

**THE EFFECTS OF ARGUMENTATION GOALS ON  
RECALL, CAUSAL UNDERSTANDING AND REASONING**

by

Liuqing Mao  
B.A., Zhejiang University, 2005

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

In the  
Faculty  
of  
Education

© Liuqing Mao 2009

SIMON FRASER UNIVERSITY

Summer 2009

All rights reserved. This work may not be  
reproduced in whole or in part, by photocopy  
or other means, without permission of the author.

# APPROVAL

**Name:** **Liuqing Mao**  
**Degree:** **Master of Arts**  
**Title of Thesis:** **The effects of argumentation goals on recall, causal understanding and reasoning**

**Examining Committee:**

**Chair:**

---

**Dr. Lucy Le Mare**  
Associate Professor of Faculty of Education

---

**Dr. John Nesbit**  
Senior Supervisor  
Associate Professor of Faculty of Education

---

**Dr. Philip Winne**  
Supervisor  
Professor of Faculty of Education

---

**Dr. Kevin O'Neil**  
External Examiner  
Associate Professor of Faculty of Education

**Date Defended/Approved:** June-15-2009



SIMON FRASER UNIVERSITY  
LIBRARY

## Declaration of Partial Copyright Licence

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection (currently available to the public at the "Institutional Repository" link of the SFU Library website <[www.lib.sfu.ca](http://www.lib.sfu.ca)> at: <<http://ir.lib.sfu.ca/handle/1892/112>>) and, without changing the content, to translate the thesis/project or extended essays, if technically possible, to any medium or format for the purpose of preservation of the digital work.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author's written permission.

Permission for public performance, or limited permission for private scholarly use, of any multimedia materials forming part of this work, may have been granted by the author. This information may be found on the separately catalogued multimedia material and in the signed Partial Copyright Licence.

While licensing SFU to permit the above uses, the author retains copyright in the thesis, project or extended essays, including the right to change the work for subsequent purposes, including editing and publishing the work in whole or in part, and licensing other parties, as the author may desire.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Simon Fraser University Library  
Burnaby, BC, Canada



SIMON FRASER UNIVERSITY  
THINKING OF THE WORLD

## STATEMENT OF ETHICS APPROVAL

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, in a research project approved in advance,

or

(d) as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at 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.

Bennett Library  
Simon Fraser University  
Burnaby, BC, Canada

## **ABSTRACT**

This research examined the effects of instructing students to use web-based text to prepare for an argument on their recall of the text information, understanding of causal relations in the text, and reasoning in an evaluation essay about theories presented in the text. Instruction to argue was compared with instructions to use text to prepare a summary or study for a recall test. A sample of 120 participants was randomly assigned into three groups: argument group, summary group and study group. The experiment also investigated the relationship of need for cognition to reasoning. The results indicated that the argument group outperformed the other two groups in the quality of reasoning, but did not produce better recall or causal understanding. Need for cognition also predicted the quality of reasoning. These findings have important implications for classroom deployment of argumentation as a learning strategy.

**Keywords:** argumentation, argument schema, learning, goals, need for cognition

## **DEDICATION**

*To my husband, Hangjun*

## **ACKNOWLEDGEMENTS**

I would like to take this opportunity to express my profound thanks to all the people who have helped make this research possible. First, I would like to express my sincere gratitude and thanks to my senior supervisor Dr. John C. Nesbit for his consistent patience and guidance to support me through this period. Thank you for always pushing me to do my best work. I would also like to thank Dr. Phil H. Winne for being part of my supervisory committee and supportive for my work.

I would also like to thank Mr Rylan Egan for his precious advice during the experiment design and data analysis, and generous help in data coding.

Last, but certainly not least, I would like to express my special thanks to my husband, Hangjun. I would not have made all this through without your encouragement and support. Thank you for being always so patient and so supportive with me.

# TABLE OF CONTENTS

<b>Approval .....</b>	<b>ii</b>
<b>Abstract.....</b>	<b>iii</b>
<b>Dedication .....</b>	<b>iv</b>
<b>Acknowledgements.....</b>	<b>v</b>
<b>Table of Contents .....</b>	<b>vi</b>
<b>List of Figures.....</b>	<b>viii</b>
<b>List of Tables .....</b>	<b>ix</b>
<b>CHAPTER 1: OUTLINE OF THE THESIS .....</b>	<b>1</b>
1.1    Rationale.....	1
1.2    Hypotheses.....	2
1.3    Findings .....	4
1.4    Structure of the Thesis.....	5
<b>CHAPTER 2: LITERATURE REVIEW.....</b>	<b>7</b>
2.1    Learning from Texts .....	7
2.2    Argument and Summary Schemas.....	10
2.3    The Impact of Argument Schemas on Learning.....	11
2.4    What is Argumentation.....	12
2.5    Four Reasons Why Argumentation May Foster Learning.....	13
2.6    The Effect of Giving Argumentation Goal Instructions .....	15
2.7    Argumentation as a Learning Strategy .....	17
2.7.1    The Effect of Reading Arguments on Learning.....	17
2.7.2    The Effect of Arguing to Learn .....	18
2.8    Need for Cognition .....	21
<b>CHAPTER 3: METHOD.....</b>	<b>23</b>
3.1    Pilot Study .....	23
3.2    Participants .....	23
3.3    Materials and Instruments.....	24
3.3.1    Demographics Questionnaire.....	24
3.3.2    Reading Materials .....	24
3.3.3    Need for Cognition Instrument .....	25
3.3.4    Instructions.....	25
3.3.5    Outcome Tasks .....	26
3.3.6    nStudy .....	27
3.4    Procedure .....	29
3.5    Latent Semantic Analysis .....	31



<b>CHAPTER 4: RESULTS.....</b>	<b>33</b>
4.1 Overview of the Types of Data Collected .....	33
4.2 Scoring Free Recall Responses.....	33
4.3 Scoring Essays.....	34
4.4 Analysis and Results.....	38
4.4.1 Outliers and Normality .....	38
4.4.2 Demographic Data and Need for Cognition .....	38
4.4.3 Log Data Analysis .....	39
4.4.4 The Analyses of Outcome Tests .....	43
<b>CHAPTER 5: DISCUSSIONS .....</b>	<b>55</b>
5.1 Discussion of the Results.....	55
5.2 Implications .....	57
5.3 Limitations and Future Work.....	58
<b>Reference List.....</b>	<b>61</b>
<b>Appendices.....</b>	<b>69</b>
Appendix 1 .....	70
Appendix 2 .....	71
Appendix 3 .....	72
Appendix 4 .....	73
Appendix 5 .....	74
Appendix 6 .....	75
Appendix 7 .....	78
Appendix 8 .....	79

## LIST OF FIGURES

Figure 1	nStudy tagging feature.....	29
Figure 2	Inter-rater reliability on all variables of the essays.....	37
Figure 3	Number of study tags, summary tags and argument tags in three groups .....	41
Figure 4	Proportions of actions containing study tags, summary tags and argument tags in three groups.....	42

## LIST OF TABLES

Table 1	Gender, English experience, education level, age and need for cognition .....	39
Table 2	Mean number of actions (hyperlink clicking and text tagging) in three groups .....	39
Table 3	One-way ANOVA of proportions of actions containing study tags, summary tags and argument tags across three groups.....	43
Table 4	Descriptive statistics of the multiple-choice test in three groups .....	44
Table 5	One-way ANOVA test of the multiple-choice test in three groups.....	44
Table 6	Descriptive statistics of the free recall test in three groups .....	44
Table 7	Descriptive statistics of word count of the free recall test in three groups .....	45
Table 8	Descriptive statistics of some important variables of the essays in three groups .....	46
Table 9	One-way ANOVA of some important variables of the essays.....	<b>47</b>
Table 10	Multiple comparisons of some important variables of the essays across three groups .....	48
Table 11	Multiple regression using NFG, group, actions and argument tags to predict the number of reasons.....	51
Table 12	Multiple regression using NFG, group, actions and argument tags to predict the number of correct and strong reasons.....	52
Table 13	Multiple regression using group, education level and English experience to predict the number of reasons .....	53
Table 14	Multiple regression using group, education level and English experience to predict the number of correct and strong reasons .....	53

# **CHAPTER 1:OUTLINE OF THE THESIS**

## **1.1 Rationale**

Many studies have indicated that active processing of text can promote better understanding of subject matter (Kintsch & Young, 1984; Mannes & Kintsch, 1987; McDaniel & Donnelly, 1996; McNamara, Kintsch, Songer & Kintsch, 1996). The research suggests that if students are assigned a writing task that requires deeper cognitive processing while reading a text, they learn more than if they are assigned other types of writing. Argumentation has been identified as an important means of improving students' understanding of subject matter in the classroom (e.g., Baker, 2003; DeVries, Lund, & Baker, 2002; Duschl & Osborne, 2002; Mason, 2001; Nussbaum & Sinatra, 2003; Schwarz, 2003; Schwarz & Glassner, 2003; Schwarz, Neuman, & Biezuner, 2000), because it engages students in constructive cognitive processing of information (Baker, 1999; Nussbaum & Sinatra, 2003). An implication of this idea is that instructing students to study text to prepare for constructing a written argument may induce a high level of cognitive engagement with the text content and, compared with other reading goals, lead to enhanced learning outcomes.

This research reported in this thesis compared learning outcomes of instructing students to use web-based text and images to prepare for (a) writing an argument, (b) writing a summary and (c) taking a recall test. The assessed learning outcomes were students' ability to recall text information, understand causal relations and develop critical analyses.

The study will also examine the effect of need for cognition as a factor that influences student cognitive engagement (Cacioppo et al., 1996). Need for cognition has been found to be related to efficiency in processing argumentative texts (Kardash & Scholes, 1996) and attraction to argumentative situations (Nussbaum & Bendixen, 2003). It is likely need for cognition would have a positive effect on the quantity and quality of reasons generated in evaluation essays.

Web-based technology has been widely used in research to aid the investigation of users' behaviors. Web-based tools can capture log data that enrich researchers' datasets and facilitate detailed analysis of student behavior. This research used a web-based tool called nStudy (Winne & Hadwin, 2009) to track participants' text tagging and hyperlink clicking actions as they studied text and images.

## **1.2 Hypotheses**

The framework for the hypotheses made in this study can be summarized in four ideas: (1) Individual differences in need for cognition, which is self-reported preference for cognitive activities, will influence participants' critical understanding. (2) Participants' performance on a recall test will not be differently affected by the instructions to argue, summarize or study. (3) Participants instructed to pursue an argumentation goal will show greater critical understanding of the material even after the impact of need for cognition is considered. (4) Regardless of different assigned instructions, the adoption of an argument schema which may be indicated by their use of argument tags in nStudy will also have a positive effect on their critical understanding.

It has been shown that text processing that is presumed to be more cognitively engaging does not always produce better recall (Kintsch & Young, 1984; Mannes & Kintsch, 1987; McDaniel & Donnelly, 1996; McNamara, Kintsch, Songer & Kintsch, 1996). There is evidence that this may be true for argumentative writing tasks (Wiley & Voss, 1999). Because recall tests require relatively little knowledge of inferences and implications that follow from presented information, it seems likely that instructing students to prepare to argue would have no effect on free recall. However, its effect on understanding inferences and implications may be educationally significant because it engages students in a deeper level of cognitive processing (Baker, 1999; Nussbaum & Sinatra, 2003).

In addition, the analysis of nStudy log data such as actions and uses of tags will further our understanding of what students did as they worked with the web-based content. The number of actions can reflect the overall level of effort students make while reading the text, which may influence their following performance. Also, because a variety of argument, summary and study tags were made available in all treatment conditions, the participants' actual adoption of the three task conditions can be tracked.

This research examined the following hypotheses:

Hypothesis 1: Students with a high need for cognition will generate higher quality of evaluation essays than students with a low need for cognition.

Hypothesis 2: Students who are instructed to argue will perform as well as students who are instructed to summarize and study in free recall test.

Hypothesis 3: Students who are instructed to argue will perform better than students who are instructed to summarize and study in understanding of causal relations.

Hypothesis 4: Students who are instructed to argue will generate a higher quality of evaluation essays than students who are instructed to summarize and study.

Hypothesis 5: Students who have more nStudy actions will produce a higher quality evaluation essays.

Hypothesis 6: Students who used more argue tags will produce higher quality evaluation essays.

### **1.3 Findings**

Participants consisted of 120 undergraduates and graduates from Simon Fraser University. They were randomly assigned to three treatment conditions which consisted of instructions to use web-based text to (a) prepare to write an argument, (b) prepare to write a summary or (c) study for a recall test. The three groups were then given a free recall test, a multiple-choice test of causal relations in the text and an evaluation essay about theories presented in the text.

Analyses of the use of tags indicated that the three groups did tend to adopt the goals assigned in the instructions they received. No differences between the three groups were statistically detected for the free recall test and the multiple-choice test of causal relations. Relative to the study and summary groups, the argument group generated many more reasons and logically sound reasons in their evaluation essays. Furthermore, a

regression analyses revealed that need for cognition, number of nStudy actions, and use of nStudy argument tags did positively predict the quality of evaluation essays.

For teachers, these results offer evidence that students' ability to reasoning about content can be promoted by asking them to use text to prepare an argument. Such argumentation interventions can also be naturally combined with other learning activities to induce and motivate students to learn more deeply.

## **1.4 Structure of the Thesis**

Chapter 2 reviews research literature on learning from texts, gist extraction in fuzzy-trace theory, schema theory, the effect of an argument schema and argumentation on learning, and studies that have been done on setting argumentation goals, reading arguments and arguing to learn. The theories of learning from text and gist extraction are reviewed with the purpose of contrasting different ways students learn from texts when they argue, summarize and study for recall. Schema theory is discussed with consideration for how an argument schema and argumentation may foster learning. Research on giving argumentation goals is reviewed because it relates to the instructions to argue used in this study. Finally, the chapter examines previous studies which investigated argumentation as a learning strategy, especially those involving students in argumentative situations such as argumentative writing. The differences between my research and that of Wiley and Voss (1999) are discussed.



Chapter 3 describes the design of the pilot study and main experiment. It provides a detailed account of the participants, learning materials, instruments, experimental procedures, and the use of latent semantic analysis for scoring the free recall responses.

Chapter 4 reports six types of data collected in the experiment including demographic data such as education level and English experience, need for cognition, log data of the number of actions and use of tags, and data for the three outcome tests. It describes descriptive statistics and analysis of variance (ANOVA) tests on the collected data. Multiple regression analyses was also performed to identify factors related to essay reasoning.

In Chapter 5, I interpret the observed results, and consider their implications. In that chapter I also examine the limitations of the research and outline productive areas of future research.

## CHAPTER 2:LITERATURE REVIEW

### 2.1 Learning from Texts

Researchers distinguish between different learning outcomes produced by reading text (Kintsch, 1994). Literal memory for a text is an encoding of the linguistic relations of the text and its semantic and rhetoric structure, which is regarded as a superficial understanding of the text. In contrast, a deeper level of understanding is acquired if prior knowledge is activated and then integrated with the incoming information. The result of literal remembering is that the reader is able to reproduce the text in some form, while the result of integrating new information with prior knowledge is a higher level of understanding which enables the reader to use the information provided by the text in a new way such as problem solving. According to the Kintsch's model (1988, 1992), people who can recall a text have constructed a *textbase*, which by itself allows only a superficial level of understanding. A deeper understanding depends on constructing a *situation model* in which the new information is embedded in a dense network of related propositions.

The structure of the text has an effect on how easily learners can build a *textbase* or a *situation model*. Students are more likely to construct a good *textbase* if a summary sentence is placed at the beginning of a paragraph they are assigned to read (Kintsch & Wipond, 1979; Weaver & Kintsch, 1990). However, the form of the text that is good for a superficial level of text-based learning is not necessarily good for a deeper level of situational learning. Kintsch (1994) argued that deeper learning can be promoted by

inducing readers to more actively participate in comprehension. Some studies have found that making texts that are less clear can benefit deep learning and understanding. For instance, McNamara, Kintsch, Songer, and Kintsch (1996) found that readers performed better on both inference and problem solving tasks after being presented with a text having lower coherence. Readers presented with a high coherence text did not perform as well. A mismatch between an introductory outline and subsequent text produced better results in problem solving compared to a well-matched one (Mannes & Kintsch, 1987). Better understanding was also found when embedding an analogy in text (McDaniel & Donnelly, 1996). One can conclude that if writing tasks encourage students to construct a situation model, it is likely that they will have a better understanding.

Kintsch's model predicts that asking students only to memorize a text will not promote situation model construction. Summarization, however, is expected to demand deeper cognitive processing. Summaries are expressions of the macrostructure of a text (Dijk & Kintsch, 1983), which are interpreted by an individual in the context of his or her background knowledge. To make a summary, a reader is required to discern and emphasize central ideas and minimize less relevant details. Theoretically, summarization involves understanding the microstructure of a text, generalizing and condensing the microstructure into a set of macropropositions, and finally expressing the macrostructure in a coherent text (Dijk & Kintsch, 1983). These characteristics of summarization can help readers better understand the text itself, but deep learning does not happen until they understand something that is not made explicit in the text. Therefore, a task that leads readers to make inferences about the underlying meaning of a text is needed to facilitate construction of a situation model. Different from simply extracting central idea for

summarization, argumentation involves more complicated activities such as understanding alternative views, comparing competing arguments and refuting opposing arguments. The advantages of argumentation on learning will be explained in detail later. Giving students an argument task might be expected to engage them in processes that explore the text at a deeper level than simply understanding the explicitly-presented text content, and constructing generalizations of that content.

Reyna and Brainerd's (1995) fuzzy-trace theory offers an alternative account of the type of knowledge learners construct as they read text. According to fuzzy-trace theory, there are two types of memory, verbatim memory and gist memory. Verbatim memory retrieves surface forms of a text such as specific details, while gist memory retrieves inferred features and meanings of a text. In addition to the distinction between verbatim and gist memory, there are two levels of gist memory, local gist and global gist (Neuschatz, Lampinen, Preston, Hawkins, & Togliani 2002). Local gist represents discrete pieces of inferred information, while global gist constructs the whole meaning of an entire text by representing relations among the discrete pieces. Global gist is a higher level of gist representation. It has been found that global gist processing has a greater influence on adults in generating false memory (Lampinen, Leding, Reed & Odegard, 2006), which is erroneous recollection of non-presented critical information. Brainerd and Reyna (1992) argued that it is gist traces, rather than verbatim inputs, that tend to be processed in reasoning. That is to say, the performance of reasoning is highly dependent on the quality of gist memory, especially global gist for adults.

It is likely that some reading tasks will facilitate gist extraction better than others. Studying for a recall test probably focuses students' attention on detailed pieces of

information, thus increasing local gist but not global gist extraction. Summarization requires students to construct a macrostructure representation of a text that consists of generalizations and abstractions made by linking discrete propositions and deleting trivia. This process focuses students' attention on the relations among small pieces of information and thus helps global gist extraction. An argumentation task may also promote global gist extraction, but in a different way from summarization tasks. To construct an argument in favor of one position, students will focus their attention on relations between positions and their supporting evidence. Consequently, the global gist extracted through an argumentation task may better facilitate reasoning than that extracted through a summarization task.

## **2.2 Argument and Summary Schemas**

Decades ago, schemas were conceived as theoretical cognitive structures to explain how humans represent and process information (e.g. Anderson & Pearson, 1984; Ausubel, 1963; Bartlett, 1932). Anderson and Pearson (1984) defined a schema as an abstract structure that can connect semantically related pieces of information. They are thought to serve many functions such as allocation of cognitive resources (West, Farmer, & Wolff, 1991), reconstruction of knowledge (Anderson, 1977), and facilitation of remembering (Rumelhart, 1980) and learning (Anderson, 1984).

In this study, the summary group will be given a written description of a summary and told to prepare to write a summary, while the argument group will be given a written description of the components of an argument and told to prepare to write an argument.

The purpose was to prompt them activate a summary or an argument schema respectively. Researchers on argumentation and reasoning have started to use the concept of schema (e.g. Chinn & Anderson, 1998; Politzer & Nguyen-Xuan, 1992; Scardamalia & Bereiter, 1986). According to Chinn & Anderson (1998), an argument schema is a network that connects argument components to represent an extended argumentative discourse. Basic argument components are a claim and a supporting reason (Angell, 1964). A complex argument includes claims, multiple reasons, counterarguments and rebuttals.

Summarization has long been investigated as a strategy for improving reading comprehension (Bean & Streenwyk, 1984; Jenkin, Heliotis, Stein & Haynes, 1987; Rinehart Stahl & Erickson, 1986). The summary schema may contribute to the beneficial effect of summarization on text comprehension by serving as guide or standard for the construction of a *textbase* or global gist. The summary schema presumably connects abstractions and generalization which are the central ideas of a text.

### **2.3 The Impact of Argument Schemas on Learning**

Reznitskaya and Anderson (2002) hypothesized that an advanced argument schema may serve a variety of functions. First, an argument schema will influence the processing of incoming information. Readers, writers or speakers with a developed argument schema will give their attention to argument-relevant information and then organize the information in a way to help them build an argument.

Then, an argument schema will help in understanding and constructing of an argumentative discourse. It is quite likely that students equipped with a strong argument

schema will have better comprehension of the presented arguments because they are able to relate individual arguments within a discourse. Moreover, argument building and repair can be facilitated by making use of the “slots” in the activated schema.

Next, having an argument schema may help retrieval of argument-relevant information from memory, because as previously discussed, an argument schema connects and relates argument-relevant information and stores it systematically in memory.

Finally, students with an argument schema may tend to be more critical of the discussed issue. Once an argument schema is activated, the students will be encouraged to compare and evaluate arguments and then find flaws in the arguments from different sides.

## **2.4 What is Argumentation**

Although the two terms “argumentation” and “argument” are sometimes used interchangeably in everyday language, they are given more restricted meanings in the argumentation research. Argumentation refers to the *process* of arguing around a specific theme in either a written form or a dialogue form (Andriessen, Baker, Suthers, 2003). In the process of argumentation people construct a series of arguments which provide evidence to support or oppose a point of view.

Argumentation is a complex activity. In a formal educational environment, students may be involved in various argumentation situations. Argumentation can be either an individual or a group activity. Argumentation is an individual activity when

students are assigned to read an argumentative text or write an argumentative essay. Individual argumentative reading requires the evaluation of competing positions and comparison of arguments. Argumentative writing is a more complicated activity because it has to deal with the addressee (Andriessen, Erkens, van de Laak, Peters & Coirier, 2003). Besides constructing reasons supporting one's position, an arguer should consider possible counterarguments put forward by the addressee and refute them with rebuttals. Unlike reading and writing by oneself, interactive argumentation engages students in interacting with one another (Nussbaum, 2005). In this process, students need to respond to the questioning of other students by making arguments and rebuttals to defend their own position. Sometimes they also need to work together to construct or critique arguments.

## **2.5 Four Reasons Why Argumentation May Foster Learning**

Why do some educational researchers believe that argumentation has potential to foster learning? First, during the process of argumentation students are confronted with claims that are contradictory and collectively induce cognitive conflict. To build an argument in support of a position, cognitive conflicts are created when writers consider the objections to their theory put forward by the adversary. Cognitive conflict is thought to be a trigger for conceptual change (Limon, 2001). Therefore, argumentation creates conditions for conceptual change.

Second, similar to the “self explanation effect” (Chi et al., 1989), the production of arguments could itself lead to reflection and re-examination of one's belief, thus



promoting understanding and knowledge restructuring (e.g. Chi, 2000; Kuhn, 1991; Neuman & Schwarz, 2000). The process of counter argumentation, which refutes opposing arguments by presenting reasons against them, may further foster conceptual understanding by broadening the learners' understanding of the web of relationships surrounding an issue.

Third, argumentation inspires high engagement from students in that it encourages mastery goal adoption, and gains and sustains their attention. Students with a mastery goal are more cognitively engaged than students with a performance goal in that they focus on the improvement of learning and understanding of materials instead of competing with others (Linnenbrink & Pintrich, 2002). Achievement goal researchers have claimed that different classroom environments may influence students' goal adoption (e.g. Ames, 1992; Blumenfeld, 1992). Reducing authority and discouraging competition and social comparison among students can promote mastery goals (Ames, 1992). According to Andriessen (2006), argumentation is not oppositional and aggressive but collaborative with the purpose to reach agreement. Individual argumentation is also collaborative in the sense of the need to consider alternative views. Thus, creating an argumentative situation in the classroom may encourage students' adoption of mastery rather than performance goals by having them considering an alternative view and possible supporting evidence for it.

An argumentative situation is able to gain and sustain students' attention. The reason is that, in contrary to explanation, arguing happens before the results are known (Nussbaum & Sinatra, 2003). That is, arguments are built on uncertain facts (Govier, 1987). Uncertainty as one characteristic of argumentation is able to motivate students by

maintaining their attention on the issue being argued (Keller, 1987). The potential of argumentation in promoting high engagement has been confirmed by several studies (Hatano & Inagaki, 1991, Herrenkohl & Guerra, 1998; Mason & Santi, 1994).

Fourth, it is thought that argumentation can contribute to critical thinking skills (Kuhn, 1992; Means & Voss, 1996). Kuhn (1992) argued that argumentation underlies peoples' every day higher order thinking and reasoning. Means and Voss (1996) also pointed out that informal reasoning requires skills in using the structures of argumentation, that is, arguments, counterarguments, rebuttals and so forth. Therefore, the development of students' argument schemas may promote their reasoning ability, which itself includes argumentation. It also explains why the development of ability to argue is so important. Our review (Nesbit, Niu & Mao, 2008) showed that there is an increasing number of studies featuring instructional activities to teach students to argue (e.g. Auriac-Peyronnet, 2001; Marttunen, 1997; Marttunen & Laurinen, 2001).

## **2.6 The Effect of Giving Argumentation Goal Instructions**

Goal setting plays an important role in writing (Hayes, 1996; Hayes & Flower, 1980; Hayes & Nash, 1996). Researchers (Ferretti, MacArthur, & Dowdy, 2000; Hayes & Flower, 1980; Scardamalia & Bereiter, 1986) found that a skilled writer devotes considerable effort to setting goals while reading in preparation for writing. Ferretti et al. (2000) argued that the poor quality often observed in students' argumentative writing is due to their inability to set appropriate goals. Better goal setting may therefore be crucial to improving the quality of argumentative writing.

Furthermore, researchers (Ferretti, MacArthur, & Dowdy, 2000; Page-Voth & Graham, 1999) argued that a specific goal is more effective than a general goal in writing because a specific goal provides students more guidance by directing their attention and efforts to particular, important aspects of writing. For argumentative writing, a number of studies have demonstrated that specific argumentation goals have a positive effect on the quality of essays. Ferretti et al. (2000) found that asking students to generate supporting reasons, counterarguments and rebuttals led to generating a greater number of argumentative elements in their essays than simply asking them to persuade, especially among six-graders, but not among fourth-graders. The reason is that specific goals helped students set appropriate subgoals, for example knowing they need to create rebuttals. This effect was replicated in the Nussbaum and Kardash study (2005) involving college students. Different from the Ferretti et al study, Nussbaum and Kardash examined the separate effect of a goal to generate reasons compared to a goal to generate counterarguments and rebuttals. They found that the counterargument condition did produce more counterarguments and rebuttals, but the reason condition did not generate significantly more reasons. Nussbaum and Schraw (2007) found that a goal instruction emphasizing argument-counterargument integration produced essays higher in holistic quality and integration scores. These investigations show that giving students an argumentative goal including argument elements is effective in increasing the number of the argument elements and promoting higher quality essays.

## **2.7 Argumentation as a Learning Strategy**

As we have discussed, argumentation has potential to foster student learning. Argumentative activities have been used by many researchers (Duschl & Osborne, 2002; Schwarz, 2003; Schwarz & Glassner, 2003; Schwarz, Neuman & Biezuner, 2000) as a way to promote student understanding of varied subject matters in classrooms. However, the investigation of argumentation as a learning strategy is still at an early stage. A review conducted by Nesbit, Niu and Mao (2008) found that very little research has been done on the effect of arguing on learning and most of the research has focused on the impact of argument reading on learning. It seems likely that the full effects of argumentation have thus far not been completely revealed.

### **2.7.1 The Effect of Reading Arguments on Learning**

Our review (Nesbit, Niu & Mao, 2008) indicated that almost all educational research on reading arguments has focused on the effect of reading refutational text. Other styles of argumentative text are seldom examined. A refutational text confronts students with scientific conceptions while at the same time acknowledging and refuting alternative conceptions (i.e., misconceptions). Refutational text addresses misconceptions and promotes conceptual change (Salisbury, Glennon, & Stevens, 1999).

Studies have found that reading a refutational text has several benefits compared to reading expository text for all levels of students (Alvermann, Hynd, & Quian, 1995; Guzzetti, Williams, Skeels, & Wu, 1997; Hynd, Alvermann, & Quian, 1997; Maria & MacGinitie, 1987). On a test of learning and conceptual change in science education, students who read a refutational text outperformed those who read an expository text (Diakidoy, Kendeou, & Ioannides, 2003). Salisbury-Glennon and Stevens (1999) found

that reading refutational texts resulted in better retention and facilitated conceptual change.

A review conducted by Hynd (2001) revealed that key features of refutational texts that induce conceptual change are moderate discrepancy with belief, understandability, credibility, usefulness, repetitiveness, and relatedness. Students found refutational texts are credible because the conflict between their prior knowledge and the scientific conception is explicitly presented (Guzzetti et al., 1997).

### **2.7.2 The Effect of Arguing to Learn**

Some researchers have started to investigate argumentation interventions such as writing arguments (Wiley & Voss, 1999) or participating in argumentative discourse (Asterhan & Schwarz, 2007; Nussbaum & Sinatra, 2003). Nussbaum and Sinatra (2003) examined the effects of argumentation on conceptual change. In their study, participants were asked to predict the path of a falling object dropped in different contexts and then explain their choices. After each prediction, the experimental group were instructed to argue in favour of the opposite side (which unbeknown to them, was the correct solution), while the control group were not asked to argue for an alternative path. They found that students in the experimental condition had significantly higher quality of explanations than those in the control condition, although they were not more accurate in making scientific predications in the domain. However, because the researchers provided prompts to help with student reasoning, it is not clear whether it is the arguments or prompts that are responsible for the improved outcomes. Asterhan and Schwarz's study (2007) confirmed their finding on the effects of argumentation intervention. They indicated that

both dialogical and monological argumentation facilitated by a confederate improved student conceptual understanding.

Researchers are also interested in the effects of argumentative writing. In their study, Wiley and Voss (1999) examined the impact of an argument writing task on student conceptual understanding of subject matter in comparison to three other writing tasks (narrative, summary and explanation). All participants were assigned to read multiple-source texts from either a web-based environment or textbook-like passages. Their analysis suggested that argument writing tasks not only produced more transformed, integrated and causal essays compared to narrative writing tasks, but also generated better conceptual understanding. However, the argumentation intervention did not yield better retention of the text information. Furthermore, Wiley and Voss indicated that the benefits of argument writing are more evident in a web-based, multiple-source environment. Writing summaries produced the same outcomes as writing narratives except that summary writers showed better integration of information. Writing explanations produced similar outcomes as writing arguments, but only in the single source condition.

The research reported in this thesis examined the effect of instructions to build an argument on students' recall of text content, understanding of causal relations in the text and quality of critical essays. It compared students instructed to prepare to (a) write an argument, (b) write a summary or (c) take recall a test. This study expands on Wiley and Voss's study (1999) in several aspects. First, the two studies are different in their treatment and dependent measures. In the Wiley and Voss study, participants were assigned to write a narrative, a summary, an explanation or an argument after finishing

reading, while my study instructed participants to use text to prepare but not actually write a summary or an argument. Another difference is the dependent measures. The dependent measures in Wiley and Voss's study were: a sentence verification task, an inference task and a principle identification task, although most of their results section focused on the analyses of the process variables, that is, the analyses of students' writing. In my study, all participants produced the same type of essay, an evaluation essay. In my study, writing the essay served as a dependent measure rather than as a treatment.

Second, my study used a web-based tool, nStudy (Winne & Hadwin, 2009), which can capture actions of the participants such as their use of tags and hyperlink clicking. The use of nStudy provided more data resources and helped me to explain students' learning by looking at their nStudy actions. Consequently it offered a better opportunity to understand the results.

Third, learner characteristics were examined as factors that may influence student achievement. Need for cognition is one factor which reflects a disposition to engage in cognitive activities (Cacioppo, et al., 1996). Nussbaum (2005) found that need for cognition had a positive effect on the number and quality of arguments. Other learner characteristics assessed in this research were level of English proficiency, level of education and major.

Fourth, this experiment was conducted under relatively well-controlled laboratory conditions, whereas Wiley and Voss's experiment was conducted in a classroom. It is true that the classroom is a more authentic setting for researching student learning. What is proved to be true in a laboratory may not occur in real classrooms. However, compared to classroom conditions, the laboratory environment enables researchers to monitor

student behaviors more accurately and gain finer control over experimental procedure. It allows minimizing the influence of some environmental or individual factors on the results.

## **2.8 Need for Cognition**

Need for cognition is a psychological construct which describes “an individual’s tendency to engage in and enjoy effortful cognitive endeavors” (Cacioppo et al., 1982, p. 197). This construct and the instrument used to measure it were developed by Cacioppo and Petty (1982), who used it to assess the general and stable disposition of an individual to involve in cognitive and intellectual activities. There are two forms of need for cognition scale, the 34-item scale and the short form of 18-item scale, which was developed by Cacioppo, Petty and Kao (1984).

The two forms of need for cognition scale have been widely used by researchers. A review conducted by Cacioppo et al (1996) investigated the relationship between need for cognition and other individual-difference variables. It found that need for cognition is positively correlated with the tendency to formulate complex attributions (Fletcher et al., 1986; Petty & Jarvis, 1996), seek relevant information in problem solving (Berzonsky & Sullivan, 1992), and be open to new ideas and actions (Berzonsky & Sullivan, 1992). Negative correlations were reported with the following variables: tendency to ignore and avoid new information (Venkatraman et al., 1990a), preference for order (Petty & Jarvis, 1996; Webster & Kruglanski, 1994) and causal uncertainty (Weary & Edwards, 1994).

Researchers also found that there is link between need for cognition and effortful information processing (Cacioppo et al., 1996). For example, students high in need for



cognition are more attentive to the quality of arguments for persuasive messages, while students low in need for cognition are more influenced by peripheral cues such as the number of arguments or attractiveness of the source messages (Cacioppo, Petty & Morris, 1983; Haugtvedt & Petty, 1992). Previous studies also indicated that need for cognition is linked to efficiency in processing argumentative texts (Kardash & Scholes, 1996) and disposition to approach arguments (Nussbaum & Bendixen, 2003).

The results of prior research suggest that students high in need for cognition are more likely to engage in cognitive and intellectual activities and are more argumentative. Therefore it is logical to assume such students will generate a higher quality of critical analyses.

## **CHAPTER 3:METHOD**

### **3.1 Pilot Study**

A pilot study was conducted before the main experiment to evaluate the experiment design and correct faults. The pilot study recruited 14 participants. Through the pilot study, several problems were discovered. Almost all the participants felt that the initial time allocated for reading and reviewing was too long. It was found that two tags provided in the pilot study (“definition” and “claim”) were seldom used. Moreover, five evaluation questions initially used as a dependent variable were too difficult. Most of the participants gave very brief and superficial answers to these questions making it difficult to identify variation in understanding. The experiment was then revised to fix these problems.

### **3.2 Participants**

A sample of 120 students at Simon Fraser University volunteered to participate in this study. Students who agreed to participate signed a consent form and were paid \$15.

Participants’ ages were between 17 and 48 years ( $M = 22.38$ ,  $SD = 5.83$ ). The sample included 56 males and 64 females. Of the 120 participants, 50 reported that English is their first language. Of the remaining 70 participants, 14 had studied in English instructional environments for more than 10 years. Eighty-five participants reported they had finished high school and were studying in undergraduate programs, an additional

eight had two-year diplomas from another institution, 10 had bachelor's degrees, 13 had master's degrees and 4 had PhD degrees. Each participant was randomly assigned to either the study group, summarize group or argue group.

### **3.3 Materials and Instruments**

#### **3.3.1 Demographics Questionnaire**

A questionnaire was used to collect information about age, gender, academic major, highest level of education attained, first language and years of studying in English-speaking schools if English was not their first language, and whether they had taken any courses in Anthropology. See Appendix 1 for the demographics questionnaire.

#### **3.3.2 Reading Materials**

In this study, students were asked to read web documents describing *Homo floresiensis* skeletons recently discovered in Indonesia and the competing theories of the origins of the Hobbit-like *Homo floresiensis*. These materials were intended to be scientifically accurate. They were obtained from Wikipedia (*Homo floresiensis*, 2008) and then revised according to the requirements of the study. Presented in a web environment called nStudy, the materials consisted of a main document (1,220 words), 6 hyperlinked mini-documents (100 to 200 words each), and 9 hyperlinked definitions of specialized terms. The mini-documents described (1) Flores Island where the skeletons were found, (2) local stories traditionally told about small human-like species, (3) the biological process of island dwarfing, (4) the disease condition of microcephaly, (5) a comparison chart showing the differences of chimpanzees and several homo species on

the location, time of existence, weight, height and brain size and (6) studies that compared microcephalic and Homo floresiensis skulls. The hyperlinked definitions were: (1) Neanderthal (2) tibia (3) pygmies (4) homo erectus (5) speciation (6) brain to body mass ratio (7) morphologist (8) homo sapiens and (9) upper paleolithic.

### **3.3.3 Need for Cognition Instrument**

Before receiving instructions, each participant completed the 18-item Need for Cognition self-report instrument (short form), that is intended to predict how much time and effort a respondent would choose to spend on intellectual activities that involve complex thinking. This instrument contains such items as “I would prefer complex to simple problems”, or “I only think as hard as I have to.” Previous researchers have found that internal consistency reliability was high ( $\alpha=0.9$ ) for this instrument (Cacioppo et al., 1984). Furthermore, Sadowski and Gulgoz (1992b) reported that a test-retest correlation was high ( $r = .88$ ) over a 7-week-period study using the 18-item scale. A number of studies also found the need for cognition scale to be gender neutral (Sadowski, 1993; Spotts, 1994; Tolentino, Curry & Leek, 1990). An investigation of the relationship between need for cognition and other individual-differences variables supported the convergent and discriminant validity of the construct (Cacioppo et al., 1996).

### **3.3.4 Instructions**

The three groups received the same reading materials but different instructions. The instructions were presented to participants on a single sheet of printed paper.

The study instructions (Appendix 2) asked participants to use the materials to prepare for a recall test. They were told that the more information from the materials they

memorized the higher score they would achieve on a recall test. The study instructions gave participants three examples of the types of information they may be asked to recall: difficult to remember information, details and examples.

The summarize instructions (Appendix 3) asked participants to use the materials to prepare for writing a summary. They were told that the better their summary the more successful they would be in the experiment. The summarize instructions gave participants three examples of the types of information suitable for including in their summary: main ideas, key terms and explanations.

The argue instructions (Appendix 4) asked participants to use the materials to prepare for writing an argument. A position was given for them to argue for: The skeletons found by researchers are of a species different from humans living today. They were told that the better their arguments the more successful they would be in the experiment. The argue instructions gave participants three examples of the types of information suitable for including in an argument: supporting claims, counterclaims and evidence.

### **3.3.5 Outcome Tasks**

I designed three tasks for participants to do after reading: a free recall test, a multiple-choice test and a theory evaluation essay. The free recall test (Appendix 5) asked participants to write everything they could remember from the material they read. They were told not to worry about spelling and grammar and that they could use point form or paragraph form.

The multiple-choice test (Appendix 6) consisted of 13 items, each having four options. Most of the multiple-choice questions evaluated participants' understanding of the causal relations presented in the material. For example, one of the questions is:

Which of the following statements is not consistent with the idea that the Hobbits were a separate homo species having small bodies and skulls?

- A. The brain to body mass ratio of the Hobbits is between that of *Homo erectus* and chimpanzees.
- B. The bones in the Hobbits' wrists were indistinguishable from a modern day African ape.
- C. Other animal bones found on Flores Island were much smaller than the animals' normal sizes.
- D. Except for the difference in size, the Hobbits were very similar to *Homo erectus*.

The correct answer of this question is A.

Participants were finally invited to write a one-page essay draft for a hypothetical issue of National Geographic magazine dealing with human evolution. The goal of the essay was to critically evaluate the competing theories about the origins of the Hobbits and persuade the hypothetical readers that one of the theories is more accurate. The evaluation question is presented in Appendix 7.

### **3.3.6 nStudy**

nStudy is a web tool being developed for researching learning (Winne & Hadwin, 2009). The participants accessed the reading materials, including text, graphics and hyperlinks through nStudy and the Firefox browser. nStudy can be used to create and

delete tags, tag information in a web page, create term definitions, make notes, and carry out various other operations pertinent to studying .

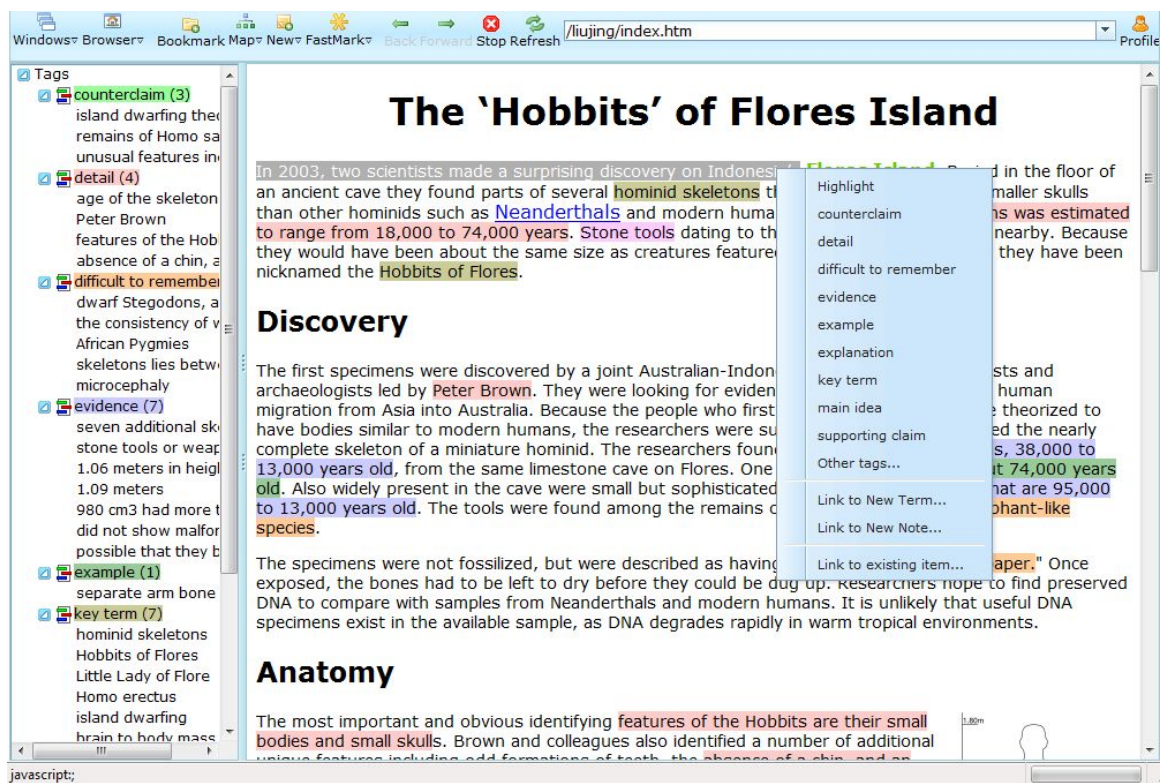
In this study, only tag features were used to help participants mark and study important text information as instructed. It is easy to tag text in nStudy. First, select the text; then, choose one of the tag names provided in a drop-down menu. Participants were allowed to delete tags by simply right-clicking the previously highlighted text and choosing “delete”.

In this experiment, the researcher created nine tags, three specifically appropriate to each experimental group, which correspond to the types of information identified in each group’s instructions. Categorized by the instructional set, the tags were:

- 1) Study tags
  - difficult to remember
  - detail
  - example
- 2) Summary tags
  - main idea
  - key term
  - explanation
- 3) Argument tags
  - supporting claim
  - counterclaim
  - evidence

Participants were told that they could use any of the nine tags to mark information needed to accomplish their assigned task. Each tag was predefined with a unique color. Participants were able to tag any text within the main document or mini-documents, and could delete a tag if they changed their mind. Figure 1 illustrates the nStudy tagging feature.

**Figure 1** nStudy tagging feature



### 3.4 Procedure

Volunteers who agreed to participate signed a consent form. After that, all participants were asked to fill out the demographic questionnaire and then the instrument measuring their Need for Cognition. Before starting reading in the nStudy environment,



participants received five minutes of training on how to tag information in the nStudy environment.

After randomly assigning participants to treatment groups, they were given a page (previously described) presenting treatment-specific instructions. Every participant was provided with a user name and password to log in to nStudy and then started reading and tagging. They were told that they had 20 minutes for reading and tagging, and then they would have 5 minutes to review because the browser window would be closed when they completed subsequent achievement tasks, thus they would not be able to see the text and tagged information after reviewing it.

After finishing reading and reviewing, participants were asked to do a crossword puzzle. Although our purpose was to wipe their short term memory, they were told to relax and have fun with the puzzle, and there was no need to worry about their answers. Five minutes later, participants were given 12 minutes to write down everything they could remember from the material they had studied before doing the crossword puzzle. They were told that spelling and grammar are not important, and they need not write in full sentences. They were given a maximum of 15 minutes to complete the free recall task. After the free recall test, participants were given up to 10 minutes to complete the multiple-choice test. Finally, participants were given up to 15 minutes to write an essay evaluating the competing theories presented in the reading materials. If they stopped writing within 15 minutes and before finishing one page, they were encouraged to continue writing.

### **3.5 Latent Semantic Analysis**

To code the free recall responses, two measuring methods were used, one is a proposition scoring method (Rewey, Dansereau, Skaggs, Hall & Pitre, 1989) and the other is latent semantic analysis (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990). The purpose of using latent semantic analysis is to check the reliability of proposition scoring.

Latent semantic analysis is one approach to the biggest challenge faced by researchers studying artificial intelligence and machine learning, that is, dealing with natural language used by human beings. It has proven extremely difficult to convert natural language statements into logical propositions that can be processed for meaning by a computer. However, by using mathematical analysis, a text document can be decomposed into a reduced dimensional representation in a so-called latent semantic space (Landauer, Foltz & Laham, 1998). By doing this, the semantic relations or the contextual-usage of words in a discourse are encoded (Landauer & Dumais, 1997).

The latent semantic space is a mathematical representation of a text document. It is constructed through several steps. First a large corpus of text is represented as a rectangular matrix with one cell representing the frequency of a given word appearing in a given passage (Landauer, Foltz & Laham, 1998). Of course, function words are excluded from the latent semantic analysis because they can be found in any context. Then, each cell frequency is transformed by weighting both the importance of the word and to what degree it carries information in a particular discourse. Next, the transformed matrix is decomposed into the product of three other matrices and finally reconstructed into a two-dimensional matrix using the matrix algebra method of singular value

decomposition (SVD). Terms, sentences and passages can then be compared within this two-dimensional semantic space. Latent semantic analysis does not make use of word order. It approximately represents the meaning of a word as the average of the meaning of passages in which it occurs and the meaning of passages as the average of the meaning of their words.

The final two-dimensional matrix is a representation of a relation between terms and concepts and a relation between these concepts and the documents. For example, a document is first described as having concept A and concept B. The correlation is then calculated between a term and concept A and B respectively. Finally, the two correlations are combined to get an estimated correlation between the term and the document. Therefore, although LSA is a computational analysis of text, it is able to capture relations of words in a context sensitive way (Landauer, Foltz & Laham, 1998). For example, let us look at the three words: eyes, story and vision in the following two sentences: “There were tears in her eyes as she listened to the story” and “Tears blurred her vision.” In LSA, although “eyes” did not appear in the second sentence, the estimated correlation between vision and eyes is higher than that between eyes and story if an infinite sample of sentences were given because there is a great chance that “eyes” and “vision” will co-occur. The chance is calculated by first extracting concepts from a larger context and then estimating the correlation between each word and the concepts. LSA thus finds that both eyes and vision have higher correlation with the central concepts than that between story and the central concepts.

## **CHAPTER 4:RESULTS**

### **4.1 Overview of the Types of Data Collected**

This study collected six types of data: (1) demographic data including education level, major, age, first language, English experience; (2) need for cognition which measures disposition to engage in cognitive tasks; (3) log data capturing number of nStudy actions and use of tags; (4) multiple-choice test scores measuring understanding of causal relations; (5) free recall scores measuring retention of the reading material; and (6) variables measuring the use of reasoning in an evaluative essay.

### **4.2 Scoring Free Recall Responses**

The free recall responses were coded using a proposition scoring method (e.g., Rewey, Dansereau, Skaggs, Hall & Pitre, 1989). A proposition contains only one piece of information. There may be several propositions in a sentence. For example, consider the sentence “There is a fairly complete skeleton proposed to belong to a 30 year old female.” It was decomposed into three propositions: (1) there is a complete skeleton, (2) the skeleton belonged to a female, and (3) the skeleton belonged to a 30-year-old. Notice that, although more propositions are literally possible (e.g., the skeleton was fairly complete), a pragmatic propositional decomposition was chosen which captured the key information. The original reading was analyzed into 328 propositions. When scoring a free recall response, if a proposition was entirely present and accurate it was assigned 1.

A partially present or partially accurate proposition was assigned 0.5. An absent or completely inaccurate proposition was assigned 0. Inferences or information not present in the original reading were not scored. The final free recall score for each participant was the sum of scores in the participant's response over all proposition in the text. A rater who was blind to treatment conditions coded all 120 free recall responses. A second rater coded 20 randomly selected responses with 92% agreement.

In addition to the use of a proposition scoring method, another measuring tool, latent semantic analysis was used to check whether proposition scoring was reliable. The LSA scores were calculated by using the online tool developed by Laham et al. (1998). A LSA score for each participant was obtained by comparing the recall response to the original text.

In addition, the word count for each response was obtained automatically from Microsoft Word to determine whether there were differences in counts across the three groups, and whether the word count of the recall was related to the recall performance score.

### **4.3 Scoring Essays**

A coding scheme consisting of 13 variables were used to score the essays:

- 1) *Position* is a categorical variable that indicates whether an essay took a position on the major theoretical dispute represented in the reading. An essay was assigned a score of 0 if it did not take a position, 2 if it did take a position. A score of 1 was given if the essay stated that more evidence is needed to decide which theory is right.

- 2) *Misunderstanding* is a dichotomous variable that indicates whether the essay writer seemed to misunderstand the instructions, for example by writing an essay on a different topic.
- 3) *Reasons* is the number of reasons presented in an essay, including those presented as counterarguments and rebuttals. Here, reasons include any supporting evidence of a claim, not just those having a causal relation with the claim. For example, “The Hobbits are the shortest hominid ever found. They are only about one meter high and shorter than even the shortest populations such as African Pygmies.” Here, the second sentence is a supporting evidence of the first claim, which is counted as a reason although it is not causally related to the claim.
- 4) *High-level reasons* is the number of first level reasons, including counterarguments and rebuttals relating directly to the major issue claims. For example, “The Hobbits showed advanced behaviors” is counted as a high-level reason because it directly supports a major issue claim, that is, “some scientists believe that the Hobbits are a new species”.
- 5) *Low-level reasons* is the number of reasons, counterarguments and rebuttals given for all non-major-issue claims and sub-claims under the major issue claims. The non-major issue claims are claims that are not related to the central theories of the reading. For instance, in the previous example, “The Hobbits are the shortest hominid ever found” is not a major issue claim, so all the reasons supporting it are counted as low-level reasons. “The Hobbits showed advanced behaviors” is a sub-claim of “the Hobbits are a new species”, so the reasons that support this sub-claim such as the use of fire for cooking are low-level reasons.
- 6) *Counterclaims* is the number of claims given which oppose a major-issue claim, a sub-claim or a non-major-issue claim. Sub-claims are claims that support the major-issue and non-major-issue claims. Counterclaims are not counted as reasons unless they are scored as reasons. For example, "Some scientists believe that the Hobbits are a new species. However, others

believe that their short stature is caused by a genetic disease." Here, the second sentence is a counterclaim but not a reason.

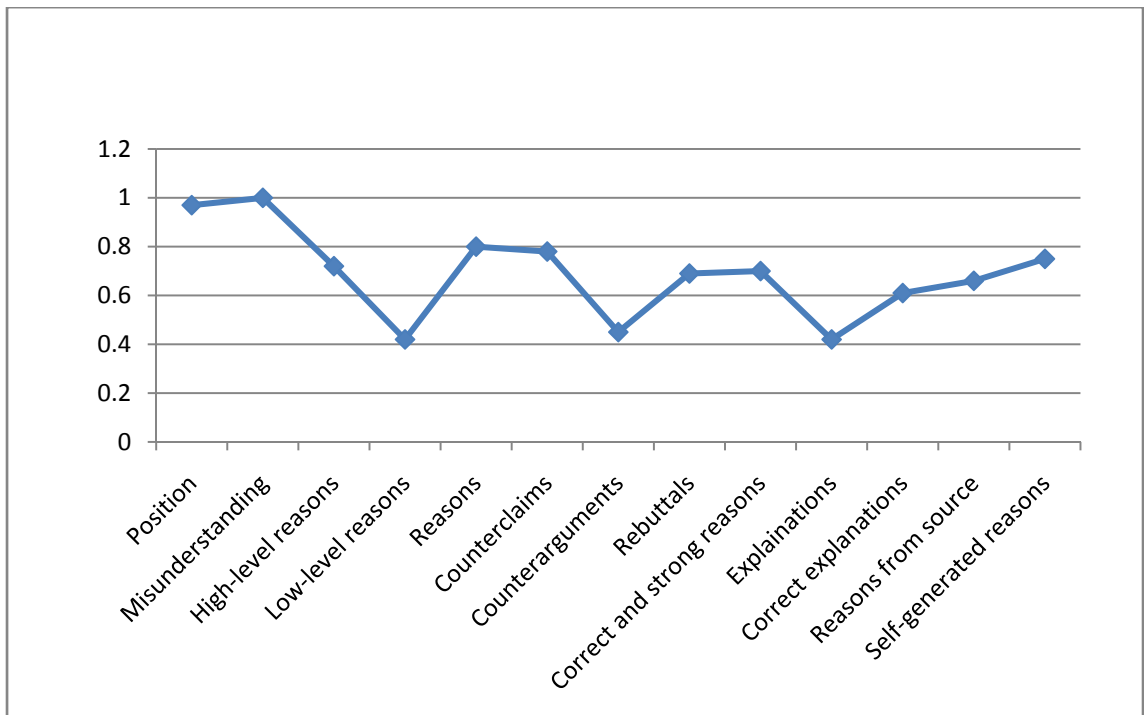
- 7) *Counterarguments* is the number of reasons given that oppose a major issue claim, a sub-claim or a non-major-issue claim. For example, "Most scientists believe that island dwarfing is a type of evolution in which limited food resources select for smaller body size. However, it is possible that island dwarfing is not a form of evolution but is instead caused by the scarcity of food restricting the diet of growing animals so that they do not grow as large. This type of change is not genetic. If the food supply suddenly increased the animals would return to their normal size after one generation." Here, "However...one generation" is a counterargument, which can be counted as a reason, but not a counterclaim.
- 8) *Rebuttals* is the number of reasons given that argue against counterclaims and counterarguments, but support the original claims at the same time.
- 9) *Correct and strong reasons* is the number of reasons given, including counterarguments and rebuttals that are both correct in themselves and bear a logically valid connection with a claim.
- 10) *Explanations* is the number of statements, different from evidence, that describe why or how something happened.
- 11) *Correct explanations* is the number of explanations that are correct.
- 12) *Reasons from the source* is the count of all reasons, including counterarguments and rebuttals that are from the source material. If a reason is not correctly presented but it is still from the source, it is assigned into this category.
- 13) *Self-generated reasons* is the number of new reasons generated by the participant or reasons from the text for which the participant generated a new connection to a claim. For example, "advanced behaviours and high self-awareness suggest that the Hobbits did not suffer from a genetic disease." This piece of information, "advanced behaviours and high self-

awareness” is from the source material, but the student created the connection between it and a theory (the non-disease theory), which is not mentioned in the original reading.

In this coding scheme, *reasons* is equal to *high level reasons* plus *low-level reasons*. It is also equal to the number of reasons from the source plus *self-generated reasons*.

To monitor reliability, another rater coded a subset of the essays (20). The reliability differs from one variable to another (See Figure 2). The inter-rater reliability is sufficient on the most important variables such as the number of reasons and correct and strong reasons. A plausible explanation for the low reliability in low-level reasons and explanations is that they tended to have quite low frequencies and thus a slight inconsistency in coding would generate large decreases in reliability.

**Figure 2** Inter-rater reliability on all variables of the essays





## **4.4 Analysis and Results**

### **4.4.1 Outliers and Normality**

Before data analysis every variable was checked for outliers and normality within each group. One outlier was found in free recall scores from the summary group ( $z = 3.22$ ). The score was changed to reduce the impact of the outlier (Tabachnick & Fidell, 2007). First, the next highest z-score was found ( $z = 2.17$ ), corresponding to a raw score of 35. The original score of the outlier was lowered to 35.5, which is one scoring unit higher than the next highest score. Because the distribution of English experience showed extremely negative skew, that variable was dichotomized by an approximate median split, to create approximately equal sized groups, representing high (more than 10 years) and low experience. Almost all other variables were approximately normally distributed except for some lower frequency essay variables such as counterclaims, counterarguments, rebuttals, explanations and correct explanations.

### **4.4.2 Demographic Data and Need for Cognition**

Table 1 shows the gender balance and percent of participants with a high level of English experience in each group. It also shows the means and standard deviations (in parentheses) for participants' age, educational level and need for cognition. An analysis of variance (ANOVA) indicated no statistically detectable difference between three conditions in education level,  $F(2, 117) = .77, p = .47$ , and age,  $F(2, 117) = .22, p = .80$ . Despite the fact that participants were randomly assigned to groups after completing the need for cognition questionnaire, the argument group was found to have higher need for cognition than the other two groups,  $F(2, 117) = 2.85, p = .06$ ; showing an advantage of

$d = .55$  in comparison with the study group and  $d = .38$  in comparison with the summary group.

**Table 1 Gender, English experience, education level, age and need for cognition**

	Study group	Summary group	Argument group	All
Female	47.5%	50%	62.5%	54%
English experience				
(greater than 10 years)	60%	47.5%	60%	56%
Education level	1.65 (1.20)	1.88 (1.29)	1.55 (1.13)	1.69 (1.20)
Age	22.35 (6.14)	22.82 (6.20)	21.95 (5.23)	22.38 (5.83)
Need for cognition	80.08 (17.33)	82.55 (17.53)	88.38 (12.48)	83.67 (16.20)

#### 4.4.3 Log Data Analysis

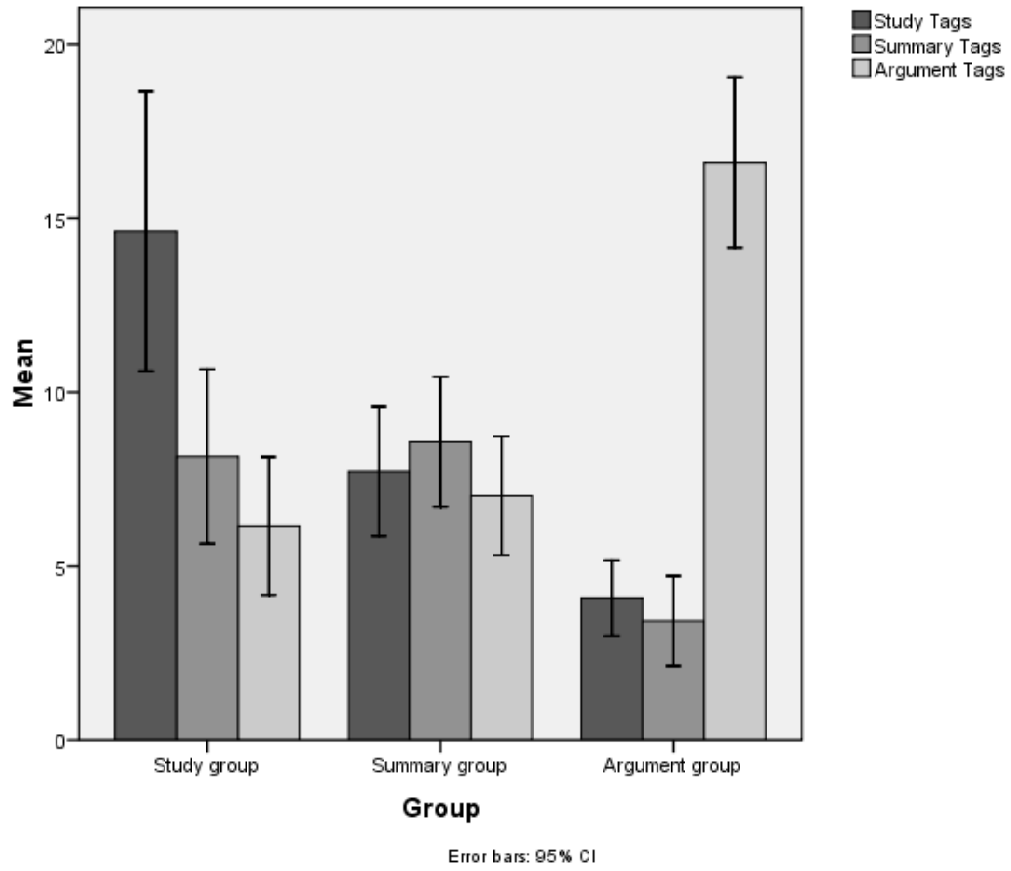
The log files automatically captured participants' activities in nStudy during the experiment sessions. The only nStudy actions analyzed were hyperlink clicking (to view mini-documents and definitions) and text tagging. In this study, I was interested to examine possible differences in the number of actions and use of tags among three

**Table 2 Mean number of actions (hyperlink clicking and text tagging) in three groups**

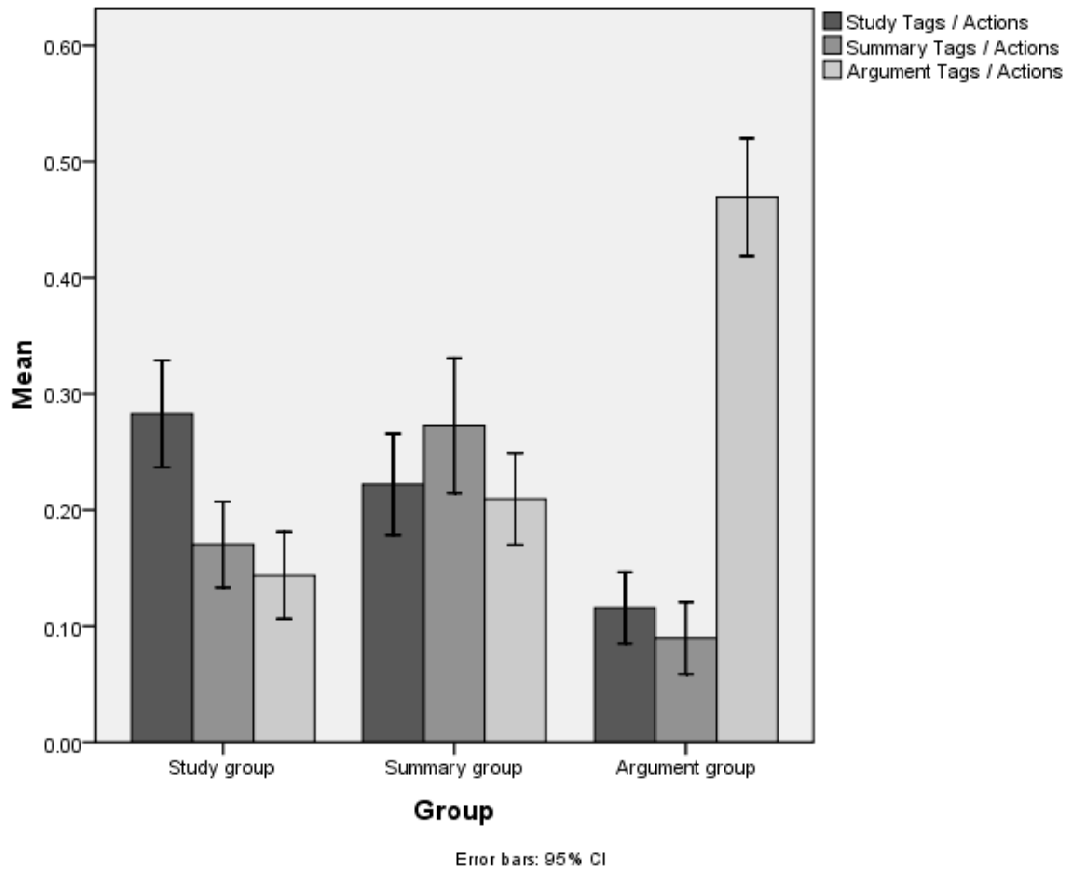
	N	Mean	Std. Deviation	Std. Error	Minimum	Maximum
Study group	40	46.60	21.00	3.32	20.00	94.00
Summary group	40	32.55	10.33	1.63	9.00	61.00
Argument group	40	35.23	10.46	1.65	13.00	64.00
Total	120	38.13	15.90	1.45	9.00	94.00

groups. An ANOVA test indicated the study group performed significantly more actions than the other two groups,  $F(2, 117) = 10.16, p < .001$ . A post-hoc test showed that there was no difference between the summary and the argument group ( $p > .99$ ), but the study group differed significantly from the other two groups in the number of actions (both  $p < .01$ ). The use of tags can be analyzed as a comparison of absolute mean frequencies (Figure 3) or mean proportions (Figure 4). From Figure 3, it can be seen that the study group did not differ from the summary group on their use of summary tags. The reason is that more actions were performed by the study group. I found that the action characteristics distinguishing each group could be more clearly shown by aggregating the number of each participant's actions of different types as a proportion of that participant's total actions. For example, if a student had 15 argument tags and 50 actions, then the proportion of argument tags to actions for this student is .3. The frequency of tagging was thus transformed into proportions of actions that were study tags, summary tags and argument tags. The groupwise means of these proportions are shown in Figure 4. Three ANOVAs shown in Table 3 statistically detected differences in the choice of tag types across the three groups: The study group used proportionally more study tags than the other two groups; the summary group used proportionally more summary tags; and that argument group used proportionally more argument tags. This finding demonstrates that the participants' studying was guided by the specific instructions they received.

**Figure 3** Number of study tags, summary tags and argument tags in three groups



**Figure 4** Proportions of actions containing study tags, summary tags and argument tags in three groups



**Table 3 One-way ANOVA of proportions of actions containing study tags, summary tags and argument tags across three groups**

		Sum of Squares	df	Mean Square	F	Sig.
Study Tags / Actions	Between Groups	.57	2	.29	17.68	< .001
	Within Groups	1.90	117	.02		
	Total	2.47	119			
Summary Tags / Actions	Between Groups	.67	2	.34	18.13	< .001
	Within Groups	2.17	117	.02		
	Total	2.85	119			
Argument Tags / Actions	Between Groups	2.37	2	1.19	65.78	< .001
	Within Groups	2.11	117	.02		
	Total	4.48	119			

#### 4.4.4 The Analyses of Outcome Tests

##### 4.4.4.1 The Multiple-choice Test

There were 13 multiple choice questions. They were scored by assigning 1 for correct answers and 0 for incorrect answers then summing scores to produce a final score for each participant. The maximum possible range of the total score is thus from 0 to 13. Cronbach's alpha for the 13 items was .43, which is unacceptably low. Table 4 shows descriptive statistics of the performance of the three groups on the multiple-choice test. A Levene test of homogeneity of variances indicated that the variance of scores on multiple choices is homogeneous ( $p = .36$ ). An ANOVA test (see Table 5) indicated that there are no statistically detectable differences among the means of three groups on the multiple-choice test as is expected given the very low reliability coefficient.

**Table 4 Descriptive statistics of the multiple-choice test in three groups**

	N	Mean	Std. Deviation	Std. Error	Minimum	Maximum
Study group	40	5.53	1.87	.30	2.00	10.00
Summary group	40	5.50	2.35	.37	1.00	10.00
Argument group	40	5.75	2.23	.35	1.00	11.00
Total	120	5.59	2.15	.20	1.00	11.00

**Table 5 One-way ANOVA test of the multiple-choice test in three groups**

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.52	2	.76	0.16	.85
Within Groups	547.48	117	4.68		
Total	549.00	119			

#### 4.4.4.2 Free Recall Test

Table 6 reports the performance of three groups on the free recall test. A test of homogeneity of variances did not detect a difference in the variances of three groups on the free recall test ( $p = .28$ ). An ANOVA test indicated that differences among the means of free recall scores for the three groups were not statistically detectable,  $F = .50$ ,  $p = .61$ .

**Table 6 Descriptive statistics of the free recall test in three groups**

	N	Mean	Std. Deviation	Std. Error	Minimum	Maximum
Study group	40	18.21	9.34	1.48	4.00	37.00
Summary group	40	16.35	8.04	1.27	3.50	37.50
Argument group	40	17.78	8.83	1.40	2.00	36.50
Total	120	17.45	8.71	.80	2.00	37.50

The reliability of the proposition scoring was confirmed by a high correlation with the latent semantic analysis ( $r = .73$ ). Homogeneity of variances on scores of latent semantic analysis scores was confirmed ( $p = .19$ ). An ANOVA conducted on participants' latent semantic analysis scores for the free recall responses revealed no statistically detectable differences across three groups,  $F(2, 117) = .89, p = .42$ .

Table 7 shows the mean length (in words) of free recall responses in each of the three groups. A homogeneity of variance test did not detect heterogeneity in the length of the free recall responses across the three groups ( $p = .36$ ). No differences in word length were statistically detected,  $F(2, 117) = .82, p = .44$ .

**Table 7** Descriptive statistics of word count of the free recall test in three groups

	N	Mean	Std. Deviation	Std. Error	Minimum	Maximum
Study group	40	207	93.42	14.77	57.00	446.00
Summary group	40	190	70.64	11.17	55.00	299.00
Argument group	40	212	80.91	12.79	86.00	421.00
Total	120	203	82.07	7.49	55.00	446.00

#### 4.4.4.3 Evaluation Essays

Means of the three groups on seven variables derived from the evaluation essays are shown in Table 8. The argument group generated many more reasons, especially high-level reasons than did the other two groups. They produced more correct and strong reasons, which suggests that the quality of their reasons was higher. They also produced



**Table 8** Descriptive statistics of some important variables of the essays in three groups

		Group Mean	Std. Deviation
high-level reasons	Study	2.25	1.89
	Summarize	2.70	2.07
	Argue	4.03	2.48
reasons	Study group	3.58	2.54
	Summary group	3.63	2.54
	Argument group	5.60	3.02
counterarguments	Study group	.35	.74
	Summary group	.75	1.06
	Argument group	.93	1.02
correct and strong reasons	Study group	2.33	1.99
	Summary group	2.33	2.19
	Argument group	4.45	2.75
Reasons from source material	Study group	1.93	1.87
	Summary group	2.43	1.96
	Argument group	3.23	2.04
Self-generated reasons	Study group	1.65	1.69
	Summary group	1.20	1.29
	Argument group	2.40	2.15
Word count	Study group	209.35	81.27
	Summary group	214.90	90.68
	Argument group	251.78	90.10

more self-generated reasons than the other two groups. Differences between the summary and study group tended to be smaller.

The ANOVAs shown in Table 9 indicated significant treatment effects on all important essay variables. Post hoc tests shown in Table 10 revealed that the argument

**Table 9 One-way ANOVA of some important variables of the essays**

		Sum of Squares	df	Mean Square	F	Sig.
high-level reasons	Between Groups	68.12	2	34.06	7.31	.001
	Within Groups	544.86	117	4.66		
	Total	612.99	119			
reasons	Between Groups	106.72	2	53.36	7.27	.001
	Within Groups	858.75	117	7.34		
	Total	965.47	119			
counterarguments	Between Groups	6.95	2	3.48	3.86	.024
	Within Groups	105.38	117	.90		
	Total	112.33	119			
correct and strong reasons	Between Groups	120.42	2	60.21	11.05	.000
	Within Groups	637.45	117	5.45		
	Total	757.87	119			
reasons from source material	Between Groups	34.40	2	17.20	4.48	.013
	Within Groups	449.53	117	3.84		
	Total	483.93	119			
self-generated reasons	Between Groups	29.40	2	14.70	4.84	.010
	Within Groups	355.10	117	3.04		
	Total	384.50	119			
word count	Between Groups	42539.32	2	21269.66	2.78	.07
	Within Groups	894891.68	117	7648.65		
	Total	937430.99	119			

group outperformed the other two groups on these variables, except that the argument group only outperformed the study group on counterarguments and reasons from source, but not the summary group. The tests also indicated no statistically detectable difference between the argument group and the study group on self-generated reasons. Overall, the findings demonstrate significant effects for the evaluative writing task on the quality of the essays, specifically on the quantity and quality of reasons.

There was a detected effect of treatment on the length of the essays (Table 9). Descriptive statistics show that there was no difference between the study group and the summary group, and the argument group's essays tended to be longer. There were fairly strong correlations between length (word count) of the essays and the number of reasons ( $r = .69$ ) and correct reasons ( $r = .62$ ).

**Table 10** Multiple comparisons of some important variables of the essays across three groups

Dependent Variable	(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
high-level reasons	Study group	Summary group	-.45	.48	1.000	-1.62	.72
		Argument group	-1.78*	.48	.001	-2.95	-.60
	Summary group	Study group	.45	.48	1.000	-.7	1.62
		Argument group	-1.32*	.48	.021	-2.50	-.15
reasons	Study group	Summary group	-.05	.61	1.000	-1.52	1.42
		Argument group	-2.03*	.61	.003	-3.50	-.55
	Summary group	Study group	.05	.61	1.000	-1.42	1.52

Dependent Variable	(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
counterarguments	Study group	Argument group	-1.98*	.61	.004	-3.45	-.50
		Summary group	-.40	.21	.186	-.92	.12
	Summary group	Argument group	-.58	.21	.023	-1.09	-.06
		Study group	.40	.21	.186	-.12	.92
correct and strong reasons	Study group	Argument group	-.18	.21	1.000	-.69	.34
		Summary group	.00	.52	1.000	-1.27	1.27
	Summary group	Argument group	-2.13*	.52	.000	-3.39	-.86
		Study group	.00	.52	1.000	-1.27	1.27
reasons from source material	Study group	Argument group	-2.13*	.52	.000	-3.39	-.86
		Summary group	-.50	.44	.769	-1.56	.56
	Summary group	Argument group	-1.30*	.44	.011	-2.36	-.24
		Study group	.50	.44	.769	-.56	1.56
self-generated reasons	Study group	Argument group	-.80	.44	.212	-1.86	.26
		Summary group	.45	.39	.751	-.50	1.40
	Summary group	Argument group	-.75	.39	.170	-1.70	.20
		Study group	-.45	.39	.751	-1.4	.50
	Summary group	Argument group	-1.20*	.39	.008	-2.15	-.25
		Study group					

#### **4.4.4.4 Regression Analyses of Essay Reasoning**

Although no differences were observed in performance on the free recall and the multiple-choice tests, analyses revealed a difference between the argue group and other two groups on the quality reasoning in the essays. However, further analysis is needed to find out whether the argumentation treatment, and presumably the elicitation of an argument schema, was causally related to reasoning in the essays and whether other factors predicted essay performance. Theoretically, need for cognition is one possible predictor for the quality of writing. Previous studies revealed that need for cognition is positively correlated with argumentativeness (Kardash & Scholes, 1996, Nussbaum, 2005). Therefore, it is predicted that need for cognition will result in more arguments and more high quality arguments. In addition, learners for which more actions were recorded may be more cognitively active and engaged in the task than other learners, and therefore may produce essays that give more reasons. The number of actions is thus considered as another predictor. Here we treat the experimental group and the use of argue tags as two different predictors. The reason is that although the argument writing instruction will affect the number and depth of reasons generated, the use of argue tags in the summary and study group will also influence their quality of arguments.

Tables 11 and 12 report the sequential multiple regressions that used the four variables to predict the number of reasons and correct reasons in the essays. A new group variable was created that used 0 to code for both the study and summary groups and 1 to code for the argument group. As can be seen from Model 1 in both tables, the hypothesis that need for cognition would positively predict the number and quality of reasons in the essays was confirmed.

**Table 11 Multiple regression using NFG, group, actions and argument tags to predict the number of reasons**

Model	R	Adjusted R Square	Change Statistics				
			R Square Change	F Change	df1	df2	Sig. F Change
1 need for cognition	.24 <sup>a</sup>	.05	.06	7.22	1	118	.008
2 need for cognition, group	.38 <sup>b</sup>	.13	.08	11.38	1	117	.001
3 need for cognition, group, actions	.47 <sup>c</sup>	.21	.08	12.53	1	116	.001
4 need for cognition, group, actions, argument tags	.51 <sup>d</sup>	.24	.04	5.66	1	115	.019

As can be seen from the change in  $R^2$  and F for Model 2 in both tables, argumentation had an effect on reasoning, even when need for cognition was statistically controlled. This is an important result because it shows that the instruction to prepare to argue contributed to the higher performance of the argue group, over and above the contribution due to their higher need for cognition.

The addition of actions in Model 3 indicates that recorded activity in nStudy was a significant predictor of the number and quality of reasons. This result occurred despite the fact that the correlation (see correlations in Appendix 8) between actions and the number of correct and strong reasons is low ( $r = .17$ ). How then can actions have a significant effect on the correct reasons? The explanation is that group is negatively

correlated with actions (see correlations in Appendix 8) because the study group has a greater number of actions than the argument group. This negative correlation suppresses the correlation between actions and reasoning. Thus, statistically removing the effect of group reveals the significant predictive relationship between study actions and reasoning.

**Table 12 Multiple regression using NFG, group, actions and argument tags to predict the number of correct and strong reasons**

Model	R	Adjusted R Square	Change Statistics				
			R Square Change	F Change	df1	df2	Sig. F Change
1 need for cognition	.23 <sup>a</sup>	.04	.05	6.28	1	118	.014
2 need for cognition, group	.42 <sup>b</sup>	.17	.13	18.49	1	117	.000
3 need for cognition, group, actions	.48 <sup>c</sup>	.21	.05	7.89	1	116	.006
4 need for cognition, group, actions, argument tags	.51 <sup>d</sup>	.23	.03	3.80	1	115	.054

The addition of argue tags in Model 4 indicates that the use of argument tags had predicted the number of reasons and the number of correct reasons even after the effect of the argumentation treatment was statistically controlled. However, note that the predictive relationship with the number of correct and strong reasons decreases. This is not surprising because statistically removing the effect of argumentation treatment decreases the relationship of argument tag usage with both reasoning variables. The findings suggest both argument writing instruction and the spontaneous use of argument

tags influence on the quality of essays even after the effects of other strong factors have been statistically removed.

Participants' education level and their years of English experience were also correlated with the quality of their evaluative essays. Sequential multiple regression

**Table 13 Multiple regression using group, education level and English experience to predict the number of reasons**

Model	R	Adjusted R Square	Change Statistics				
			R Square Change	F Change	df1	df2	Sig. F Change
1 group	.33 <sup>a</sup>	.10	.11	14.66	1	118	.000
2 group, education level	.37 <sup>b</sup>	.12	.03	3.46	1	117	.066
3 group, education level, English experience	.43 <sup>c</sup>	.17	.05	7.22	1	116	.008

**Table 14 Multiple regression using group, education level and English experience to predict the number of correct and strong reasons**

Model	R	Adjusted R Square	Change Statistics				
			R Square Change	F Change	df1	df2	Sig. F Change
1 group	.40 <sup>a</sup>	.15	.16	22.29	1	118	.000
2 group, education level	.44 <sup>b</sup>	.18	.03	4.54	1	117	.035
3 group, education level, English experience	.49 <sup>c</sup>	.22	.05	7.59	1	116	.007



analyses shown in Tables 13 and 14 showed that these factors still predicted reasoning after the effect of instruction to argue was statistically removed.

## CHAPTER 5:DISCUSSIONS

### 5.1 Discussion of the Results

The results indicated that there were no statistically detectable differences in the three groups' free recall scores. This is convergent with previous studies which demonstrated that more elaborative processing of text does not always increase recall of text information beyond that obtained by less elaborative processing (Kintsch & Young, 1984; Mannes & Kintsch, 1987; McDaniel & Donnelly, 1996; McNamara, Kintsch, Songer & Kintsch, 1996). It is also consistent with the Wiley and Voss study (1999) which revealed that argumentative writing was no more effective than narrative writing tasks for enhancing students' recall. The results show that the argument group did as well as the study group whose task was to memorize as much information as possible. Thus, although the argument group was not instructed to memorize, the cognitive processing induced by argumentation, and perhaps by an argument schema, competes favorably with the cognitive processing induced by other learning strategies. The results are consistent with the idea that understanding is an effective aid to memorization of propositions.

The study failed to confirm the hypothesis that students who are instructed to argue will perform better in understanding of causal relations than students who are instructed to summarize and study. Differences were not detected in the three groups' multiple-choice test scores. A possible reason is that the design of the multiple-choice test is problematic, which is suggested by its low internal consistency ( $\alpha = .43$ ). Therefore,

one cannot conclude whether the instruction to use text to prepare for an argument offers advantages for student understanding of causality.

It is promising that the instruction to use text to prepare an argument affects the number and strength of reasons in essays. These results suggest that invoking an argument schema, which presumably occurs in argumentative situations, is an effective instructional method for focusing learning on reasons. This is further confirmed by the finding that students who used more argue tags outperformed those who did not on the number and quality of reasons. It is quite likely that these students were also engaged in the process of arguing although they were not instructed to do so. It remains unclear whether the argue tags invoked an argument schema or the learners who used the argue tags would have engaged in argument processing without them.

Need for cognition, which is the tendency or desire to involve in cognitive activities (Cacioppo et al., 1996), has been found in previous research to positively correlate with argumentativeness (Nussbaum & Bendixen, 2003). This study is consistent with that finding participants with higher need for cognition produced more reasons in the evaluative writing task.

The relationship between studying actions and quality of essays is somewhat complicated. There is evidence that students who spontaneously engage in more studying actions write higher quality essays. However, the treatment group with the greatest number of studying actions, the study group, produced lower quality essays. One explanation is that because participants in the study group were told to memorize as much information as possible, they likely tagged often to try to remember every piece of information. Indeed, the study group performed many more actions than the other two

groups. However, using tags for the purpose of remembering may not be linked to deep cognitive engagement. After researchers develop an understanding of which tags are associated with deeper cognitive engagement, it may be possible to use observations of the type of tagging to predict the quality of reasoning.

## **5.2 Implications**

The exciting message conveyed by this research is that reasoning about previously studied materials can be greatly improved by instructing students to study with the goal of preparing an argument without actually writing while materials are available. The study-to-argue intervention offers greater flexibility in the classroom deployment of argumentation as a learning strategy. To mentally plan an argument while reading is less time-consuming than writing an argumentative essay, and presumably presents fewer barriers to students who are poor writers. In addition, there are many opportunities for teachers to offer this type of intervention because reading is a common classroom practice.

The study-to-argue interventions can be readily and naturally combined with other learning activities. For example, they can be used in individual pre-reading prior to argumentation-based collaborative learning activities. Collaborative learning situations require two or more students work together to achieve a joint product. During this process, it is critical to discuss with partners to reach agreement. By embedding study-to-argue interventions into a pre-reading stage before collaborative learning activities, participants are predicted to be able to develop and understand different positions and

thus facilitate their subsequent negotiation. Such interventions can also be involved in individual or collaborative problem solving as we know that problem solving includes such important steps as information gathering, and developing and selecting alternatives. Study-to-argue interventions can be used in all these steps to engage students in weighing gathered materials and alternatives developed from them.

This research suggests that study-to-argue interventions can help learners to think more deeply about what they are reading. Students with an argumentation goal are predicted to focus their attention less on surface features and superficial meaning of the text and more on deeper processing of the positions and arguments presented in the text. Ultimately it may also lead them to construct their own point of view and provide reasons and evidence to support it.

### **5.3 Limitations and Future Work**

One limitation of this study is that the design of the multiple-choice test that assessed understanding of causal relations was problematic because of its low reliability. Perhaps because of this, the instruction to argue was not found to influence students' understanding of causality. Further studies should avoid this problem.

Another limitation is that, although students' understanding of text is strongly related to their reading ability, the role of reading ability in moderating the effects of argumentation was not investigated in this study. Besides the education level and English experience, reading speed likely impacts on learners' comprehension of a text, especially for the text in this study which contained many unfamiliar words. The time limit for

reading constrains the full processing of reading materials for some students. In a future study, it is important to examine the effect of reading speed.

Many students are unwilling to engage in effortful cognitive activities due to low need for cognition, lack of interest, motivation, and so on. Further analysis of the data collected in this research could be performed to determine whether study-to-argue interventions can especially motivate students with low need for cognition to think more deeply. It might be expected that engaging students in argumentation would motivate them to actively seek information useful to support a position while at the same time refuting the opposite position.

Because this study mainly focused on promoting students' near transfer, that is, their understanding of causality and ability to make critical analyses related to the studied materials, future research should examine whether the study-to-argue intervention would improve far transfer on tasks which require solving problems in a different setting, for example, using learned knowledge to solve another dispute in research on homo species. Interventions effective for near transfer tasks do not always produce same similar effects on far transfer tasks. It is therefore possible that our results may not be replicated on far transfer tasks.

In addition, this study provided participants with a verbal model of arguments in the written instructions. This is a simple way of presenting argument elements. The same results may hold if a procedure or a physical model of argument building is available because it will show students a clearer structure of arguments (Nussbaum & Schraw, 2007). Future research should attempt to use models that substantially include non-verbal elements and see whether the effects would hold or be even greater.

Finally, participants in the study were involved in the process of individual argumentation, so it is not known whether the same results would hold in interactive argumentation. Researchers have found that dialogical argumentation can promote conceptual understanding (Asterhan & Schwarz, 2007). It is thought that interactive argumentation may stimulate deeper thinking because participants have to respond to the questioning from others by making rebuttals, and face a great variety of ideas (Anderson, Chinn, Waggoner, & Nyguyen, 2002). Future studies that investigate interactive argumentation may find that it has even greater effects.

## REFERENCE LIST

- Alvermann, D. E., Hynd, C. E., & Quian, G. (1995). Effects of interactive discussion and text type on learning counterintuitive science concepts. *Journal of Educational Research*, 88, 146 - 154.
- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, 84, 261-271.
- Anderson, R. C. (1977). The notion of schemata and the educational enterprise. In R. C. Anderson, R. J. Spiro, & W. E. Wontague (Eds.), *Schooling and the acquisition of knowledge* (pp. 415-431). Hillsdale, NJ: Erlbaum.
- Anderson, R. C. (1984). Role of the reader's schema during comprehension, learning, and memory. In R. C. Anderson, J. Osborn, & R. J. Tierney (Eds.), *Learning to read in American schools* (pp. 243-258). Hillsdale, NJ: Erlbaum.
- Anderson, R. C., Chinn, C., Waggoner, M., & Nyguyen, K. (2002). Intellectual stimulation story discussions. In J. Osborn & F. Lehr (Eds.), *Literacy for all: Issues in teaching and learning* (pp. 172-186). New York: Guilford.
- Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading comprehension. In P. D. Pearson, R. Barr., M. L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 255-291). New York: Longman.
- Andriessen, J. (2006). Arguing to learn. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 443-459). New York: Cambridge University Press.
- Andriessen, J., Baker, M., & Suthers, D. (2003). Argumentation, computer support, and the educational context of confronting cognitions. In J. Andriessen, M. Baker, & D. Suthers (Eds.), *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments* (pp.1-25). Norwell, MA: Kluwer Academic.
- Andriessen, J., Erkens, G., Van de Laak, C., Peters, N., & Coirier, P. (2003). Argumentation as negotiation in electronic collaborative writing. In J. Andriessen, M. Baker, & D. Suthers (Eds.), *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments* (pp.79-115). Norwell, MA: Kluwer Academic.
- Angell, R. B. (1964). *Reasoning and logic*. New York: Appleton-Century-Crofts.



- Asterhan, C. S. C., & Schwarz, B. B. (2007). The effects of monological and dialogical argumentation on concept learning in evolutionary theory. *Journal of Educational Psychology, 96*, 626-639.
- Auriac-Peyronnet, E. (2001). The impact of oral training on argumentative texts produced by ten-and eleven-year-old children: Exploring the relation between narration and argumentation. *European Journal of Psychology of Education, 16*(2), 299-317.
- Ausubel, D. P. (1963). *The psychology of meaningful verbal learning*. New York: Grune & Stratton.
- Baker, M. (1999). Argument and construction integration. In G. Rijlaarsdam & E. Esperet (Series Eds.), J. Andriessen & P. Coirier (Vol. Eds.), *Foundations of argumentation text processing* (pp. 179 -201). Amsterdam: Amsterdam University Press.
- Baker, M. (2003). Computer-mediated interactions for the co-elaboration of scientific notions. In J. Andriessen, M. Baker, & D. Suthers (Eds.), *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments* (pp. 47-78). Norwell, MA: Kluwer Academic.
- Barlett, F. C. (1932). *Remembering*. Cambridge, England: The Cambridge University Press.
- Bean, T. W., & Steenwyk, F. L. (1984). The effect of three forms of summarization instruction on sixth graders' summary writing and comprehension. *Journal of Reading Behavior, 16*(4), 297-306.
- Berzonsky, M. D., & Sullivan, C. (1992). Social-cognitive aspects of identity style: Need for cognition, experiential openness, and introspection. *Journal of Adolescent Research, 7*, 140-155.
- Blumenfeld, P. (1992). Classroom learning and motivation: Clarifying and expanding goal theory. *Journal of Educational Psychology, 84*, 272-281.
- Brainerd, C. J., Reyna, V. F. (1992). The memory independence effect: What do the data show? What do the theories claim? *Developmental Review, 12*, 164-186.
- Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin, 119*, 197-253.
- Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*, 306-307.
- Cacioppo, J. T., Petty, R. E., & Morris, K. J. (1983). Effects of need for cognition on message evaluation, recall, and persuasion. *Journal of Personality and Social Psychology, 45*, 805-818.

- Chi, M. T. H. (2000). Self-explaining expository texts: The dual process of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 161–238). Mahwah, NJ: Erlbaum.
- Chi, M. T. H., Bassock, M., Lewis, M. W., Reimann, P. & Glaser, R. (1989). Self-Explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*(2), 145-182.
- Chinn, C. A., & Anderson, R. C. (1998). The structure of discussions that promote reasoning, *Teachers College Record*, *100*(2), 315-368.
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society For Information Science*, *41*, 391-407.
- De Vries, E., Lund, K., & Baker, M. (2002). Computer-mediated epistemic dialogue: Explanation and argumentation as vehicles for understanding scientific notions. *Journal of the Learning Sciences*, *11*, 63–103.
- Diakidoy, I. N., Kendeou, P., & Ioannides, C. (2003). Reading about energy: The effects of text structure in science learning and conceptual change. *Contemporary Educational Psychology*, *28*, 335 – 356.
- Duschl, R. A., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education*, *38*, 39–72.
- Ferretti, R. P., MacArthur, C. A., & Dowdy, N. S. (2000). The effects of an elaborated goal on the persuasive writing of students with learning disabilities and their normal achieving peers. *Journal of Educational Psychology*, *92*, 694–702.
- Fletcher, F. J. O., Danilovics, P., Fernandez, G., Peterson, D., & Reeder, G. D. (1986). Attributional complexity: An individual difference measure. *Journal of Personality and Social Psychology*, *51*, 875-884.
- Govier, T. (1987). *Problems in argument analysis and evaluation*. Providence, RI: Foris.
- Guzzetti, B. J. Williams, W. O., Skeels, S. A., & Wu, S. M. (1997). Influence of text structure on learning counterintuitive physics concepts. *Journal of Research in Science Teaching*, *34*(7), 701–719.
- Hatano, G., & Inagaki, K. (1991). Sharing cognition through collective comprehension activity. In L. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 331-348). Washington, DC: American Psychological Association.
- Haugtvedt, C. P., & Petty, R. E. (1992). Personality and persuasion: Need for cognition moderates the persistence and resistance of attitude changes. *Journal of Personality and Social Psychology*, *63*, 308-319.

- Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 1-28). Mahwah, NJ: Erlbaum.
- Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. Gregg & E. Steinberg (Eds.), *Cognitive processes in writing: An interdisciplinary approach* (pp. 3-30). Hillsdale, NJ: Erlbaum.
- Hayes, J. R., & Nash, J. G. (1996). On the nature of planning in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 29-56). Mahwah, NJ: Erlbaum.
- Herrenkohl, L. R., & Guerra, M. R. (1998). Participant structures, scientific discourse, and student engagement in fourth grade. *Cognition and Instruction, 16*, 432-473.
- Homo floresiensis. (2008). Retrieved May 2, 2008, from [http://en.wikipedia.org/wiki/Homo\\_floresiensis](http://en.wikipedia.org/wiki/Homo_floresiensis)
- Hynd, C. R. (2001). Refutational Texts and the Change Process. *International Journal of Educational Research, 35*, 699-714.
- Hynd, C. E., Alvermann, D. E., & Quian, G. (1997). Preservice elementary school teachers' conceptual change about projectile motion: Refutation text, demonstration, affective factors, and relevance. *Science Education, 81*, 1-27.
- Jenkins, J. R., Heliotis, J., Stein, M. L., & Haynes, M. (1987). Improving reading comprehension by using paragraph restatements. *Exceptional children, 54*, 54-59.
- Kardash, C. M., & Scholes, R. J. (1996). Effects of pre-existing beliefs, epistemological beliefs, and need for cognition on interpretation of controversial issues. *Journal of Educational Psychology, 88*, 260-271.
- Keller, J. M., & Kopp, T. W. (1987). An application of ARCS model of motivational design. In C. M. Reigeluth (Ed.), *Instructional design theories in action: Lessons illustrating selected theories and models* (pp. 289-320). Hillsdale, NJ: Erlbaum.
- Kintsch, W. (1988). The use of knowledge in discourse processing: A construction-integration model. *Psychological Review, 95*, 163-182.
- Kintsch, W. (1992). A cognitive architecture for comprehension. In H. L. Pick, P. van den Broek, & D. C. Knill (Eds.), *The study of cognition: Conceptual and methodological issues* (pp. 143-164). Washington, DC: American Psychological Association.
- Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist, 49*, 294-303.
- Kintsch, W., Vipond, D. (1979). Reading comprehension and reading ability in educational practice and psychological theory. In L. G. Nilsson (Ed.), *Perspectives of memory research* (pp. 325-366). Hillsdale, NJ: Erlbaum.

- Kintsch, W., & Young, S. (1984). Selective recall of decision-relevant information from text. *Memory and Cognition*, *12*(2), 112-117.
- Kuhn, D. (1991). *The skills of argument*. Cambridge, England: Cambridge University Press.
- Kuhn, D. (1992). Thinking as argument. *Harvard Educational Review*, *62*, 155–178.
- Laham, D., & Steinhart, D. (2003). Latent semantic analysis at the University of Colorado Boulder. Retrieved November, 5, 2008, from <http://lsa.colorado.edu>
- Lampinen, J. M., Leding, J. K., Reed, K.B., & Odegard, T. N. (2006). Global gist extraction in children and adults. *Memory*, *14*, 952-964.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211-240.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes*, *25*, 259-284.
- Limon, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning and Instruction*, *11*, 357–380.
- Linnenbrink, E. A., & Pintrich, P. R. (2002). Achievement goal theory and affect: An asymmetrical bidirectional model. *Educational Psychologist*, *37*, 69-78.
- Mannes, S., & Kintsch, W. (1987). Knowledge organization and text organization. *Cognition and Instruction*, *4*, 91-115.
- Maria, K., & MacGinitie, W. (1987). Learning from texts that refute the reader's prior knowledge. *Reading Research and Instruction*, *26*, 222–238.
- Marttunen, M. (1997). Argumentation course by electronic mail. *Scandinavian Journal of Educational Research*, *40*(1), 15-32.
- Marttunen, M., & Laurinen, L. (2001). Learning of argumentation skills in networked and face-to-face environments. *Instructional Science*, *29*, 127–153.
- Mason, L. (2001). Introducing talk and writing for conceptual change: A classroom study. *Learning and Instruction*, *11*, 305–329.
- McDaniel, M., & Donnelly, C. (1996). Learning with analogy and elaborative interrogation. *Journal of Educational Psychology*, *88*, 508-519.
- McNamara, D., Kintsch, E., Songer, N., & Kintsch, W. (1996). Are good texts always better? *Cognition and Instruction*, *14*, 1-43.
- Means, M. L., & Voss, J. F. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability, and knowledge levels. *Cognition and Instruction*, *14*, 139–179.

- Nesbit, C. J., Niu, H., & Mao, L. (2008). *Argumentation in education: A comprehensive review*. Unpublished manuscripts.
- Neuman, Y., & Schwarz, B. B. (2000). Substituting one mystery for another: The role of self-explanations in solving algebra word-problems. *Learning and Instruction, 10*, 203–220.
- Neuschatz, J. S., Lampinen, J., Preston, E. L., Hawkins, E. R., & Toglia, M. P. (2002). The effect of memory schemata on memory and the phenomenological experience of naturalistic situations. *Applied Cognitive Psychology, 16*, 687-708.
- Nussbaum, E. M. (2005). The effect of goal instructions and need for cognition on interactive argumentation. *Contemporary Educational Psychology, 30*, 286-313.
- Nussbaum, E. M., & Bendixen, L. M. (2003). Approaching and avoiding arguments: The role of epistemological beliefs, need for cognition, and extraverted personality traits. *Contemporary Educational psychology, 28*, 573-595.
- Nussbaum, E. M., & Kardash, C. M. (2005). The effects of goal instructions and text on the generation of counterarguments during writing. *Journal of Educational Psychology, 97*, 157–169.
- Nussbaum, E. M., & Schraw, G. (2007). Promoting argument-counterargument integration in students' writing. *The Journal of Experimental Education, 76*, 59-92.
- Nussbaum, E. M., & Sinatra, G. M. (2003). Argument and conceptual engagement. *Contemporary Educational Psychology, 28*, 384–395.
- Page-Voth, V., & Graham, S. (1999). Effects of goal setting and strategy use on the writing performance and self-efficacy of students with writing and learning problems. *Journal of Educational Psychology, 91*, 230-240.
- Petty, R. E., & Jarvis, B. G. (1996). An individual difference perspective on assessing cognitive processes. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 221-257). San Francisco: Jossey-Bass.
- Politzer, G., & Nguyen-Xuan, A. (1992). Reasoning about conditional promises and warnings: Darwinian algorithms, mental models, relevance judgements or pragmatic schemas. *Quarterly Journal of Experimental Psychology, 44A(3)*, 401-421.
- Rewey, K. L., Dansereau, D. F., Skaggs, L. P., Hall, R. H., & Pitre, U. (1989). Effects of scripted cooperation and knowledge maps on the processing of technical information. *Journal of Educational Psychology, 81*, 604-609.
- Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning & Individual Differences, 7*, 1-75.

- Reznitskaya, A., & Anderson, R. C. (2002). The argument schema and learning to reason. In C. C. Block & M. Pressley (Eds.), *Comprehension instruction: Research-based best practices* (pp.319-334). New York: The Guilford Press.
- Rinehart, S. D., Stahl, S. A., & Erickson, L. G. (1986). Some effects of summarization training on reading and studying. *Reading Research Quarterly, 21*(4), 422-438.
- Rumelhart, D. E. (1980). Schemata; The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading and comprehension* (pp. 33-58). Hillsdale, NJ: Erlbaum.
- Sadowski, C. J. (1993). An examination of the short Need for Cognition Scale. *Journal of Psychology, 127*, 451-454.
- Sadowski, C. J., & Gulgoz, S. (1992b). Internal consistency and test-retest reliability of the Need for Cognition Scale. *Perceptual and Motor Skills, 74*, 610.
- Salisbury-Glennon, J. D., & Stevens, R. J. (1999). Addressing preservice teachers' conceptions of motivation. *Teaching and Teacher Education, 15*, 741-752.
- Scardamalia, M., & Bereiter, C. (1986). Research on written composition In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 778-803). New York: Macmillan.
- Schwarz, B. B. (2003). Collective reading of multiple texts in argumentative activities. *International Journal of Educational Research, 39*, 133–151.
- Schwarz, B. B., & Glassner, A. (2003). The blind and the paralytic: Supporting argumentation in everyday and scientific issues. In J. Andriessen, M. Baker, & D. Suthers (Eds.), *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments* (pp. 277–260). Utrecht, the Netherlands: Kluwer Academic.
- Schwarz, B. B., Neuman, Y., & Biezuner, S. (2000). Two wrongs may make a right . . . if they argue together! *Cognition and Instruction, 18*, 461–494.
- Spotts, H. (1994). Evidence of a relationship between need for cognition and chronological age: Implications for persuasion in consumer research. *Advances in Consumer Research, 21*, 238-243.
- Tabachnik, B. G., & Fidell, L.S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson Education.
- Tolentino, E., Curry, L., & Leak, G. (1990). Further validation of the Short Form of the Need for Cognition Scale. *Psychological Reports, 66*,321-322.
- Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

- Venkatraman, M. P., Marlino, D., Kardes, F. R., & Sklar, K. B. (1990a). Effects of individual difference variables on response to factual and evaluative ads. *Advances in Consumer Research*, *17*, 761-765.
- Weary, G., & Edwards, J. A. (1994). Individual differences in causal uncertainty. *Journal of Personality and Social Psychology*, *67*, 308-318.
- Weaver, C. A., III, & Kintsch, W. (1990). Expository text. In R. Barr, M. L. Kamil, P. B. Rosenthal, & P. D. Pearson (Eds.), *Handbook of reading* (pp. 230-245). New York: Longman.
- Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology*, *67*, 1049-1062.
- West, C. K., Farmer, J. A., & Wolff, P. M. (1991). *Instructional design: Implications from cognitive science*. Englewood Cliffs, NJ: Prentice-Hall.
- Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology*, *91*, 301-311.
- Winne, P. H., & Hadwin, A. F. (2009). *nStudy: A web application for researching and promoting self-regulated learning* (version 1.01) [computer program]. Simon Fraser University, Burnaby, BC, Canada.

## **APPENDICES**



## Appendix 1

### Demography

Age:

Gender\*:  Male  Female

Highest Level of education\*:  high school  two-year diploma  Bachelor  
 Master  PhD

Major:

Have you taken any courses in anthropology\*:  Yes  No

First language:

Years of studying in English-speaking countries if English is not your first language:

Note: \* please tick the box that fits you

## Appendix 2

### The study instructions

In this experiment you will study information about ancient skeletons recently discovered in Indonesia. The purpose is to prepare for a test covering all the information. The more information you memorize the more successful you will be in this experiment. Your goal while working with the materials is to get the highest possible score on a recall test. To achieve your goal you should selectively tag information. You may use any of the tags provided.

The test will cover all types of information in the materials, including information that is difficult to remember, details, and examples. Here are some examples of the types of information you will be asked to recall.

Some information is difficult to remember because it is not familiar. Paying close attention to information that is **difficult to remember** may help you to recall it later.

A **detail** is a specific fact that relates to more general ideas. The following statement is an example of detail information.

“A study conducted in 1987 found higher levels of aggression among teenagers who watched violent television programs.”

An **example** is a specific instance of a general concept or idea. The following statement gives an example of the idea that teenagers sometimes copy television violence.

“The day after watching a violent television program in which a person was stabbed with a knife, a 14 year old boy in California hurt another boy with a knife.”

Please read the materials to study for a test. Tag information that will help you to remember during the test.

## Appendix 3

### The summarize instructions

In this experiment you will study information about ancient skeletons recently discovered in Indonesia. The purpose is to make a summary of the information. The better your summary the more successful you will be in this experiment. Your goal while working with the materials is to prepare an accurate and complete summary. To achieve your goal you should selectively tag information. You may use any of the tags provided.

Effective summaries contain the central concepts of the original text, leaving out the less important details. A summary often contains a main idea, key terms and explanations. Here are examples of the types of information you may use to prepare your summary.

The **main idea** of a passage is what it is all about. The following statement is an example of a main idea.

“Violent television programs increase aggression among teenagers.”

A **key term** is a word that represents an important concept. The following statement contains the key term vicarious learning.

“Vicarious learning occurs when someone copies the behaviour of another, such as when a person performs an aggressive act after seeing it on television.”

An **explanation** describes the cause or meaning of something. The following statement is an example of an explanation.

“Aggressive teenagers may prefer watching violent television programs, but that does not mean the programs are causing them to be aggressive.”

A good summary is a shorter version of the original text with nothing added. Do not go beyond the information in the original text by adding your own ideas, opinions or conclusions.

Notice that your goal is to write a summary. Tag information that will help you to build a good summary.

## Appendix 4

### The argue instructions

In this experiment you will study information about ancient skeletons recently discovered in Indonesia. The purpose is to make an argument supporting a particular theory about the origin of the skeletons. The better your argument the more successful you will be in this experiment. Your goal while working with the materials is to prepare an effective and convincing argument. To achieve your goal you should selectively tag information. You may use any of the tags provided.

Effective arguments support a position with reasoning. An argument often contains supporting claims, counterclaims, evidence, and rebuttals. Here are some examples of the types of information you may use to prepare your argument.

A **supporting claim** is a statement that supports the main position you are arguing for. It is logically consistent with that position. If you are trying to argue that violent television programs increase aggression among teenagers, then the following statement is an example of a supporting claim.

“Teenagers often learn by observing what others do and copying them.”

A **counterclaim** is a statement that opposes the main position you are arguing for. The following statement is an example of a counterclaim.

“Teenagers don’t copy the violence they see in television shows because they understand it is only entertainment and not real.”

**Evidence** is an observed event or scientific data used to support a position, a supporting claim or a counterclaim. The following statement is an example of evidence.

“A study conducted in 1987 found higher levels of aggression among teenagers who watched violent television programs.”

A rebuttal is a supporting claim or evidence that opposes a specific counterclaim. The following statement is an example of a rebuttal to the counterclaim given above.

“Research has found that teenagers who watch a specific violent action on television are more likely to perform that action within the following 24 hours than teenagers who do not see the violent action on television.”

A good argument supports a position using supporting claims and evidence. A good argument also identifies counterclaims and refutes them with rebuttals.

Please read the materials to build an argument supporting the position that the skeletons found by researchers are of a species different from humans living today. Tag information that will help you to build a strong argument.

## Appendix 5

### The free recall test

Write down everything you can remember from the material you studied in this experiment before doing the crossword puzzle. Please type your answers in the text box.

A large, empty rectangular box with a thin black border, intended for the user to type their free recall answers.

## Appendix 6

### The multiple-choice test

1. What is the main idea of the article?
  - A. How the Hobbit skeletons were discovered
  - B. A comparison of the Hobbits to other hominids
  - C. Competing theories about the origins of the Hobbits
  - D. The effects of microcephaly and island dwarfing
  
2. How does the small body of the Island Fox help it survive on the Channel Islands of California?
  - A. Small bodies are more easily hidden from predators.
  - B. Small bodies can move more quickly to get food.
  - C. Small bodies have greater immunity to disease.
  - D. Small bodies require less food to survive.
  
3. Which of the following does **not** support the theory that Hobbits evolved smaller bodies?
  - A. Flores Island is separated from neighbouring islands by a deep channel.
  - B. The Hobbit skulls showed evidence of microcephaly.
  - C. Dwarf Stegodons were found nearby the skeletons of the Hobbits.
  - D. The food resources on Flores Island are limited.
  
4. The purpose of mentioning Laron syndrome is to
  - A. suggest that the small bodies of the Hobbits were caused by a disease.
  - B. cast doubt on the theory that the Hobbits were microcephalic.
  - C. explain how island dwarfing could have occurred.
  - D. compare different types of disease that cause short stature.
  
5. The purpose of telling the stories about the Ebu Gogo and Orang Pendek is to give information
  - A. about the culture of Indonesia and Flores Island.
  - B. consistent with the theory that a genetically diseased population may have existed for many years.
  - C. consistent with the theory that Homo erectus once inhabited Sumatra and Flores Island.

- D. consistent with the theory that the Hobbits were a species separate from Homo sapiens.
6. Which of the following statements is **not** consistent with the idea that the Hobbits were a separate homo species having small bodies and skulls?
- A. The brain to body mass ratio of the Hobbits is between that of Homo erectus and chimpanzees.
  - B. The bones in the Hobbits' wrists were indistinguishable from a modern day African ape.
  - C. Other animal bones found on Flores Island were much smaller than the animals' normal sizes.
  - D. Except for the difference in size, the Hobbits were very similar to Homo erectus.
7. Which of the following would most strongly indicate that the Hobbits are Homo sapiens?
- A. Considering the normal brain to body mass ratio in the genus Homo,  $400 \text{ cm}^3$  is too large for a one meter tall homo.
  - B. The Hobbit skulls showed odd formations of teeth and the absence of a chin.
  - C. The Hobbits' brain shape is similar to that of a microcephalic brain.
  - D. Skeletons of human-like creatures with small bodies have been found on other small islands.
8. The strongest argument against the theory that the Hobbits are descendants of Homo erectus would be
- A. The Hobbits became extinct only recently.
  - B. The Hobbits' skeletons were considerably smaller than that of Homo erectus.
  - C. The estimated brain to body mass ratio of the Hobbits' skeletons lies between that of Homo erectus and chimpanzees.
  - D. No remains of Homo erectus or transitional forms between Homo erectus and the Hobbits have been found on Flores Island.
9. A good argument against the theory that the Hobbits were microcephalic Homo sapiens is
- A. They showed advanced behaviours.
  - B. The bones in the Hobbits wrists look like an African ape.
  - C. The Hobbits resemble Homo erectus, which became extinct 100,000 years ago.
  - D. All of the above.
10. What caused the inconsistency in the results of the CT scan?

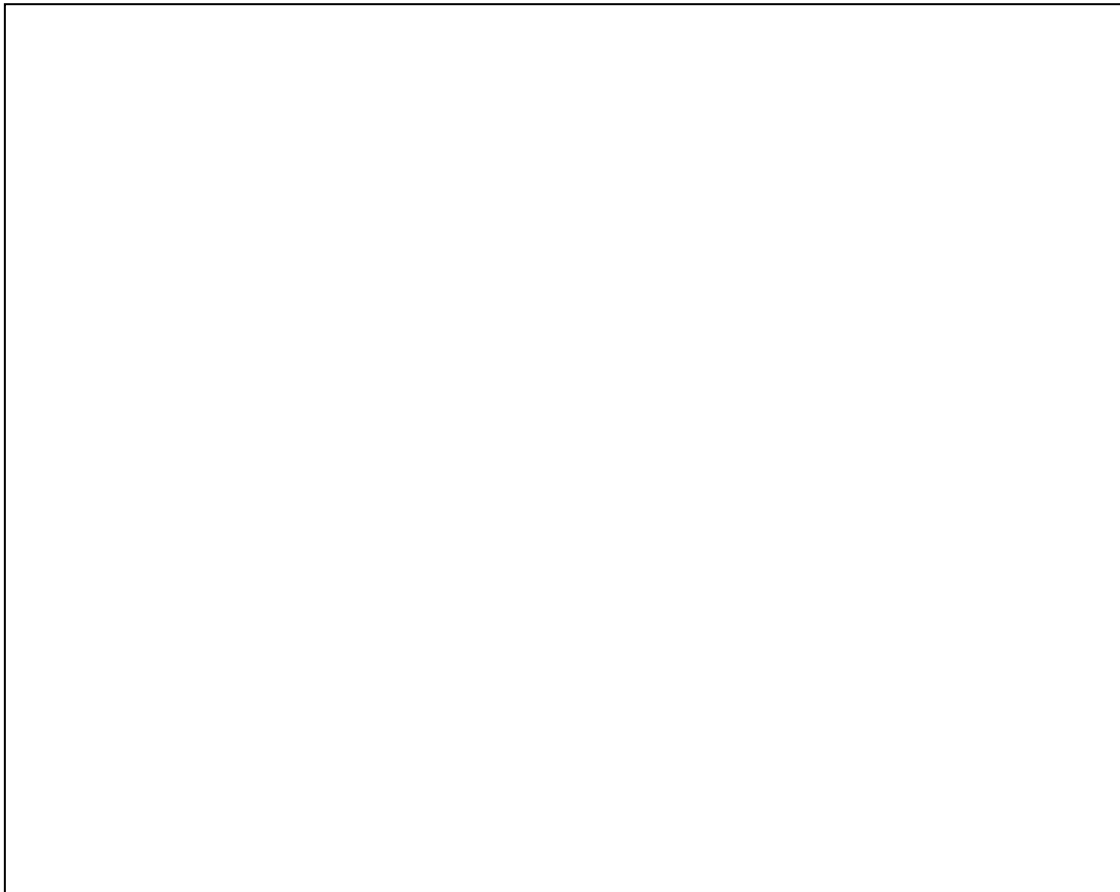
- A. The limited number of Hobbit specimens and the absence of Hobbit DNA.
  - B. The high variability of the microcephalic brain shapes and the absence of Hobbit DNA.
  - C. The limited number of the Hobbit specimens and the high variability of the microcephalic brain shapes.
  - D. The limited number of Hobbit specimens, the high variability of the microcephalic brain shapes and the absence of Hobbit DNA.
11. The scientists, who were looking for evidence on Flores Island of the original human migration from Asia to Australia, were surprised by the discovery of the skeletons of the Hobbits because
- A. they found that that the Hobbits migrated to Australia before modern humans.
  - B. no one had ever seen a homo species as small as the Hobbits.
  - C. they discovered the Hobbits were related to the modern aboriginal people of Indonesia.
  - D. despite their small size, the Hobbits were apparently able to cross the wide sea channel to get onto the island.
12. The Hobbits had a brain size between Homo erectus and chimpanzees, but also showed advanced behaviours. Given what you know about the theories of the Hobbits' origins, choose the most reasonable explanation of that fact.
- A. The Hobbits were microcephalic Homo sapiens who were still capable of the intelligence required to create sophisticated tools.
  - B. The Hobbits were the ancestors of Homo erectus.
  - C. The Hobbits kept the advanced behaviours of their Homo erectus ancestors while evolving a smaller brain size.
  - D. The Hobbits developed advanced behaviours earlier than other Homo species.
13. The materials you read had no conclusion. If you were the writer, which of the following conclusions do you think would be most appropriate for it?
- A. Most anthropologists in this area believe that it is more reasonable to put the Hobbits into a separate homo species.
  - B. It is too early to say which theory is right. Further research is needed to find more evidence.
  - C. Some scientists believe that none of the existing theories are correct and therefore a new theory needs to be developed.
  - D. Recently, new research has begun which compares the skull shapes of the Hobbits, modern microcephalics and Homo erectus. This new study will tell us more information about the origins of the Hobbits.



## Appendix 7

### Evaluation question

You have been invited to write a one-page essay for an issue of National Geographic Magazine dealing with human evolution. Your job now is to write a draft of the essay. The one-page draft should critically evaluate the competing ideas about the origins of the Hobbits. Be sure to identify the theory you believe is most accurate and help the readers to understand your choice.



## Appendix 8

### Correlations

	1	2	3	4	5	6	7	8	9	10	11	12
1. Group	—	-.03	.00	.21	-.29	.54	.04	-.02	.29	.35	.03	.20
2. Education level		—	-.27	.09	-.02	.02	.10	-.23	-.19	-.21	-.35	-.37
3. English Experience			—	.15	.02	.17	.14	.39	.28	.29	.38	.47
4. Need for Cognition				—	-.03	.18	.29	.22	.24	.23*	.18	.21
5. Actions					—	.22	.26	.32	.24	.17	.35	.25
6. Argument Tags						—	.14	.12	.44	.43	.24	.31
7. Multiple choices							—	.47	.35	.40	.20	.23
8. Recall test								—	.58	.59	.76	.55
9. Reasons									—	.86	.58	.69
10. Correct and strong reasons										—	.56	.62
11. Word count of the recall test											—	.72
12. Word count of the essays												—