to use the instrument without prior empirical assessment of validity and reliability. The results showed that it had poor internal consistency. Although the commercially available CCTST is claimed to be a reliable instrument, administering the whole CCTST would have taken too much time, and the company that owns it did not allow only part of the test to be deployed. Although the instrument I developed was modeled on the CCTST, which uses multiple-choice questions, it is possible that a better-designed assessment tool (such as asking students to write justifications for answer options that they consider as correct or incorrect) might have shown that using the DM improves critical thinking.

The absence of a pronounced group difference on the critical thinking measure may also be due to participants’ fatigue. Participants on average spent 1.5 hours completing the whole task, and the critical thinking questions were positioned in the later section of the task. Although the sessions were similar to the duration of final examinations in many undergraduate courses, the participants were working without the powerful incentive of academic grades. Due to the relatively lengthy descriptions of scenarios in each question, it is quite likely that the participants rushed to finish and might not have been paying due attention to the questions or answers. It is also possible that critical thinking skills may take longer to develop than studying one text.

**Research Question 4: Effects on Recall and Comprehension**

The fourth research question asked whether there was an effect of using the DM on recall and comprehension. This question is answered by the group effects on the number of Idea Units and Main Ideas included in the summaries and cued-recall responses, as well as the number of errors made in the summaries.

Previous research has demonstrated benefits AVTs have on memory and comprehension, measured by prior knowledge activation (Salminen et al., 2010), cued-recall (Dwyer et al., 2010; Dwyer et al., 2013), and deeper understanding (Salminen et
al., 2010). Results from the present study are consistent with these effects. Although no group differences were detected for the Idea Units measures, the DM Group used more Main Ideas in both their summary and cued-recall responses than either the Argue Group or the Control Group. Participants in the DM Group also made fewer errors in their summary than those in the Control Group. The Argue Group was not statistically significantly different from the other groups. Noting that the DM Group made more errors in their pre-test free recall, the number of errors decreased significantly for this group compared to the Control Group in a repeated measure test. This suggested that the DM helped improve the accuracy of students’ memory. Even after statistically controlling for other contributing factors, such as Years in English, Need for Cognition, and pre-test measures (Judgment of Learning for Text 1 and Free Recall of Text 1), the DM was still a significant predictor of the Main Idea measures. Note that the MANOVA test showed no significant differences between 1) the DM Group and the Control Group on Main Ideas Used in the summary and 2) the DM Group and the Argue Group on Cued-Recall Main Ideas. After controlling for the four predictor variables (Years in English, JOL1, NfC, and Free Recall Idea Units), the DM Group outperformed both the other two groups on the two Main Ideas measures.

There are several possible explanations for these findings. First, because completing the DM generally requires condensing information to fit in the text boxes, it may function as an exercise for gist extraction. According to the fuzzy-trace theory (Reyna & Brainerd, 1995), there are two types of memory as learners process text information, that is, verbatim memory and gist memory. Verbatim memory refers to memory of specific details of a text. Gist memory, on the other hand, contains inferred meanings. Main ideas involve synthesis of information into a coherent statement that represents the inferred meaning of a paragraph. Therefore, Main Idea usage can be seen as an indicator of gist extraction. The DM may have facilitated gist extraction by focusing students’ attention on the relationships among components of their arguments, such as reasons and their supporting evidence. Such relationships are essential for making inferences and are usually implicit in written text due to its linear representation format. In addition, since Main Ideas are not readily given in the text, and extracting Main Ideas involve synthesizing information, identifying Main Ideas can also be viewed as an indicator of comprehension (Durkin, 1978). Therefore, in this study recall and
comprehension are discussed together in terms of Main Ideas, because one cannot recall the gist of a paragraph or text without understanding it.

Second, it is likely that constructing the DM helped students focus their attention on relevant information, because the DM features signalled or directed their attention to important information. This is what Mayer calls the signalling principle for multimedia learning, defined as “highlighting essential material” (Mayer, 2014; Mayer & Moreno, 2003). In other words, the DM may have drawn students’ attention away from content that is less important or trivial. The DM also signals the principal structural relationships among argumentation components. These signals can be used to extract useful information in order to understand the structure of the text. When students were signalled to identify a reason, in order to fill out the reason text box in the DM, they needed to synthesize the information into a concise statement. These exercises closely resemble the extraction of Main Ideas for summarization and answering cued-recall questions. Specifically, in the summary only important information was included, and short-answer questions required students to include information that was pertinent to answering the questions. It is important to note that caution is needed when discussing student attention before it can be verified by measurements such as eye-tracking.

The DM may also have helped students activate, strengthen, and refine their existing schemas or build new ones. The DM, by visually displaying a structure with two sides to fill in arguments and counterarguments, an issue to solve on top, and a conclusion to reach on the bottom, would help students activate, construct, or polish an argument schema. Researchers have noted that schemas guide remembering and interpretation (Reznitskaya & Anderson, 2002). Andersen (1984) also noted that schemas assist tasks like editing and summarization. The reason is that a schema includes conditions of what information is important within the structure; the reader can construct summaries that include significant ideas and exclude negligible ones based on such a schema. Students are also more likely to recall information that is deemed important. Moreover, argument schemas enable individuals to interact with texts differently (Reznitskaya & Anderson, 2002). Once a text is recognized as argumentative, a reader would process the text by using the slots in the schema. For example, they would search for claims, evidence, counterclaims, and rebuttals. Researchers believe
that argument schemas direct student attention to argument-relevant information and facilitate argumentation construction (Reznitskaya & Anderson, 2002). Comprehension and recall of the argumentative text are also enhanced (Anderson, 1984; Armbruster, Anderson, & Ostertag, 1987; Carrel, 1992). Although the cued-recall and summary writing tasks were not explicitly argumentative, they shared elements of the basic argument schema (Reznitskaya, et al., 2007).

Another plausible explanation is that graphic organizers are believed to lower extraneous cognitive load. They have been proven effective in enhancing reading comprehension (Moore & Readence, 1984). The DM, as a type of graphic organizer, was designed based on several cognitive theories including the cognitive load theory. Information regarding logical relations among argument components in the DM is organized spatially with high saliency. As a result, fewer cognitive resources are required to make inferences.

The group differences on the number of Main Ideas used may also be comparable to Salminen et al.’s (2010) findings on prior knowledge activation in students’ argument diagrams. They reported that students included more prior knowledge when they freely constructed the AVTs than when they were asked to modify existing ones. The authors attributed this outcome to the benefits of multimedia learning. Taking our cue from Salminen et al., a plausible explanation of my results suggested by multimedia learning theory is that students were able to make more efficient use of cognitive resources when using the DM. For example, they may have been able to offload some cognitive processing of evidentiary relationships to the visual processing channel and leave more processing capacity in the verbal channel (Mayer, Bove, Bryman, Mars, & Tapangco, 1996; Mayer, 1999).

While it was expected that the DM Group would outperform the Argument Group, because the latter did not have the opportunity to study with the AVT, it was somewhat surprising to discover that the Argument Group was also outperformed by the Control Group on Main Ideas used in the summary. It is possible that the Argument Group, by receiving the verbal argument training alone, did not get a full argument schema activated or updated. Wolfe, Britt, and Butler (2009) presumed that people use schemas
to different extents when reading or writing. Based on this assumption, it is reasonable to assume that participants in my experiment came to the lab with argument schemas of different complexity. According to the research showing that students are often poor arguers, it is highly probable that they possess and use a deficient argument schema when reading and writing. Consequently, they need additional scaffolding such as the DM in order to improve or activate a more advanced argument schema for enhanced learning performance.

The verbal argumentation training included ideas such as claims and evidence, arguments and counterarguments with short examples explaining how each component functioned in an argument. Research has shown moderate effectiveness of such training in learning about argumentation (Nussbaum & Kardash, 2005). However, my results showed no positive learning effect from verbal training alone for recall and comprehension. Nussbaum and Schraw (2007) stated that verbal instructions do not lower the cognitive load required in generating two-sided arguments. They did not test for other learning outcomes, but it could help explain the lack of desirable learning effects by verbal training alone from the cognitive load perspective. For desirable learning outcomes, students may need more scaffolding in addition to verbal argumentation training such as using an AVT.

Similar to the DM Group and the Control Group, when the Argument Group constructed the summary, they needed to retrieve the text content from their memory. Argue Group participants may also have tried to retrieve information from the argumentation training and attempt to make a connection between the training and the task. That is, participants may have done preliminary processing of information such as deleting some trivia, reorganizing some content, or paraphrasing some text. In addition to this processing, they may also have been trying to make use of what they learned from the argumentation training. This processing and effort to link information may have increased the memory load. While they did not have the DM to offset this overload, students may have resorted to a reduction of effort in synthesizing raw information into Main Ideas in order to decrease the cognitive burden.
Research Question 5: Effects on Argumentative Writing

The Fifth research question asked whether there was an effect of using the DM on argumentative writing. The overall pattern of the data suggested that participants in the DM Group had a stronger tendency to argue (made more argumentative moves) when given writing tasks, including summarization and short-answer questions. In this thesis, argumentative writing was operationally defined by 1) in the summary, argumentative moves such as adopting the introduction-body-conclusion organization that is conventional for rhetorical writing, 2) in the summary, more use of Argument Markers, 3) in the summary, infrequent use of non-argumentative information (Neutral Idea Units), and 4) in the short-answer questions, a greater number of reasons used. Group comparisons showed that the DM participants were more likely to organize their summary in an introduction-body-conclusion type of structure compared to participants in the other two groups. Multiple regression analysis showed that the DM intervention was a significant predictor of such organization of information even after effects brought by the four predictor variables (Years in English, JOL1, NfC, and Free Recall Idea Units) were statistically controlled for. Although a MANOVA test did not show group differences on the number of Argument Markers used in participants’ summary, the DM treatment was a significant predictor of the use of Argument Markers after controlling for the four predictor variables. A follow up regression showed that the difference was between the DM Group and the Argue Group. The Control Group did not differ from the other two groups on this measure. In addition, the DM Group used significantly fewer Neutral Idea Units than the Control Group. In summary, participants in the DM group outperformed at least one of the other two groups on the above four defining measures of argumentative writing.

Schema theory may offer some explanations for this tendency to argue. As discussed earlier, it is likely that the DM helped students construct, refine an argument schema or activate an existing one. The argument schema includes the components of an argument, the relations among the components, and what information is important in the schema. As described in Chapter 2 and 3, based on the design goals and features of the DM, the argument schema associated with the DM would be affiliated with the Toulmin model (Toulmin, 1958), Walton’s argument schemes (Walton, 1996), and a
dialectical schema (Basseches, 1979). It may also be seen as an amalgamation of some or all of these known schemas. This argument schema includes an overall issue to be resolved, a series of claims and counterclaims to be evaluated with their supporting evidence, warrants that support the evidence when necessary, and a conclusion as an ending point. Walton described logical argumentation with a remarkably similar structure that contains a starting point, a middle stage with dialogues between agents who take turns to make assertions and ask questions, and an end point (Walton, 2013). Such a structure is likely to facilitate students in organizing their writing in a similar a way, where the main issue is stated in an introduction, the argumentation process that involves weighing claims and evidence forms the body, and an integrated conclusion is stated at the end.

As described previously, schemas also contain the criteria to judge what information is important. This may help explain why fewer Neutral Idea Units were used by the DM participants in their summary. Neutral Idea Units by definition do not contribute to either side of the argument. They merely provide factual background information on the topic. As described earlier, once students activate an argument schema, they process the text as argumentative and try to fill in the slots in the schema, for example by looking for claims and evidence. They are more likely to make more argumentative moves, such as using Argument Markers, when they write about the argumentative source text to express the schema.

Another possible explanation may be based on the relationship between schemas and remembering (Rumelhart, 1980). People remember their interpretation or the gist of a text rather than the details of the text and the interpretation is guided by schemas. When we recall information, it is schemas again that are used to reinterpret the stored data/memory and reconstruct the original interpretation. Akin to this process, participants who used the DM interpreted the study material as argumentative and encoded the information accordingly in their memory. When completing tasks that required recalling/remembering information from the text, including writing a summary of the text, the argument schema was activated to reinterpret and reconstruct the meaning as argumentative. This may have been the main reason that the DM participants used more Argument Markers to reflect this interpretation and reinterpretation.
There are other theories that could help explain the results. For example, Robinson and Kiewra (1995) found that graphic organizers using visual argument promote understanding the text structure (Loman & Mayer, 1983) of reading material, which includes the coordination and hierarchy among concepts. Although the text structure was intentionally made implicit in the reading material used in my research, the DM may have helped participants in recognizing the hidden structure and in turn helped students write summaries reflecting such features. The DM may have facilitated students’ understanding of the relations among concepts through visually positioning and connecting them in a spatial/schematic way. This is believed to enable students to build a mental network that indexes and retrieves information better than reading a linear text alone (Larkin & Simon, 1987). Thus, they included more organizational features in their writing. This may also help explain the better recall performance the DM participants demonstrated on Main Ideas.

Previous studies showed that AVTs help students’ reasoning skills (Carr, 2003; Easterday et al., 2007; Okada & Buckingham Shum, 2008). However, contrary to my initial expectations, the number of reasons used did not differ across groups. I noticed that the answers were rather short (on average 46.3 words per answer), and the numbers of reasons were very low (on average 1.9 reasons per answer). This may have been because these questions were positioned at the end of the task and participants were fatigued so they rushed to finish. This is highly likely especially when students were not graded for the tasks. Some participants even left some answers blank. The overly brief responses obscured any inclination participants may have had toward providing more reasons.

In summary, due to the characteristics of argument schemas, individuals with a developed argument schema have different experiences when interacting with a text. Results from the present study showed that using the DM increased students’ confidence in their learning and facilitated information recall and comprehension. Participants also demonstrated a greater tendency to argue and made more argumentative moves in their writing after the DM intervention. These results suggested that the readers were not merely encoding information they found in the texts. The argument schema helped them monitor their own learning, extract more relevant
information, and produce more organized summaries. Verbal argument training alone did not seem sufficient to generate such desirable learning outcomes, possibly due to the lack of visualization scaffolding. Participants appeared to need more time to complete the tasks adequately, but overall the DM showed great potential in facilitating learning.

6.2. Discussion on Classroom Implementations

In Chapter 3, I presented three classroom applications of the DM. Overall, both the students and the instructor expressed favourable experiences with the DM. I am aware of the selective nature of these implementations and the lack of control conditions. Therefore, I am cautious in making any generalizations from these applications. These three in situ studies and the lab experiment began a larger research enterprise on how to improve the design features and learning outcomes of the DM with iterations of research designs and data collection. Through preliminary instructor reports with the current implementations, I believe there is cause for optimism that the DM facilitates the learning of argumentation. More research is needed to confirm these findings. Moreover, the instructor observed possible transfers of the argumentation skills students learned from using the DM to other critical thinking tasks in the writing samples and on the exam questions. Future endeavours could involve more in depth data analyses to confirm results of such observation.

My colleagues and I had presented the DM at educational conferences. Many fellow attendees expressed interest with using the DM tool in their courses. For example, at the STHLE conference (Niu et al., 2015), an instructor from a local university described the growing importance of argumentation in their curriculum. She had been experimenting with argumentation scaffolding on paper and asked if we could make the DM available to her university.

In a simplistic view, this new AVT may also be seen as a hybrid between a typical Toulmin style argument map (see Figure 3.1) and the Argumentation Vee Diagram (see Figure 3.2). It may be true that the DM capitalizes on both the graphical representation of the relations among components of an argument with boxes and
arrows and the scaffolding of claim-counterclaim resolution. However, any changes in learning outcomes that may result from seemingly small tweaking of design features from existing tools should not be seen as necessarily additive and linear due to the complexity of human thought and its interactions with instructional technologies. A series of randomized comparison studies are needed to confirm or reject the hypothesis that the DM offers scaffolding for argumentation that is significantly enhanced relative to prior tools such as the Argumentation Vee Diagram. This sets the horizon for part of our future work with the DM.

In its infancy, the DM is currently undergoing preliminary testing and modifications for improvement. Ultimately, we would like to redirect our focus and energy back to authentic instructional settings after we gather abundant data from trial implementations and controlled experiments. After all, the fundamental goal for developing instructional tools is not to improve results obtained from manipulating lab parameters in a vacuum, but rather to facilitate meaningful and authentic teaching and learning in real life.

6.3. Theoretical Implications

Findings from this thesis complemented previous research on AVTs and improved learning outcomes (Carrington et al., 2011; Lund et al., 2007; Suthers, 2003). Prior research has shown the importance of schemas for learning (Bigenho, 1992; Sadoski et al., 1991). This thesis provided empirical evidence that students can benefit from having a well-developed argument schema and that the acquisition and development of such a schema can be facilitated by studying with an AVT. Particularly, the DM showed great potential in improving learning in several areas: recall and comprehension and argumentative writing. The use of the DM also increased learners’ confidence in learning. The newly developed DM, which integrated several argumentation theories and cognitive theories, bridged a gap in the AVT field. It addressed several design challenges many AVTs encounter, which potentially can lower the instructional efficacies of such tools, such as lack of support to generate counterarguments and rebuttals, lack of signalling to conclude the argument, and lack a balance between flexibility and guidance.
In addition, the present study extended previous research in a number of ways. First of all, most AVT research has focused on teaching students how to argue. The present study examined the effects of using the DM on argument skills and also on participants’ tendency to argue. In the instructions for writing the summary the participants were not asked to argue. The increased tendency to use argumentation detected by this measure is potentially a very powerful outcome, because the inclination to argue is likely to transfer to many other learning situations. Second, summarization is not usually used in AVT studies. This shows a theoretical gap that neglects the connections between argumentation and the cognitive processes involved in summarization. Summarization is an organizational process (Kintsch & Kozminskey, 1977) in addition to recall of information. This overlaps with the rhetorical significance of the argumentation process. As explained by Hidi and Anderson (1986), when writing a summary, readers need to not only choose what to include or delete from the original text, but also decide how to transform and organize the ideas to make coherent sense. Argumentation also requires the organization of ideas for coherence in order to make sense to a potential “rational judge” (van Eemeren et al., 1996, p 5). The organizational pattern from the present study, which included an introduction that gives an overview of the summary, a body that expands the introduction, and a conclusion indicating a well-reasoned solution, has not been particularly investigated for argumentative writing. However, van Eemeren, Houtlosser, and Snoeck Henkemans (2007) did include indicators of the beginning and end of an argument as important markers for argumentative discourses. This study helped confirm and extend this model with empirical data. Thus, the present study suggested that summarization could be a useful method to characterize argumentation in addition to the traditional approaches such as argumentative essays and critical thinking measures.

The present study also made an initial connection between the AVT and JOL literature. It provided insights to understanding the nature of argumentation and its influence on learners’ metacognitive processes and increased our knowledge about the human mind.
6.4. Practical Implications

Overall, this thesis provided additional evidence that AVTs can be instructionally effective. They can benefit an average student with not only recall and comprehension, but also in strengthening their tendency to argue. This is useful in many, if not all, subjects in school. This thesis provides sufficient basis for instructional designers to further explore the applications of the DM in multiple settings.

The results showed that students were able to use most of the functions provided in the DM with minimal difficulty. Previous research suggested that students are less engaged to fully use AVTs in laboratory experiments, because they are not receiving grades for doing so. As a consequence, they do not learn better (Pinkwart et al., 2009). However, my findings suggested that most participants in the DM Group used the map to a substantial degree. Moreover, previous research contended that warrants as specified by Toulmin are often complex and difficult to use, unless extensive scaffolding is provided (Knudson, 1994; Stein & Albro, 2001; Nussbaum & Schraw, 2007; Yeh, 1998). In the present study, the warrant boxes in the DM were frequently used and in most cases accurately. This suggested that the DM offered the scaffolding necessary to understand the complex concepts involved in argumentation.

Sometimes instructional strategies that are proven successful in laboratory experiments can unfortunately be difficult for instructors or students to apply. However, the three classroom implementations in this thesis suggested possibilities to use the DM centrally in classes for full course-length. The teacher who implemented the DM in her courses expressed satisfaction with the ease of use and how well the DM fit with her regular teaching routine. Both the teacher and students expressed very positive attitudes toward the DM. The teacher also suggested promoting the DM to the whole university.

This study also logged participants' interactions with the study material and the DM through Time-on-Task measures, which helped the researcher to better understand students studying processes. When it comes to classroom implementations, I advise instructors to give students enough time to read and study with the DM in order to facilitate the best learning outcomes. Students should also be given the opportunity to study with the DM as they normally would with other materials. For example, this would
apply to the DM by allowing students to make revisions to improve their DM over time and reviewing their DMs before exams.

6.5. Limitations

There are several inherent limitations of this study. First, I made observations on how participants behaved during the experiment sessions and took notes. The observations were not as thorough as they could have been due to the fact that I needed to travel between sites to train participants so as not to disturb other participants. Therefore, the observation data were not used for inferences. However, interesting information sometimes arose. For example, a student was found to be taking notes even though the instructions clearly said no note taking was allowed. The notes were likely used for recall responses and that may have inflated her scores. An examination of this participant’s data did not reveal much difference compared to participants from the same group, therefore I decided to keep her data as they were.

Second, due to the time limit, the reasoning answers were overly brief, averaging about 46.3 words or three to four sentences in length and some answers were left blank. It is likely that longer responses would yield different outcomes. The reasoning measure I used was consequently limited to a one-dimensional measure (Total Reasons). If there were more reasons, I could have categorized them based on their quality as in Means and Voss (1996), or measure specific argumentative moves such as counterarguments and rebuttal. However, the numbers of reasons were too scarce to carry on additional analyses and the responses were too short to develop measurable counterarguments and rebuttals. Therefore, it is highly probable that given more time participants would have written longer answers, which would allow more sophisticated analyses that can capture some rich argumentative activities.

Third, this research inherited the common limitations of laboratory experiments. Variables are better controlled than in an authentic classroom, but students are usually less motivated or motivated by different factors from classroom applications (monetary incentives vs. grades). Although I obtained positive results from the three classroom implementations that showed high engagement and improvement in learning, they are
less generalizable because of the lack of control conditions. Nevertheless, it is useful to discuss some key features of these implementations to compliment the experimental study conducted for this thesis.

6.6. Future Research

Based on the findings and limitations of this study, I have several recommendations for future research. First, it is useful to investigate how the DM works with users from various age groups and grade levels. Both my experimental study and the three classroom implementations worked with university students. Furthermore, based on the essential role argumentation has on learning, it is important to examine whether the DM can provide effective support to learners of different ages and grade levels. Since argumentation skills are often viewed as developmental (Kuhn, 1991), it is highly probable that the DM will benefit younger learners as well as or more than was indicated for university students.

Second, more classroom implementations should be carried out. As suggested by Pinkwart et al. (2009), students may not be as motivated to use AVTs when they are not being graded. Even though my study results showed satisfactory DM usage, it is important to investigate if there is room for improvement. Using the DM as a central feature of course learning activities, as well as grading students on their DM work, could potentially result in better or possibly creative uses of the DM. Therefore, it is useful to keep exploring the DM’s efficacy with more classroom implementations.

Third, I recommend a better critical thinking measure. Future research could extend the Critical Thinking scale developed in this study and establish its psychometric properties. Due to the large amount of time and large sample sizes needed to validate a new measure, I opted to improve the critical thinking measure in future endeavours. Although no effects were found using the current measure, I am hopeful that with more rigorous testing and calibration, we could develop a tool that consistently measures the construct of critical thinking. This scale will be useful for further research on argumentation and argument visualization.
Fourth, I recommend collecting more qualitative data such as student feedback to compliment the numerical data. I did not include student feedback from the experiment, because 1) it would have increased the completion time significantly if written feedback was required and 2) students completed their tasks at different times in multiple sessions that I ran simultaneously. The procedural limitations prevented me from systematically collecting verbal feedback from each participant. As a result, I only spoke with participants who were available to talk about their experiences. Several participants from the DM Group expressed satisfaction with the experience and favourable attitudes toward the tool. A graduate student, however, remarked that the DM might be more helpful for students who were less experienced with argumentation. Some research has shown that AVTs tend to benefit lower aptitude students more (Harrell, 2011; Pinkwart et al., 2009). Other research has reported that even students majoring in Philosophy lack sufficient argumentation skills (Harrell, 2005). It is useful to incorporate such feedback for the betterment of the DM design or its implementation.

Fifth, laboratory experiments usually constrain learning assessments to immediate testing. Although it is harder to keep track of participants after they have left the laboratory, it is important to investigate the long-term effects the DM may have on student learning. For future research when there are fewer time or resource limitations than the present study, I recommend follow-up delayed testing for such effects. Research has yielded discordant results on whether learning effects diminish over time when students study with graphic organizers (Robinson & Schraw, 1994). It appears that the effects are contingent on the relational structure of the graphic organizers. On one hand, the structure provided by graphic organizers helps students learn relationships among concepts rather than a list of facts. As a consequence, students produce better delayed performances. On the other hand, if too much structure is provided students tend to minimize their cognitive effort and process information in a shallow way producing worse delayed performances. Prior AVT research typically did not incorporate delayed testing, even when the interventions were a semester long. Assessments are usually carried out at the end of the interventions as an ending point of a research project. I am hopeful that due to the balanced design features of the DM, it may offer students the desirable amount of flexibility and structure needed for longer-term benefits.
6.7. Conclusion

Despite its limitations, this thesis offers both theoretical and practical contributions going beyond previous research on AVTs and learning. It contributes significantly to the design and implementation of AVTs by bringing together research on argumentation, AVT, and learning theories. This thesis also presents the first empirical study that investigated the pedagogical efficacy of a newly developed AVT, the Dialectical Map. Findings from this work can help instructors, instructional designers, and students to better apply, advance, and learn with AVTs in various settings. This thesis also starts the momentum for an on-going research project to continuously improve the design and implementation of the DM.
References


tation on concept learning in evolutionary theory. *Journal of Educational

Barth, E.M. & E.C.W. Krabbe (1982). From axiom to dialogue: A philosophical study of


Bell, P. (2001). Using argument map representations to make thinking visible for
individuals and groups. In T. Koschmann, R. Hall, & N. Miyake (Eds.), *CSCL 2,
carrying forward the conversation* (pp. 449–485). Mahwah, N.J.: Lawrence
Erlbaum Associates.

When retrieval fluency is misleading as a metamnemonic index. *Journal of
Experimental Psychology. General, 127*(1), 55–68.


Bloomington.*


41–56.


In R. Harper (Ed.), *Computer supported cooperative work* (pp. 3–24). Springer.

Improving critical thinking using web based argument mapping exercises with
268–291.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and


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


Appendix A.

Reading Material Text 1

Shale gas extraction: Horizontal drilling and hydraulic fracturing

Conventional versus unconventional gas

Natural gas is widely used for heating homes, generating electricity, and as fuel for manufacturing. Conventional natural gas is generally free gas trapped in relatively small, porous locations beneath rock formations and flows freely to the surface once the well is drilled. Unconventional natural gas is trapped in rock foundations with very low permeability and does not flow freely to the surface after the well is drilled. Most of the growth in supply from today’s recoverable gas resources is found in unconventional forms such as shale gas. Shale gas is usually located further underground than conventional gas and is more difficult to extract, but the potential supply is thought to be huge. Nations like China, Argentina, Canada and U.S. each hold hundreds of trillion cubic feet of technically recoverable shale gas.

Horizontal drilling and hydraulic fracturing

The combination of horizontal drilling and hydraulic fracturing, called fracking, has made unconventional gas supplies commercially feasible. Oil and gas companies have used fracking on vertical wells since the 1950s. Horizontal drilling or directional drilling, which increases the exposed length and surface of the gas reservoir by drilling through it at an angle, was first developed in the 1920s to improve efficiency. The fracking method that is often discussed today only surged a decade ago when oil companies began fracking along horizontal wells and producing gas in dramatically larger quantities.

The shale gas extraction process involves drilling the well vertically down to a maximum of 2.5 kilometres below the surface and then drilling a few hundred metres horizontally. The well is then cased with cement to protect ground water from contamination from fracking fluids. The fracking fluids, which are water containing a small percentage of chemicals, are injected at a high pressure to create small fissures of about 1mm in the shale formation. Finally, sand or solid, ceramic beads are pumped in to hold the fissures open in order to release the gas from the pores of the shale rock.
Appendix B.

Reading Material Text 2

Shale gas extraction: Horizontal drilling and hydraulic fracturing

Conventional versus unconventional gas

Natural gas is widely used for heating homes, generating electricity, and as fuel for manufacturing. Conventional natural gas is generally free gas trapped in relatively small, porous locations beneath rock formations and flows freely to the surface once the well is drilled. Unconventional natural gas is trapped in rock foundations with very low permeability and does not flow freely to the surface after the well is drilled. Most of the growth in supply from today's recoverable gas resources is found in unconventional forms such as shale gas. Shale gas is usually located further underground than conventional gas and is more difficult to extract, but the potential supply is thought to be huge. Nations like China, Argentina, Canada and U.S. each hold hundreds of trillion cubic feet of technically recoverable shale gas.

Horizontal drilling and hydraulic fracturing

The combination of horizontal drilling and hydraulic fracturing, called fracking, has made unconventional gas supplies commercially feasible. Oil and gas companies have used fracking on vertical wells since the 1950s. Horizontal drilling or directional drilling, which increases the exposed length and surface of the gas reservoir by drilling through it at an angle, was first developed in the 1920s to improve efficiency. The fracking method that is often discussed today only surged a decade ago when oil companies began fracking along horizontal wells and producing gas in dramatically larger quantities.

The shale gas extraction process involves drilling the well vertically down to a maximum of 2.5 kilometres below the surface and then drilling a few hundred metres horizontally. The well is then cased with cement to protect ground water from contamination from fracking fluids. The fracking fluids, which are water containing a small percentage of chemicals, are injected at a high pressure to create small fissures of about 1mm in the shale formation. Finally, sand or solid, ceramic beads are pumped in to hold the fissures open in order to release the gas from the pores of the shale rock.

Economic implications

Lower natural gas prices. The use of fracking and horizontal drilling to access shale gas is expected to lower natural gas prices worldwide. There is evidence that the price for natural gas is already starting to drop in the US. Increased gas extraction from fracking led to greater supply of gas in the U.S. market, and greater supply may have caused lower gas prices. When the price for natural gas is lower, it sometimes happens that the cost of drilling wells is greater than the value of the gas that is extracted.

Investment and job creation. In the UK, shale gas production could potentially attract almost £4 billion investment per year and create 74,000 jobs. In the US, it is estimated
that shale gas could draw $71 billion in investment, and President Obama claimed that
fracking is likely to support 600,000 jobs by the end of the decade.

Lower income tax. Taking Canada as an example, residents in New Brunswick, Nova
Scotia and Quebec pay significantly more in provincial taxes than residents of the oil and
gas producing provinces of Saskatchewan and Alberta. Treasuries of the latter provinces
are less dependent on personal income taxes because they receive revenues from
companies in the energy industry. In 2012–2013, the proportions of provincial revenue
from oil production were 0.1% for Nova Scotia, 2.1% for New Brunswick, and 0.6% for
Quebec. In contrast, the proportions for Alberta and Saskatchewan were 20% and 23%
respectively.

Political implications

The expansion of shale gas in North America is transforming the natural gas market. For
example, 75% of Pennsylvania’s everyday natural gas consumption is currently
imported. However, the Marcellus Shale site in the United States alone holds enough
recoverable natural gas to not only meet Pennsylvania’s own needs, but to turn the U.S.
into a significant exporter of energy. In 2012, the U.S. president announced that the
country had enough gas supply for the next 100 years. Since domestic production of
shale gas enables nations like the U.S. to be independent from oil producing countries
like Russia and Iran, these exporting countries will no longer be able to use oil and gas
as political leverage on issues such as nuclear weapon testing.

Environmental considerations

Natural gas burning emits 50% less greenhouse gas than coal fire generation. It could
greatly decrease the carbon footprint and be a transitional energy source until cleaner
renewable energies, such as wind, solar, and nuclear become more widely used.
Carbon emissions have reduced by 9% since 2005, and 50% of the reduction has been
attributed to shale gas use. Additionally, shale gas mass production provides an
important boost to the production of nitrogen-based fertilizers for agricultural purposes.

People usually have favourable attitudes toward renewables due to environmental
considerations. They tend to overlook the cost effectiveness and the fact that any kind of
energy generation has certain impact on the environment and human life, for example
wind power systems are associated with pollution caused by the use and mining of
neodymium; solar powering requires very large amount of land, biofuel generation
competes with food supplies over farmland use, and hydro powering system can disrupt
natural aquatic ecosystems. And the ever-growing nuclear powering runs the risk such
as the Fukushima meltdown, which left radiation contaminations that will keep damaging
the global environment and human health for decades and longer.

In recent years, much of the media attention, debates and conflicts have centred on
fracking and its risks to the environment and human health. On October 18th, 2013, a
protest led by Elsipogtog First Nation members in New Brunswick against a proposed
hydraulic fracturing site on their land evolved into a violent encounter with the RCMP.
Soon the protest went nationwide. Currently some states in the US, such as New York,
and countries, such as Northern Ireland and France, have put moratoriums or even bans
on fracking. Hamilton ON, also passed resolution against fracking recently to protect
their drinking water.
Fracking wellbores are drilled through drinking water sources in order to access the gas further below. Well drilling activities may cause disturbance to groundwater quality. Chemicals and natural gas may leak from the wellbore if it is not properly sealed and cased. While there are requirements for well casing and integrity, accidents may occur, such as the explosion of a water well in Dimock, Pennsylvania. Some groundwater contamination is demonstrated in YouTube videos that show kitchen tap water being set on fire in areas close to a fracking site.

There are also serious challenges associated with storing, transporting, treating, and disposing of fracking wastewater. It must be processed at facilities that are not always equipped to remove harmful contaminants. Although the fluid used in fracking is mostly water, various kinds of chemicals are added to make fracking more effective. Fracking also releases naturally occurring salts, metals, radioactive elements like barium and strontium and carcinogens like benzene into the wastewater. These wastewater pollutants could potentially leak into the natural environment.

Shale gas extraction requires more water than traditional vertical wells, ranging between 3 to 5 million gallons of water per well. Although the industry claims that most of the water can be recycled, depending on the data sources, it is reported that only 30%–80% of the water flows back to the surface after fracking. This recovered water is unfit for further use before being treated for contaminants. Much of the recovered water is disposed directly in an underground wastewater well and not recycled. Such massive water consumption can divert water from other important uses such as agriculture, especially if the area is suffering from drought.

During the extraction process, some of the methane gas may escape. Methane is a greenhouse gas that can cause global warming. Some believe that damage caused by the fugitive methane outweighs the environmental benefits of using natural gas. Some people who live near fracking sites complain of noxious fumes that cause headaches, nausea and other symptoms that they attribute to the chemicals released during fracking. The *New York Times* reported that parts of Texas where fracking is common have higher rates of asthma. However, these areas generally have high air pollution from sources other than fracking.

The U.S. Geological Survey reported an increase of earthquakes in the U.S. midcontinent since 2001. Although it is not certain whether or not these earthquakes are caused by the fracking process, earthquakes have been linked to the wastewater disposal. U.K. experts believe that recent earthquakes in northwest England were caused by wastewater injection into a nearby fault zone.

**Standards and regulations**

Research has identified key factors that cause the leakage from fracking, which is the main environmental concern. In particular, the number of casings and how these casings were cemented is an important factor. Some of the leaky wells in Canada had only one casing or were uncased except in the section from the surface to just below the aquifer. Others had not been cemented at all or the cementation had not reached the required height. At a minimum, the cement should extend above exposed water or methane bearing zones. In some states, such as Pennsylvania and Texas, there is a requirement to cement casing of approximately 75 ft below any aquifers. Failure to do so can lead to groundwater contamination as occurred in Dimock. In the U.K., the standard practice is
to have three casing with at least 2 casings passing through in order to isolate groundwater zones.
Appendix C.

List of Idea Units

This file contains all the Idea Units with their tags. Tags that start with a letter P indicate Pro Idea Units; tags that start with a letter C indicate Con Idea Units; and tags that start with a letter N indicate Neutral Idea Units. The number following the letter P, C, or N represents the number of the sentence that the Idea Unit is in. The letter following this number indicates that there is more than one Idea Unit in a specific sentence.

N1: (this means sentence #1 is an Idea Unit, and it is a Neutral Idea Unit) Natural gas is widely used for heating homes, generating electricity, and as fuel for manufacturing.

N2a: (this means sentence #2 Idea Unit a, and it is a Neutral Idea Unit) Conventional natural gas is generally free gas trapped in relatively small, porous locations beneath rock formations.

N2b: (this means sentence #2 Idea Unit b, it is neutral) …flows freely to the surface once the well is drilled.

N3a: Unconventional natural gas is trapped in rock foundations with very low permeability…

N3b: …does not flow freely to the surface after the well is drilled.

N4: Most of the growth in supply from today’s recoverable gas resources is found in unconventional forms such as shale gas.

N5a: Shale gas is usually located further underground than conventional gas…

C5b: (this is a Con Idea Unit) …more difficult to extract…

P5c: (this is a Pro Idea Unit) …the potential supply is thought to be huge.

P6: Nations like China, Argentina, Canada and U.S. each hold hundreds of trillion cubic feet of technically recoverable shale gas.

N7a: The combination of horizontal drilling and hydraulic fracturing, called fracking…

C7b: …has made unconventional gas supplies commercially feasible.

N8: Oil and gas companies have used fracking on vertical wells since the 1950s.

N9: Horizontal drilling or directional drilling, which increases the exposed length and surface of the gas reservoir by drilling through it at an angle, was first developed in the 1920s to improve efficiency.

N10a: The fracking method that is often discussed today only surged a decade ago.

N10b: …oil companies began fracking along horizontal wells, producing gas in dramatically larger quantities.
N11: The shale gas extraction process involves drilling the well vertically down to a maximum of 2.5 kilometres below the surface, then drilling a few hundred metres horizontally.

N12: The well is then cased with cement to protect ground water from contamination from fracking fluids.

N13a: The fracking fluids, which are water containing a small percentage of chemicals…

N13b: …are injected at a high pressure to create small fissures of about 1mm in the shale formation.

N14: …sand or solid, ceramic beads are pumped in to hold the fissures open in order to release the gas from the pores of the shale rock.

P15: The use of fracking and horizontal drilling to access shale gas is expected to lower natural gas prices worldwide.

P16: There is evidence that the price for natural gas is already starting to drop in the U.S.

P17: Increased gas extraction from fracking led to greater supply of gas in the U.S. market, and greater supply may have caused lower gas prices.

C18: When the price for natural gas is lower, it sometimes happens that the cost of drilling wells is greater than the value of the gas that is extracted.

P19: In the U.K., shale gas production could potentially attract almost £4 billion investment per year and create 74,000 jobs.

P20: In the U.S., it is estimated that shale gas could draw $71 billion in investment, and President Obama claimed that fracking is likely to support 600,000 jobs by the end of the decade.

P21: Taking Canada as an example, residents in New Brunswick, Nova Scotia and Quebec pay significantly more in provincial taxes than residents of the oil and gas producing provinces of Saskatchewan and Alberta.

P22: Treasuries of the latter provinces are less dependent on personal income taxes because they receive revenues from companies in the energy industry.

P23: In 2012–2013, the proportions of provincial revenue from oil production were 0.1% for Nova Scotia, 2.1% for New Brunswick, and 0.6% for Quebec.

P24: In contrast, the proportions for Alberta and Saskatchewan were 20% and 23% respectively.

P25: The expansion of shale gas in North America is transforming the natural gas market.

N26: For example, 75% of Pennsylvania’s everyday natural gas consumption is currently imported.
P27a: However, the Marcellus Shale site in the United States alone holds enough recoverable natural gas to not only meet Pennsylvania’s own needs…

P27b: …but to turn the U.S. into a significant exporter of energy.

P28: In 2012, the U.S. president announced that the country had enough gas supply for the next 100 years.

P29a: Since domestic production of shale gas enables nations like the U.S. to be independent from oil producing countries like Russia and Iran…

P29b: …these exporting countries will no longer be able to use oil and gas as political leverage on issues such as nuclear weapon testing.

P30: Natural gas burning emits 50% less greenhouse gas than coal fire generation.

P31a: It could greatly decrease the carbon footprint…

P31b: …and be a transitional energy source until cleaner renewable energies, such as wind, solar, and nuclear become more widely used.

P32a: Carbon emissions have reduced by 9% since 2005…

P32b: 50% of the reduction has been attributed to shale gas use.

P33: Additionally, shale gas mass production provides an important boost to the production of nitrogen-based fertilizers for agricultural purposes.

N34: People usually have favourable attitudes toward renewables due to environmental considerations.

P35a: They tend to overlook the cost effectiveness…

P35b: …and the fact that any kind of energy generation has certain impact on the environment and human life…

P35c: …wind power systems are associated with pollution caused by the use and mining of neodymium…

P35d: …solar powering requires very large amount of land…

P35e: …biofuel generation competes with food supplies over farmland use…

P35f: …hydro powering system can disrupt natural aquatic ecosystems.

P36: And the ever-growing nuclear powering runs the risk such as the Fukushima meltdown, which left radiation contaminations that will keep damaging the global environment and human health for decades and longer.

C37a: In recent years, much of the media attention, debates and conflicts have centred on fracking and its risks to the environment…

C37b: …and human health.
On October 18th, 2013, a protest led by Elsipogtog First Nation members in New Brunswick against a proposed hydraulic fracturing site on their land evolved into a violent encounter with the RCMP.

Soon the protest went nationwide.

Currently some states in the U.S., such as New York, and countries, such as Northern Ireland and France, have put moratoriums or even bans on fracking.

Hamilton ON, also passed resolution against fracking recently to protect their drinking water.

Fracking wellbores are drilled through drinking water sources in order to access the gas further below.

Well drilling activities may cause disturbance to groundwater quality.

Chemicals…

…natural gas may leak from the wellbore if it is not properly sealed and cased.

While there are requirements for well casing and integrity…

…accidents may occur, such as the explosion of a water well in Dimock, Pennsylvania.

Some groundwater contamination is demonstrated in YouTube videos that show kitchen tap water being set on fire in areas close to a fracking site.

There are also serious challenges associated with storing, transporting, treating, and disposing of fracking wastewater.

It must be processed at facilities…

…that are not always equipped to remove harmful contaminants.

Although the fluid used in fracking is mostly water, various kinds of chemicals are added to make fracking more effective.

Fracking also releases naturally occurring salts, metals, radioactive elements like barium and strontium and carcinogens like benzene into the wastewater.

These wastewater pollutants could potentially leak into the natural environment.

Shale gas extraction requires more water than traditional vertical wells, ranging between 3 to 5 million gallons of water per well.

Although the industry claims that most of the water can be recycled…

…depending on the data sources, it is reported that only 30%–80% of the water flows back to the surface after fracking.

This recovered water is unfit for further use before being treated for contaminants.
C55: Much of the recovered water is disposed directly in an underground wastewater well and not recycled.

C56: Such massive water consumption can divert water from other important uses such as agriculture, especially if the area is suffering from drought.

C57: During the extraction process, some of the methane gas may escape.

N58: Methane is a greenhouse gas that can cause global warming.

C59: Some believe that damage caused by the fugitive methane outweighs the environmental benefits of using natural gas.

C60: Some people who live near fracking sites complain of noxious fumes that cause headaches, nausea and other symptoms that they attribute to the chemicals released during fracking.

C61: The New York Times reported that parts of Texas where fracking is common have higher rates of asthma.

N62: However, these areas generally have high air pollution from sources other than fracking.


C64: Although it is not certain whether or not these earthquakes are caused by the fracking process, earthquakes have been linked to the wastewater disposal.

C65: U.K. experts believe that recent earthquakes in northwest England were caused by wastewater injection into a nearby fault zone.

C66: Research has identified key factors that cause the leakage from fracking, which is the main environmental concern.

N67a: In particular, the number of casings...

N67b: ...how these casings were cemented is an important factor.

N68a: Some of the leaky wells in Canada had only one casing...

N68b: ...uncased except in the section from the surface to just below the aquifer.

N69a: Others had not been cemented at all...

N69b: ...the cementation had not reached the required height.

N70: At a minimum, the cement should extend above exposed water or methane bearing zones.

P71: In some states, such as Pennsylvania and Texas, there is a requirement to cement casing of approximately 75 ft below any aquifers.

N72: Failure to do so can lead to groundwater contamination as occurred in Dimock.
N73: In the U.K., the standard practice is to have three casing with at least 2 casings passing through in order to isolate groundwater zones.
Appendix D.

Critical Thinking Items

Question 1. Lawyers are expected to treat what a client tells them as confidential and not reveal any information that could harm the client. It is possible, though, that a client may reveal to his lawyer that he intends to hurt someone, for example, a client might describe a plan to throw acid on the face of a spouse who is suing him for abuse. If the lawyer reports the client’s intention to the police the lawyer would be revealing confidential information. Such cases can pose a difficult ethical dilemma for lawyers because they pit the client’s right to confidentiality against the rights of a third party. Sometimes, there is no obvious solution to such dilemmas and reasonable people can disagree about the proper course of action. The reasoning presented in this question is:

- Good. Client confidentiality should always be preserved.
- Good. Depending on the circumstances, the decision to break client confidentiality is sometimes right, sometimes wrong, and sometimes too close to decide with certainty.
- Poor. There is no difficult ethical dilemma. The lawyer obviously must break confidentiality and stop the attack on the spouse.
- Poor. There is no difficult ethical dilemma. Protecting other people from harm should take precedence over preserving client confidentiality.

Question 2. In arguing for the legalization of marijuana, Tom said: “The reasons generally given for why marijuana should be illegal are that it is often a gateway drug leading to the use of more dangerous drugs and it has negative health effects. However, research has found using marijuana does not actually cause the use of more dangerous drugs.” Tom’s argument is:

- Good. Marijuana is a drug with few negative side effects.
- Good. Tom refers to research not just his own opinion.
- Poor. Tom only deals with one of the reasons for laws against marijuana.
- Poor. Tom does not explain the absence of a link between marijuana and more dangerous drugs.

Question 3. Jake and Tyson are two 17-year old classmates. Jake’s father had a gun that was not licensed and not secured in a locker to prevent misuse as is required by law. During an argument, Jake shot Tyson with his father’s gun. Jake’s father should be given a severe punishment for breaking the law. The rational for this conclusion is:

- Good. The illegal actions of the father were a factor in the shooting of Tyson.
- Good. Jake’s father should have taught his son about the proper handling of guns.
Poor. Even if the gun were licensed and secured, Jake could have found the key and used the gun to shoot his classmate.

Poor. It is Jake who pointed the gun at Tyson and pulled the trigger. His father had no intent to hurt Tyson.

Question 4. Because silver and gold are both precious metals, silver prices tend to follow the fluctuation of gold prices. If the demand of gold (such as in jewelleries) goes up, the price of gold will increase. It has been predicted that silver prices will rise dramatically to a point where they exceed gold prices because, in addition to being a precious metal used for investment and currency, silver is a product consumed in manufacturing, health and other industries. The justification given for the prediction is:

- Good. Past correlation can be a predictor of future trends. If the price of gold goes up, the price of silver probably will go up too.
- Good. More than one factor determines the price of silver.
- Poor. There is no consideration of whether gold might be consumed by manufacturing, health and related industries.
- Poor. The price of gold will go down because gold jewelleries are going out of style.

Question 5. In ancient times, pagan priests conducted special rituals at mid-winter to invoke the coming of spring. An observant villager had seen these rituals performed every year since he was a child. He decided that the rituals must indeed bring about the lengthening days of spring because every year after the rituals were performed the days stopped getting shorter and began to get longer. The villager’s thinking is:

- Good. He was doing the best he could with the limited information he had.
- Good. He recognized that days begin to get longer after the winter solstice.
- Poor. A priest’s rituals can have no effect on seasonal change.
- Poor. An association between two things does not mean that one necessarily causes the other.

Question 6. John owns a construction company. He said: “I am the owner, I decide whom I hire and whom I don’t. To ensure success of the company, all my employees are young, strong males because of the nature of their jobs. They are more efficient and less likely to get injuries doing their job than female workers and those who are less strong.” John’s reasoning is:

- Good. The company is his private property, and it’s reasonable that he uses strategies to make sure his investment gets the best return.
- Good. The law protects citizens’ properties, and other people have no right to interfere with the owner’s decisions.
- Poor. He is ignoring employment equity laws, and he may be ordered by a court to pay a large fine that could hinder the success of his company.
- Poor. Some women can be stronger than men.
Appendix E.

List of Main Ideas

M1: a) There are two types of natural gas: conventional and unconventional that b) has many uses, among which c) shale gas is a form of unconventional gas that is hard to obtain but exists in large quantity.

M2: a) The combination of horizontal drilling and hydraulic fracturing, called the fracking method for shale gas b) has a history of decades, c) which allows energy companies to access and produce large amount of unconventional gas.

M3: The fracking process involves a) vertical and horizontal drilling, b) using water and chemicals at high pressure, to release shale gas c) when protecting ground water from contaminations.

M4: Fracking shifts the demand and supply: a) due to the large quantity of gas produced by fracking, b) natural gas price are lowered, c) demonstrated by the market pattern in the U.S.

M5: Shale gas production could potentially a) attract investment and b) create jobs.

M6: a) Shale production could lower income tax b) due to revenues from energy companies, c) use one example.

M7: a) Shale production shifts the gas market and b) free energy dependent countries form oil generating countries for c) political leverage such as nuclear weapon testing.

M8: Shale gas can benefit the environment by a) having less emission, b) being a good transition to renewables, and c) boosting fertilizer production.

M9: a) While renewable sources of energy are generally favoured, b) they each have some problems, c) use one example.

M10: a) There are environmental and health concerns form the fracking process, b) consequently people developed negative attitudes toward fracking, c) therefore some states and countries restrict or ban fracking.

M11: a) Fracking process could cause ground water contamination (mention any of: disturbance, chemical, gas leaks), b) if not properly sealed, c) even under regulation, therefore hard to protect.

M12: a) Wastewater poses risks, b) due to the chemicals added and possible leakage to the environment, but c) wastewater is often mishandled because facilities are unequipped.

M13: a) Fracking consumes massive amount of water, b) which is not always recycled, c) thus competing with other important water needs.
M14: a) The fracking process could leak fugitive methane, b) which is a threat to the environment, c) and human health.

M15: a) Fracking has been linked to earthquakes, b) although no hard evidence available on the correlation, c) other than earthquakes and the disposal of wastewater into fault zones.

M16: a) There are known issues with fracking such as leakage that b) countries are trying to address with standards and regulations in order to lower such risks, c) but no agreements so far and not all countries are enforcing the regulations to the same extent.
Appendix F.

Argument Markers

Additionally
Agree/Disagree
All in all
Although/Even though
Analysis/Analyze/Analyzing
Argue/Argument
As a result
Back up
Because
Besides
But
Claim
Compare/Contrast
Conclude/Conclusion
Consequently
Controversial/Controversy
Data/Evidence
Debate
Despite
Divisive
Due to
Evaluate
Finally
First of all
First, second, third… /Firstly, secondly…
Fist and foremost
Furthermore
Hence
However
In fact
In turn
Including/include
Meanwhile
On one hand…On the other hand
On the plus side
Outweigh/Outweighing
Proof
Proven/Prove
Reason/Reasons
Show/Shown
Similarly
Since
Support/Supporting
Then
Therefore
Thus
To top it off
Trade off
Weight/Weighted
Whereas
While
Why
Appendix G.

Cued-Recall Questions with Scoring Rubrics

How does the combination of horizontal drilling and hydraulic fracturing make unconventional gas supplies commercially feasible?

The correct answer should include:
Main Ideas: #1, #2, #3, #4, #5, #6, and #7.

Idea Units: N2a, N2b, N3a, C3b, N4, N5a, C5b, P5c, N7a, P7b, N9, N10b, N11, N13a, N13b, N14, P17.

What are the economic implications of fracking?

The correct answer should include:
Main Ideas: #4, #5, #6, #7, #10, #11, #12, and #14.

Idea Units: P15, P16, P17, C18, P19, P20, P21, P22, P23, P24, P25, P27a, P27b, P34, C37a, C37b, C38, C39, C40, C41, N42, C43, C44a, C44b, N45a, C45b, C46, C47, N48a, C48b, C49, C50, C51, C52, N53a, C53b, C54, C55, C56, C57, N58, C59, C60, C61, N62.

What does wastewater processing involve?

The correct answer should include:
Main Ideas: #11, #12, and #13.

Idea Units: C47, C48a, C48b, C49, C50, C51, C52, N53a, C53b, C54, C55, C64.

What is the government’s role in fracking?

The correct answer should include:
Main Ideas: #6, #10, #11, #12, #14, and #16.

Idea Units: N12, P24, C38, C40, N48a, N67a, N67b, N70, N71, N73.

What are the standards to minimize environmental negative impact by fracking?

The correct answer should include:
Main Ideas: #12 and #16.

Idea Units: N12, C44b, N48a, N67a, N70, N71, N73.
Appendix H.

Reasoning Questions

1. How is fracking connected to global warming?

2. If you worked for an energy firm that was advertising to gain public acceptance of fracking, how would you present the advantages of fracking?

3. How do you think fracking could affect YOUR life and WHY?

4. If you were appointed by your local government to investigate possible fracking sites proposed by energy companies, what aspects or issues would you focus on? What policies and decisions would you make regarding fracking and why?

5. Explain your opinion on whether or not companies should be allowed to expand oil pipelines in BC that transport crude oil from Alberta?