

THE EVALUATION OF ARGUMENT MAPPING AS A LEARNING TOOL

Christopher Peter Dwyer, B.A., Ph.D.

Thesis submitted to the National University of Ireland, Galway in fulfilment of the requirements for the Degree of Doctor of Philosophy in Psychology



School of Psychology,
National University of Ireland, Galway

October 2011

Supervised By:

Dr. Michael J. Hogan, BA, CNA, PhD
School of Psychology, National University of Ireland, Galway.

&

Dr. Ian Stewart, BA, HDipPsych, PhD
School of Psychology, National University of Ireland, Galway.

Table of Contents

Table of Contents.....	ii
Funding.....	vi
Acknowledgments.....	vii
Publications & Research Presented.....	viii
Abstract.....	x
List of Tables.....	xi
List of Figures.....	xiii
List of Appendices.....	xv
List of Acronyms.....	xvi
Preface.....	1
Chapter 1: Thinking Processes in Instructional and Educational Settings.....	10
1.1 Frameworks for Thinking.....	10
1.1.1 Bloom’s Taxonomy of Educational Objectives (1956).....	14
1.1.2 Romiszowski’s (1981) Framework for Knowledge & Skills.....	18
1.1.3 Anderson & Krathwohl’s (2001) Revision of Bloom’s Taxonomy.....	24
1.1.4 Marzano’s (2001) New Taxonomy of Educational Objectives.....	27
1.1.5 Summary of the Frameworks for Thinking.....	29
1.2 Memory.....	32
1.2.1 Baddeley’s (2000) Model of Working Memory.....	33
1.2.1.1 The Slave Systems.....	34
1.2.1.2 The Episodic Buffer.....	38
1.2.2 Long-term Memory.....	39
1.2.2.1 Schemas.....	40
1.2.2.2 Comprehension as LTM.....	43
1.3 Metacognition.....	45
1.3.1 Critical Thinking.....	48
1.3.1.1 Analysis.....	51
1.3.1.2 Evaluation.....	53
1.3.1.3 Inference.....	54
1.3.1.4 Reflective Judgment.....	55
1.4 Cognitive Load.....	62
1.4.1 Intrinsic Cognitive Load.....	64
1.4.2 Extraneous Cognitive Load.....	68
1.4.2.1 The Problematic Nature of Text-Based Reading.....	70
1.5 Summary.....	71

Chapter 2: Review of the Literature Examining Strategies Aimed at Improving Memory, Comprehension & Critical Thinking.....	74
2.1 Improving Memory and Comprehension Ability.....	74
2.1.1 Hierarchical Summarisation.....	77
2.1.2 Graphic Organisers.....	80
2.1.2.1 Mapping Strategies.....	82
2.2 Improving Critical Thinking Performance.....	85
2.3 Summary of the Previous Research.....	101
Chapter 3: Argument Mapping.....	103
3.1 Argumentation.....	103
3.2 What is an Argument Map?	106
3.3 A Brief History of Argument Maps.....	108
3.4 Why use Argument Maps?	111
3.5 Previous Research Conducted on Argument Mapping.....	117
3.6 Rationale for the Current Research.....	124
Chapter 4: Study 1: A Multi-Experiment Analysis of the Effects of Argument Map Studying & Construction on Students' Memory and Comprehension.....	130
4.1 Purpose.....	130
4.2 Experiment 1.....	135
4.2.1 Method.....	136
4.2.1.1 Design.....	136
4.2.1.2 Participants.....	137
4.2.1.3 Materials & Measures.....	137
4.2.1.4 Procedure.....	138
4.2.2 Results.....	138
4.2.3 Discussion of Experiment 1.....	140
4.3 Experiment 2.....	144
4.3.1 Method.....	146
4.3.1.1 Design.....	146
4.3.1.2 Participants.....	146
4.3.1.3 Materials & Measures.....	147
4.3.1.4 Procedure.....	147
4.3.2 Results.....	148
4.3.3 Discussion of Experiment 2.....	149
4.4 Experiment 3.....	152
4.4.1 Method.....	154
4.4.1.1 Design.....	154
4.4.1.2 Participants.....	154
4.4.1.3 Materials & Measures.....	155
4.4.1.4 Procedure.....	156
4.4.2 Results.....	156
4.4.3 Discussion of Experiment 3.....	159
4.5 Experiment 4.....	161
4.5.1 Method.....	162
4.5.1.1 Design.....	162

4.5.1.2. Participants.....	162
4.5.1.3 Materials & Measures.....	162
4.5.1.4 Procedure.....	163
4.5.2 Results.....	164
4.5.3 Discussion of Experiment 4.....	165
4.6 Discussion of Study 1.....	166
4.6.1 Interpretation of Results.....	166
4.6.2 Limitations & Future Research.....	170
Chapter 5: Study 2: The Effects of Argument Mapping on Critical Thinking & Reflective Judgment Performance.....	175
5.1 Purpose.....	175
5.2 Method.....	180
5.2.1 Design.....	180
5.2.2 Participants.....	182
5.2.3 Materials & Measures.....	182
5.2.4 Procedure.....	189
5.3 Results.....	194
5.3.1 Group differences in CCTST performance.....	194
5.3.2 Group differences in LRJA performance.....	197
5.3.3 Group differences on the CCTDI.....	198
5.3.4 Correlations.....	198
5.4 Discussion of Study 2.....	200
5.4.1 Interpretation of Results.....	200
5.4.2 Limitations & Future Research.....	206
Chapter 6: Study 3: The Effects of Argument Mapping on Critical Thinking Performance in e-Learning Environments.....	215
6.1 Purpose.....	215
6.1.1 Improvements in Design & Implementation from Study 2.....	215
6.1.1.1 e-Learning.....	216
6.1.1.2 Sample Size.....	217
6.1.1.3 Feedback.....	218
6.1.1.4 Engagement in AM Training.....	219
6.1.1.5 Method of Assessment.....	220
6.2 Method.....	223
6.2.1 Design.....	223
6.2.2 Participants.....	224
6.2.3 Materials & Measures.....	224
6.2.4 Procedure.....	227
6.3 Results.....	234
6.3.1 Group differences in critical thinking, need for cognition, and motivation.....	234
6.3.2 High-engagement versus low engagement differences in the AM group.....	238
6.3.3 Correlations.....	239
6.3.4 Focus Group.....	241
6.4 Discussion of Study 3.....	241
6.4.1 Interpretation of Results.....	241

6.4.2 Limitations & Future Research.....	251
6.4.3 Summary & Conclusion.....	256
Chapter 7: General Discussion.....	258
7.1 Interpretation of Results.....	258
7.1.1 Memory & Comprehension.....	258
7.1.2 Critical Thinking.....	263
7.1.3 Dispositional Factors towards Thinking.....	267
7.2 Future Research.....	269
7.3 Broader Implications of Argument Mapping.....	278
7.4 Summary & Conclusion.....	281
References.....	283

Funding

The work associated with this thesis was made possible by a Doctoral Research Scholarship awarded to the author by the Irish Research Council for Humanities and Social Sciences (IRCHSS).



The author would like to extend his sincere gratitude to the IRCHSS for the council's continued support of this work.

Acknowledgments

This PhD thesis is dedicated to the eternal memory of Tony Nappi – the smartest man I have ever known. He was a brilliant teacher and also a brilliant friend.

First and foremost, I would like to thank my parents, Peter and Maura Dwyer, for their encouragement, guidance, undying support and most importantly, their love. I would also like to extend to them my appreciation for the thousands, upon thousands, of dollars and euros spent on my education. Their financial investment is a testament to the faith they have had in me, for which I am also forever grateful. Mom and Dad, it was your driving encouragement in my grammar school years, your guidance in my high school years and your loving support in my college years that I credit for being where and who I am today. The phrase ‘thank you’ simply cannot emphasise the amount of gratitude that I would like to express to you. I love you both dearly and I thank you whole-heartedly for everything I have and for everything that I am.

I would also like to extend my appreciation and thanks to my primary supervisor, Dr. Michael Hogan, for his wonderful advice, guidance, constant encouragement and his faith in me, for which I am truly grateful. I would also like to thank my co-supervisor, Dr. Ian Stewart for his invaluable contributions and support. I would further like to thank Dr. AnnMarie Groarke and all the School of Psychology staff for their aid and support over the years.

I would also like to express my gratitude and deepest appreciation to Mr. Declan Coogan for always going the extra mile. From helping me ‘check the parking’ to ‘un-crashing’ my computer and everything in between - especially your good will and friendship, I am forever in your debt.

A debt of gratitude is also owed to Vincent McDermott. Thank you for *always* being there, for all your help, for all your support and for being a truly great friend.

Furthermore, I would like to give special thanks to my very good friends who aided me in my research throughout the years, especially Andrea Gibbons, Dr. Sarah Roddy, Rachel Carey, Dr. Michael Keane, Brian Slattery, Audrey Hanly, Dr. Catriona Martyn, Dr. Darragh McDermott and Triona Tammemagi. I would also like to thank all my other wonderful friends for their support and friendship over the years, especially Travis, Ronan, Lorraine, Aidan, Micheál, Dee, Rose, Davina, Sinéad, Niall, Cormac, Niamh, Rob and Paul. For those not mentioned, please know that I am, and will always be, unreservedly grateful for your help, support and friendship.

Last, but truly the furthest from least, I would like to thank Lise Hanly. For once in my life, I am near lost for words, because finding the right ones to say how thankful I am for you is near impossible. Nevertheless, I’ll try my best. Lise, I thank you and appreciate you for not only what you have done for me, but also for everything you are – beautiful, intelligent, funny and most importantly, my best friend. From the bottom of my heart, thank you so very much for your support and your love over the past few years. It has not been an easy journey. However, you had faith in me and managed to see the light at the end of the tunnel, even at times when I could not. I am eternally grateful for this and everything you have done for me – both the small and the large: helping me with grading papers and referencing, cooking and cleaning, not to mention chauffeuring! I love you with all my heart. I appreciate everything you do, how you make me feel and again, most importantly, everything you are.

Publications & Research Presented

Publications:

Dwyer, C.P., Hogan, M.J., & Stewart, I. (2010). The evaluation of argument mapping as a learning tool: Comparing the effects of map reading versus text reading on comprehension and recall of arguments. *Thinking Skills and Creativity*, 5, 1, 16-22.

Dwyer, C.P., Hogan, M.J., & Stewart, I. (2011). The promotion of critical thinking skills through argument mapping. In C.P. Horvart & J.M. Forte (Eds), *Critical Thinking*. Nova Science Publishers, New York.

Dwyer, C.P., Hogan, M.J., & Stewart, I. (2011). An evaluation of argument mapping as a method of enhancing critical thinking performance in e-learning environments. *Metacognition and Learning*. Submitted.

Dwyer, C.P., Hogan, M.J., & Stewart, I. (2011). An examination of the effects of argument mapping on students' memory and comprehension performance. *Thinking Skills & Creativity*. Submitted.

Dwyer, C.P., Hogan, M.J., & Stewart, I. (2011). The evaluation of argument mapping as a method of enhancing critical thinking and reflective judgment performance. In preparation.

Conference Presentations:

Dwyer, C.P. (2007). The evaluation of argument mapping as a learning tool. *29th Annual Psychological Society of Ireland Student Conference*, Dublin, Ireland, 29th & 30th March 2007.

Dwyer, C.P., Hogan, M.J. & I. Stewart (2007). Argument mapping with Rationale™: The evaluation of argument mapping as a learning tool. *5th Annual Conference on Teaching & Learning (CELT)*, Galway, Ireland, 7th & 8th June 2007.

Dwyer, C.P. (2008). Argument mapping increases recall performance: The evaluation of argument mapping as a learning tool. *30th Annual Psychological Society of Ireland Student Conference*, Limerick, Ireland, 4th & 5th April 2008. (Awarded and recommended).

Dwyer, C.P. (2010). Using argument mapping to promote critical thinking skills. *21st National Conference on Artificial Intelligence and Cognitive Science*, Galway, Ireland, 30th August – 1st September 2010.

Dwyer, C.P., & Hogan, M.J. (2011). Halpern critical thinking assessment predicts real-world outcomes of critical thinking. *15th International Conference on Thinking*, Belfast, Northern Ireland, 20th – 24th June 2011.

Seminar Presentations:

Dwyer, C.P. (2008). The evaluation of argument mapping as a learning tool: The effects on recall memory and comprehension. *NUI Galway School of Psychology Research Seminar Series*, Galway, Ireland, 25th April 2008.

Dwyer, C.P. (2009). Critical thinking: What is it? *NUI Galway School of Psychology Research Seminar Series*, Galway, Ireland, 3rd April 2009.

Abstract

Argument mapping (AM) is a method of visually diagramming arguments using a 'box and arrow' format with the aim of simplifying the reading of an argument structure and facilitating the assimilation of core statements and relations. The overall aim of the current programme of research was to evaluate the use of AM as a learning tool. Over the course of three studies, this research examined the effects of AM, in comparison with other traditional, educational methods, on immediate recall, delayed recall, comprehension and critical thinking (CT). Study 1 involved four experiments. The collective findings from these four experiments suggest that AM reading and construction can facilitate better immediate recall of propositions from arguments when compared with more traditional learning strategies, such as text-reading and text-summarisation. Study 1 experiments revealed that when compared with traditional text-based study materials, AM reading significantly enhanced the immediate recall of arguments, regardless of (1) the presence or absence of colour to demarcate reasons and objections in AMs, (2) the environmental setting in which AMs were studied and (3) the study topic used in the experiment. Results also revealed that those who actively learned through AM and hierarchical outline (HO) construction performed significantly better on immediate recall testing than those who actively learned through text-summarisation. Study 2 compared the effects of a six-week AM-infused CT training course with those of a HO-infused CT training course and a no-CT training control condition. Study 2 findings revealed that participants in the AM training group performed significantly better on inductive reasoning than controls, as did the HO training group. When analysed together, the CT training attendees (i.e. both AM and HO groups combined) outperformed the control group on the CT skills of analysis, evaluation and inductive reasoning. Study 3 examined the effects of an AM-infused CT e-learning course, in comparison with a no-CT course control group, on measures of CT ability. Results from Study 3 revealed that those who participated in the AM-infused CT training condition outperformed those in the control group on overall CT, argument analysis and verbal reasoning. Study 3 results also revealed that performance on overall CT and all CT sub-scales (i.e. hypothesis testing, argument analysis, verbal reasoning, assessing likelihood and uncertainty, and problem-solving) of those in the AM-infused CT group were significantly enhanced from pre-to-post-testing. Overall, the results suggest that AM is an efficacious learning method, as it was shown to facilitate both recall and CT ability. Empirical and theoretical implications of these results and future research possibilities are discussed.

List of Tables

Table 1.1: Past Definitions and Descriptions of Critical Thinking.....	50
Table 1.2: Core CT Skills According to the Delphi Report (adapted from Facione, 1990).....	52
Table 1.3: The Reflective Judgment Model (Adapted from King & Kitchener, 2002).....	58
Table 2.1: Topics Assessed by the Ennis-Weir Critical Thinking Essay Test (adapted from Ennis & Weir, 1985).....	87
Table 2.2: Critical Thinking skills taught to students in Hitchcock (2004).....	94
Table 2.3: Solon’s Conceptualisation of Critical Thinking.....	96
Table 2.4: Paul’s Elements of Reasoning (adapted from Paul, 1993).....	97
Table 3.1: Course outline used in Butchart et al. (2009).....	123
Table 4.1: Core manipulations and Hypotheses associated with each of the Four Memory Experiments.....	131
Table 4.2: Descriptive statistics for Experiment 1.....	139
Table 4.3: Experiment 1 MANCOVA Summary.....	140
Table 4.4: Descriptive statistics for Experiment 2.....	148
Table 4.5: Experiment 2 MANCOVA Summary.....	149
Table 4.6: Descriptive statistics for Experiment 3.....	157
Table 4.7: Experiment 3 ANCOVA Summary.....	158
Table 5.1: Example of a concept as it changes across the complexity orders (adapted from Hogan & Stein, 2010).....	187
Table 5.2: Critical Thinking Course Outline	192
Table 5.3: Means and standard deviations for the three groups on the CCTST, LRJA & CCTDI.....	195
Table 5.4: ANCOVA Summary for Study 2.....	198
Table 5.5: Correlations Among CT Skills and Disposition at Pre-testing and Post-testing.....	199

Table 6.1: e-Learning CT Course Outline.....	230
Table 6.2: Means and standard deviations for CT Performance by Condition.....	235
Table 6.3: ANOVA Summary for Study 3.....	237
Table 6.4: Correlations Among CT performance, Need for Cognition and Motivation Sub-scales at Pre-testing and Post-testing.....	240

List of Figures

Figure 1.1: Bloom’s Taxonomy (1956).....	16
Figure 1.2: Romiszowski’s (1981) Skill Cycle (adapted from Moseley et al., 2005).....	19
Figure 1.3: Bloom’s Taxonomy and Anderson and Krathwohl’s (2001) Revision.....	25
Figure 1.4: Marzano’s (2001) ‘New’ Taxonomy of Educational Objectives.....	28
Figure 1.5: Baddeley’s (2000) Model of Memory.....	34
Figure 1.6: Mayer’s (1997) Model of the Integration of Visual & Verbal Information.....	36
Figure 1.7: Expert Schema for Principles of Mechanics (Chi, Glaser & Rees, 1982).....	42
Figure 1.8: Working Memory as a buffer between informational input and storage in LTM as knowledge.....	45
Figure 1.9: The Interdependencies Among Critical Thinking Skills and Reflective Judgment.....	60
Figure 1.10: The proposed system of cognitive processes for educational objectives used in this research.....	62
Figure 1.11: The effect of Cognitive Load on Thinking Processes.....	63
Figure 2.1: An example of an outline (hierarchical summary) as used in Taylor & Beach (1984, p. 139).....	78
Figure 2.2: An example of a graphic organiser as used in Robinson & Kiewra (1995).....	82
Figure 2.3: An example of a mapping strategy, as used in Berkowitz (1986, p. 165).....	84
Figure 3.1: An example of Toulmin’s Model of Mapping.....	104
Figure 3.2: An example of an argument map using Rationale™ (van Gelder, 2007).....	107
Figure 3.3: An example of Whately’s Model of Mapping.....	109
Figure 3.4: An excerpt from Horn’s Map “Can Computers Think?”.....	110
Figure 3.5: An example of a Concept Map created using QuestMap™ (Carr, 2003) in Reed, Walton & Macagno (2007).....	114

Figure 3.6: John H. Wigmore’s (1913) Classically used Concept Mapping Strategy - An example of a concept map using numbers as reference points.....	115
Figure 3.7: An example of a Mind Map (Budd, 2004).....	116
Figure 3.8: An example of a branch from an argument map.....	128
Figure 4.1: Immediate and Delayed Memory performance in text reading and AM reading groups.....	159
Figure 5.1: Inductive and Deductive Reasoning problems on the CCTST Form 2000 (adapted from Facione et al., 2002).....	185
Figure 5.2: Example of how a response to an LRJA item might be mapped for purposes of scoring.....	186
Figure 6.1: Question 21 on the HCTA (of the Problem-Solving Sub-Scale) with Scoring Protocol (Halpern, 2010).....	226

List of Appendices

Appendix A: Memory Test Scoring Manual	318
Appendix B: Study Materials & Tests Used in Study 1.....	327
Appendix C: Participant Consent Form & Study Information.....	354
Appendix D: Customised Student Questionnaire, Exercise Handouts and DVD Recordings of Lectures & Exercises Used in Study 2.....	356
Appendix E: e-Learning Lectures, Exercises and an Example of Feedback Provided to Participants in Study 3.....	364
Appendix F: Programme of Questions used to Guide the Focus Group Interview in Study 3.....	368

List of Acronyms

- AM(s): Argument Mapping / Argument Map(s)
- ANCOVA: Analysis of Covariance
- ANOVA: Analysis of Variance
- CCTDI: California Critical Thinking Dispositions Inventory
- CCTST: California Critical Thinking Skills Test
- CCTT: Cornell Critical Thinking Test
- CT: Critical Thinking
- EWCTET: Ennis-Weir Critical Thinking Essay Test
- HCTA: Halpern Critical Thinking Assessment
- HO(s): Hierarchical Outlining / Hierarchical Outline(s)
- MANCOVA: Multivariate Analysis of Covariance
- MSLQ: Motivated Strategies towards Learning Questionnaire
- RJ: Reflective Judgement
- WGCTA: Watson-Glaser Critical Thinking Appraisal

PREFACE

“It had always been necessary to read a mountain of books, take folders of notes and try to form a mental picture of the matter. The latter is what, at the end of the day, we try to do with any argument we enter into. This is why argument mapping is a breakthrough.”

- Monk (2001, p.3)

Psychology has long influenced our thinking about teaching and learning (Folsom-Kovarik et al., 2010). Both cognitive and educational psychology are at the forefront of this influence given that the relationship between the success of teaching and learning is dependent upon students’ cognitive processing in educational settings (Romiszowski, 1981; Sweller, 1999; 2010). In order for students in school or university to achieve their academic requirements, it is both important and often necessary for them to use different cognitive processes to acquire knowledge from a range of sources, including textbooks, didactic instruction, class notes, and websites. An important goal for teachers, therefore, must be to aid students in their acquisition of knowledge. Such aid can be supplied through, for example, the use of teaching strategies that help improve students’ memory (Sweller, 1999) and comprehension (Meyer, Brandt & Bluth, 1980).

At the same time, it is often argued that higher-order forms of thought, including critical thinking and metacognitive self-regulation of thinking and learning processes, need to be cultivated in the classroom to facilitate both the acquisition and application of knowledge (Folsom-Kovarik et al., 2010; Huffaker & Calvert, 2003; U.S. National Research Council, 2002). Bloom (1956) describes a hierarchy of learning objectives in this context, where teachers and students seek to develop their memory, comprehension, analysis, synthesis and evaluation skills. Designing and

evaluating educational tools and strategies that facilitate the teaching and learning of skills that map onto this hierarchy of learning objectives is an area of research and development where educational and cognitive psychologists can continue to work together with teachers and students to advance our collective knowledge. The research presented in the current thesis seeks to advance our understanding of the utility of one increasingly popular educational tool - argument mapping. As outlined below, this thesis examines the efficacy of argument mapping in facilitating students' memory, comprehension, analysis, evaluation, inference and reflective judgement skills.

Beyond Didacticism

Part of the impetus for this research was a simple observation - that too much class and study time in school and university is devoted to didactic instruction and reading of textbooks while not enough time is devoted to actively analysing and evaluating knowledge as it is acquired and constructed (Hogan, 2006). While traditional means of learning used by students to assimilate knowledge (e.g. rote learning) may be helpful for short-term educational goals, more meaningful learning skills (Good & Brophy, 1986), such as critical thinking, are necessary for deeper comprehension, broader application, and comprehensive synthesis of knowledge across domains (Halpern, 2003a; Darling-Hammond, 2008; King & Kitchener, 2004; King, Wood & Mines, 1990; Kuhn, 1991; 1999; Sweller, 1999).

Furthermore, the development of instructional strategies aimed at improving meaningful learning skills such as critical thinking is necessary because, in today's world, where there is an exponential increase in the annual output of scientific knowledge, it is not only the ability to draw upon knowledge, but also the capacity to engage in enquiry and constructively solve problems (Darling-Hammond, 2008). For example, it is estimated that 500,000 times the volume of information contained in the

U.S. Library of Congress print collection was created in 2002 alone; and more amazingly, from the years 1999 to 2002, the amount of new information created equalled the amount of information previously developed throughout the history of the world (Varian & Lyman, 2003). It is further estimated that the creation of new information is doubling every two years (Jukes & McCain, 2002). Consequently, not only is successful learning reliant on students' attainment of knowledge, but also on student's ability to adapt both to new information and to new situations. Thus, "nations around the world are reforming their school systems to meet these new demands by revising curriculum, instruction and assessment, in order to support the critical thinking skills necessary in the 21st century - skills needed for framing problems; seeking and organising information and resources; and working strategically with others to manage and address dilemmas and create new products" (Darling-Hammond, 2008 p. 2).

The teaching of critical thinking skills in higher education has been identified as an area that needs to be explored and developed (Association of American Colleges & Universities, 2005; Australian Council for Educational Research, 2002; Higher Education Quality Council, 1996). Such skills are vital in educational settings because they allow students to go beyond simply memorising information, to actually gaining a more complex understanding of the information being presented to them (Halpern, 2003a). Critical thinking skills are not only important in the academic domain, but also in social and interpersonal contexts where adequate decision-making and problem-solving are necessary on a daily basis (Ku, 2009). Good critical thinkers are more likely to get better grades and are often more employable as well (Holmes and Clizbe, 1997; National Academy of Sciences, 2005).

Argument Mapping

The development of knowledge, comprehension, and critical thinking skills, including analysis, evaluation, inference and reflective judgment, can derive from many sources. Teachers often impart knowledge to students using text-based presentation of information. However, text-based presentation of information may not be as conducive to memory, comprehension and critical thinking development as is often implicitly assumed. Traditionally, in western culture, propositional knowledge is presented in linear left to right lines of text with one sentence following another, one paragraph following another, etc., often with relatively few cues to facilitate recognition of the logical structure of the text. Traditional text-based reading in classroom settings may result in significant cognitive load for students. Cognitive load refers to the demands placed on working memory resources during information processing; and high levels of cognitive load while reading may impede memory, comprehension and critical thinking (Sweller, 1988; 1999; 2010; van Gelder, 2003).

On the other hand, a core, often untapped feature of human intelligence is the ability to process large, integrated chunks of visual-spatial information (Gardner, 1985; Kosslyn, 1980; Mayer, 1997; 2003; 2005); and it is often possible to translate text-based information into a visual representation that may much better facilitate use and development of these visual-spatial and graphicacy skills. Argument mapping is one such method of visually representing arguments using a 'box and arrow' diagrammatic format, with the aim of simplifying the reading of an argument structure and facilitating the assimilation of core statements and relations. Argument mapping has been developed with the explicit intention to lessen cognitive load and facilitate both the learning and the cultivation of critical thinking skills (van Gelder, 2000; 2003). In a standard argument map constructed using *Rationale*TM software

(van Gelder, 2007), boxes are colour-coded to indicate the nature of propositions (e.g. reasons, objections, rebuttals) and arrows are labelled so as to specify the nature of the relationship between the propositions (e.g. *but*, *because* or *however*). Argument mapping may facilitate the ability to process large, integrated chunks of propositional information using a visual-spatial form of representation, which may subsequently allow for the enhancement of memory, comprehension and critical thinking ability. The core aim of the current research was to test this claim, specifically, by comparing argument mapping as a means of facilitating learning with a number of alternative, more traditional educational strategies (i.e. text-reading, text summarisation and hierarchical outlining). Argument mapping was compared with these other strategies across three experimental studies: Study 1 focused on the learning outcomes of memory and comprehension; Study 2 focused on the learning outcomes of critical thinking and reflective judgement in a classroom setting; and Study 3 focused on the learning outcome of critical thinking in an online e-learning environment.

Thesis Structure

The structure of the thesis is as follows. Chapter 1 introduces and defines a number of key learning outcomes that educational methods should positively affect. These outcomes are memory, comprehension and critical thinking; including the critical thinking sub-skills of analysis, evaluation, inference and reflective judgement. The rationale for the choice of these specific outcomes is detailed in Chapter 1. The theoretical context supporting this rationale involves the description of a series of cognitive frameworks (i.e. Anderson & Krathwohl, 2001; Bloom, 1956; Marzano, 2001; Romiszowski, 1981). These cognitive frameworks highlight the importance and interdependence of memory, comprehension, and critical thinking as learning outcomes in educational contexts. In addition, each learning outcome, and related

thinking process, is further discussed in light of recent theory and empirical research in cognitive psychology. Specifically, Baddeley's model of working memory (1986; 2000), models of schema construction in long-term memory (e.g. Chi, Glaser and Rees, 1982; Chase and Simon, 1973; Kotovsky, Hayes & Simon, 1985), the Delphi model of critical thinking (Facione, 1990b) and the reflective judgment model (Kitchener & King, 1981) will be discussed. This analysis serves to deepen our understanding of the cognitive processes associated with different learning outcomes. Furthermore, factors that may impede memory, comprehension and critical thinking ability are reviewed and discussed, specifically, cognitive load and the problematic nature of text-based reading. This review and discussion is important for the purposes of contextualising the potential benefits of argument mapping and situating the core aims of the thesis in the broad fields of cognitive and educational psychology.

Chapter 2 presents a critical review of previous research conducted on strategies designed to improve memory, comprehension and critical thinking performance in educational contexts. More specifically, previous research on the effects of organisational strategies on memory and comprehension will be discussed, as will research on the effects of various training interventions on critical thinking performance. Thus, this review will present preliminary (indirect) empirical evidence concerning the potential of methods such as argument mapping as means for enhancing memory, comprehension and critical thinking.

Chapter 3 defines argument mapping and provides a brief history of its use. The nature of argumentation and its principles are discussed, as is the nature of argumentation using the principles of argument mapping. Previous research conducted on the effects of argument mapping on learning outcomes is also explored. Notably, this research focused mainly on critical thinking rather than on memory and

comprehension. In addition, many of the intervention studies in this area have been limited in certain respects (e.g. non-randomized controlled trials; insufficient control of experimenter bias). These deficiencies suggested the need for a targeted programme of empirical research in this domain. A general rationale for why argument mapping is hypothesised to enhance learning is also presented in Chapter 3, which is an effort to synthesise and extrapolate from the research and theory reviewed in Chapters 1 – 3.

Chapter 4 presents the rationale for Study 1 (which included four experiments), in which argument map reading and construction was hypothesised to enhance memory and comprehension performance. Chapter 4 also presents the results and discussion of the four experiments that were conducted. Experiments 1 and 2 examined the effect of argument map reading in comparison with text reading on memory and comprehension performance through the manipulation of three variables: (1) the colour of argument maps (i.e. coloured v. monochrome), (2) the size of arguments maps (i.e. 30-proposition v. 50-proposition), and (3) the environmental setting in which studying and testing took place (i.e. a lecture hall setting v. an isolated booth). Experiment 3 examined the main and interaction effects of argument size (i.e. 30-proposition v. 50-proposition) and study material (i.e. argument map v. text reading) on both immediate and delayed recall performance. Experiment 4 compared the effects on recall performance of actively learning with three different methods, namely, argument map construction, text summarisation, and hierarchical summarisation (outlining) of text. In summary, the results of these experiments revealed that argument map reading and argument map construction facilitated better immediate recall performance when compared with more traditional methods of learning.

The promising results of Study 1 prompted the development of an argument mapping-infused critical thinking course. Chapter 5 presents the rationale for why an argument mapping-infused critical thinking course was hypothesised to enhance overall critical thinking performance and the critical thinking sub-skills of analysis, evaluation, inference and reflective judgment in Study 2. The results of Study 2 are also presented, which more specifically compared the effects of a six-week argument mapping-infused critical thinking (CT) course, a CT course taught through traditional means (i.e. bullet-points and outlining) and a no-intervention control condition on CT and reflective judgment performance. Results revealed that participation in a CT training course significantly enhanced CT skills of analysis, evaluation, and inductive reasoning. Participants in the argument mapping-infused course scored significantly higher than controls on tests of inductive reasoning, while participants in the traditional CT course scored significantly higher than controls on tests of analysis and inductive reasoning. The results of Study 2 prompted the development of an e-learning version of the AM-infused CT training course, which was designed to open the CT training opportunity to a larger audience, improve upon the design of the first training course in a number of ways and examine the effects of engagement on the development of CT skills.

Chapter 6 explains the rationale for Study 3 in detail. Study 3 was designed to advance upon Study 2 and test the hypothesis that an argument mapping-infused CT course taught through e-learning enhances overall CT ability and critical thinking sub-scale performance. The issues that needed to be considered in translating the AM training from a classroom setting to an online environment are discussed. In addition to significant modification of lecture recordings and exercises, Study 3 also made use of online communication systems to provide students with feedback on their work.

Ultimately, Study 3 examined the effects of a six-week argument mapping-infused CT course taught through e-learning on CT performance, in comparison with a no-intervention control condition. Results revealed that there was a significant gain from pre-to-post-testing on all aspects of CT performance for those who attended the e-learning course; and that those who took part in the course scored significantly higher on overall CT and multiple CT sub-scales than those in the control group.

Chapter 7 concludes the thesis by presenting a general discussion of the empirical findings from this research. Findings are evaluated with respect to their theoretical and empirical implications. Limitations of the research and possible future research are also discussed. Finally, a general summary is presented which discusses the beneficial effects of argument mapping on memory and critical thinking performance as well as its potential contribution to the field of education.

Chapter 1

Thinking Processes in Instructional and Educational Settings

This chapter introduces and defines a number of key thinking processes that educational strategies should positively affect. These processes are memory, comprehension, critical thinking, and the critical thinking sub-skills of analysis, evaluation, inference and reflective judgment. Given that the aim of this thesis is to examine the effects of argument mapping on these thinking processes, the rationale for the selection of these processes as targets for educational intervention is first explained by reference to thinking frameworks for instructional and educational objectives. Research and theory from the field of cognitive psychology is also drawn upon to further describe these thinking processes and their interdependence. Finally, the latter section of the chapter will investigate factors that may negatively impact memory, comprehension, and critical thinking in educational settings, such as cognitive load and the problematic nature of text-based learning.

1.1. Frameworks for Thinking

In the past, cognitive and metacognitive processes such as memory, comprehension and critical thinking have been described in the context of various thinking frameworks, developed specifically for the purpose of facilitating instructional design and the advancement of educational objectives (Moseley et al., 2005). A framework for thinking is a “general pool of constructs for understanding a domain” (i.e. cognition), “but is not tightly enough organised to constitute a predictive theory” (Anderson, 1983, pp. 12-13). Simply, a framework for thinking organises the identified cognitive processes in an effort to explain the structure of thought (Moseley et al., 2005). The identification and description of thinking processes vary from

framework to framework, given that the term ‘framework’ covers a wide variety of structures (Moseley et al., 2005). More specifically, frameworks differ in complexity, in that some may be very simple (e.g. lists of cognitive processes that students may need to use in different learning situations), while others may be quite complex (e.g. taxonomies that hierarchically structure lower-order to higher-order thinking processes according to complexity or importance).

For example, a framework for thinking that is represented as a list may simply present an inventory of thinking processes (e.g. identify, name, describe construct, order and demonstrate; cf. Gerlach & Sullivan, 1967), which may be helpful given that an inventory can be used to check that a certain cognitive activity has been given attention in the classroom (Moseley et al., 2005). However, a list of thinking processes offers no explicit description of the relationship between thinking processes (e.g. hierarchical relationships among the entries), nor does it rank items on the list according to complexity, importance or the context in which they might be applied.

A more complex thinking framework may be represented as a group, which organises thinking processes according to shared similarities or inter-relationships (e.g. Jonassen & Tessmer, 1996). A group of thinking processes may be subdivided into smaller groups. These groups may stand alone, co-exist alongside other groups, or hierarchically subsume other smaller groups (e.g. according to complexity or interdependency) - thus creating organised ranks (Moseley et al., 2005).

Hierarchical classification of groups is a common feature of thinking frameworks that are represented as taxonomies. In a taxonomy, groups within groups are created via the classification and organisation of thinking processes, which produces an organised hierarchy. Many current frameworks for thinking are taxonomies (e.g. Allen, Feezel, Kauffeld, 1967; Anderson & Krathwohl, 2001;

Bloom, 1956; Ennis, 1998; Marzano, 2001; Romiszowski, 1981). Taxonomies of thinking processes may be developed for a number of specific reasons (e.g. to address instructional design, productive thinking, or cognitive development). For example, instructional design taxonomies aim to provide frameworks and guidance on the practical task of designing learning experiences (Moseley et al., 2005). The focus of instructional design taxonomies are the thinking processes (i.e. both *lower-order* and *higher-order thinking skills*; Bloom, 1956) that can be improved through educational instruction. Common instructional design frameworks are those created by Bloom (1956), Romiszowski (1981), Gagne (1985), Anderson and Krathwohl (2001), Marzano (2001) and Presseisen (2001).

On the other hand, frameworks which focus more on productive thinking deal specifically with thought processes geared towards the generation of ideas and consequences of thought-based actions. Productive thinking frameworks specifically focus on the *higher-order levels of thinking* identified by Bloom (1956), such as analysis, synthesis and evaluation; as well as the various metacognitive (i.e. thinking about thinking; Flavell, 1976; Ku & Ho, 2010b) aspects of thought including: critical thinking, disposition towards thinking, self-regulation of thinking processes, and creative thinking (e.g. Allen, Feezel & Kauffeld, 1967; Ennis, 1987; De Bono, 1985; Halpern, 2003a; Paul, 1993; Petty, 1997). Frameworks may also be devised specifically to address and model the manner in which cognition develops (e.g. Carroll, 1993; Demetriou, 1993; Fischer, 1980; King & Kitchener, 1994; Kitchener & King, 1981; Piaget, 1952).

Though frameworks that focus specifically on productive thinking and cognitive development are important to consider, the remainder of this section will focus on a series of frameworks that were developed for the purpose of instructional

design - more specifically, frameworks that align with the goals of this thesis, which is to examine the effects of argument mapping on the learning outcomes of memory, comprehension, analysis, evaluation, inference, and reflective judgement skills. In this context, a number of frameworks are usefully considered in more detail, such as those developed by Bloom (1956), Romiszowski (1981), Anderson & Krathwohl (2001) and Marzano (2001). These frameworks have been chosen for discussion because the cognitive processes they describe as being necessary in educational settings (1) have been demonstrated in empirical research as cognitive processes that can be improved by educational intervention (e.g. Berkowitz, 1986; Butchart et al., 2009; Oliver, 2009; Marzano, 1998; Robinson & Kiewra, 1995; Taylor, 1982; van Gelder, 2001; van Gelder, Bissett & Cumming, 2004; as discussed in Chapter 2); and more specifically, (2) are hypothesised to be improved through the use of argument mapping (as discussed in Chapters 3-7).

Frameworks will be presented in chronological order, beginning with Bloom's (1956) influential taxonomy of educational objectives, which identifies six core thinking processes and learning outcomes that need to be cultivated in educational settings (i.e. knowledge, comprehension, application, analysis, evaluation and synthesis). Each of these educational objectives, and corresponding thinking process, will then be discussed in relation to Romiszowski's (1981) 'skill cycle', which elaborates upon how the thinking processes identified by Bloom both function and develop in educational settings. Next, Anderson and Krathwohl's (2001) revision of Bloom's taxonomy will be discussed in order to provide a more up-to-date perspective on the kinds of cognitive processes that Bloom had previously identified as important in educational settings. Finally, Marzano's (2001) taxonomy will be discussed in light of his meta-analysis of educational interventions (Marzano, 1998).

1.1.1 Bloom's Taxonomy of Educational Objectives (1956)

Bloom's taxonomy of educational objectives was developed for purposes of enhancing education and the manner in which thinking is examined in the classroom via classifying "mental acts or thinking [resulting from] educational experiences" (Bloom, 1956, p.12). Bloom's taxonomy, which actually emerged from consensus agreement and work completed by a group of university professors (i.e. B.S. Bloom, M.D. Engelhart, F.J. Furst, W.H. Hill & D.R. Krathwohl), has been a cornerstone of educational practice for many years, as it was one of the first frameworks to characterise thinking as an array of both *lower-order* and *higher-order* thinking processes. In addition, the taxonomy has been claimed to be "the most pervasive in curriculum development and provides the clearest definition of educational goals expressed in terms of descriptions of student behaviour" (Reeves, 1990, p. 609).

In the past, Bloom's taxonomy has been used as a pedagogical aid in enhancing writing performance (Granello, 2001), as a basis for performance evaluation in various academic domains (Scott, 2003), and as a guide for planning course curricula and designing instruction (Krathwohl, 2002). Research has also shown that making students aware of Bloom's taxonomy increases their ability to apply the appropriate skills (i.e. knowledge, comprehension, application, analysis, synthesis and evaluation) when critically thinking (Athanassiou, McNett & Harvey, 2003). More generally, when it comes to designing interventions that seek to cultivate Bloom's lower-order and higher-order thinking processes, there is a vast educational literature that has identified a plethora of educational strategies that are more or less effective (Berkowitz, 1986; Chi, Glaser & Rees, 1982; Gadzella, Ginther & Bryant, 1996; Hitchcock, 2004; Reed & Kromrey, 2001; Rimiene, 2002; Robinson & Kiewra, 1995; Solon, 2007; Taylor, 1982; Taylor & Beach, 1984; see Chapter 2).

Bloom's taxonomy of educational objectives consists of six major categories of thought (see Figure 1.1). The first category pertains to the *knowledge* of specifics (e.g. facts); ways and means of dealing with specifics (e.g. procedures); and the abstract nature of some information (e.g. abstract concepts). This may include knowledge of specific terminology, facts, conventions, patterns, classifications, criteria, methodologies, principles, generalisations, theories and structures; and the ability to *recall* this knowledge upon demand. The second category of thought, *comprehension*, is the ability to understand or grasp the meaning of information. This implies the ability to interpret and extrapolate information for purposes of developing understanding; and also to translate information from one level of abstraction to another, one symbolic form to another, or one verbal form to another (e.g. for purposes of summarising, paraphrasing or explaining information). The third category, often referred to as the least well understood category of Bloom's taxonomy (Moseley et al., 2005) is *application*, which is the use of learned information in new and concrete situations. The fourth category, *analysis* is the ability to break down ideas or arguments into their component parts (e.g. analysing a complex idea by reference to the organisational principles, abstractions, and representations expressed by a speaker or author). *Synthesis*, the fifth category, refers to the production of a unique communication or plan (i.e. the ability to put parts of information together to form a new whole). The final category of thought in Bloom's taxonomy is *evaluation*, which is defined as the ability to make judgments in terms of the value of internal evidence and/or external criteria.¹

¹ Despite what seem like clear definitions (Moseley et al., 2005), Wood (1977) found that educators often find it hard to differentiate among the higher-order thinking skills (especially analysis, synthesis and evaluation). These three processes will be discussed later in greater detail with regards to both Romiszowski's 'skill-cycle' and Facione's (1990) conceptualisation of critical thinking, in order to clarify distinctions among them.

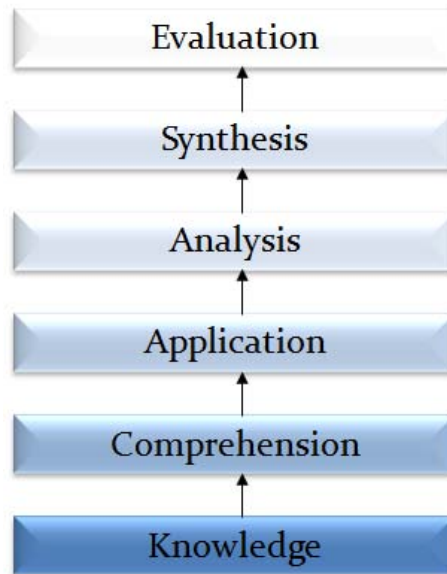


Figure 1.1: Bloom's Taxonomy (1956)

The categories of Bloom's taxonomy are hierarchically arranged starting at the bottom with the lower-order thinking skills of knowledge/recall and comprehension; and proceeding through the progressively higher-order thinking skills of application, analysis, synthesis, and evaluation. Even though memory-based knowledge and comprehension are labelled *lower-order* thinking skills, this does not suggest that they are ultimately less important for education than the higher-order thinking skills. For example, the operation of the higher-order processes is dependent upon the existence of knowledge (e.g. comprehending, applying, analysing, synthesising and evaluating knowledge of specific facts, conventions, patterns and methodologies). More specifically, though Bloom's six categories are organised according to the complexity of each mental act (i.e. from *Knowledge* to *Evaluation*), this organisation is based on a hierarchical interdependence between levels. For example, in order to *Evaluate* a theory [level six], one must first be able to *analyse* its propositions [level four] and *synthesise* their interdependence [level five]. Furthermore, to be able to *analyse* a theory [level four] one must be able to *comprehend* its propositions [level two], and

remember them [level one]). However, there has been debate over whether Bloom's six categories imply a strict form of hierarchical interdependence, or constitute a rigid or interdependent form of hierarchical complexity (e.g. Anderson & Krathwohl, 2001). For example, Kreitzer & Madaus's (1994) review of Bloom's (1956) taxonomy included discussion of research conducted by Kropp, Stoker and Bashaw (1966), which assessed the performance of students on the thinking processes in Bloom's taxonomy and found no evidence of a difficulty or complexity hierarchy for the higher level skills of evaluation and synthesis. Based on these and other findings, Kreitzer & Madaus (1994) suggest that *evaluation* is not more complex than *synthesis*. In other words, evaluation does not readily subsume synthesis in a higher-order/low-order hierarchical relationship. This issue is further explored in the discussion of Anderson and Krathwohl's (2001) revision of Bloom's taxonomy presented below.

Apart from the influence that Bloom's taxonomy has had on the development of subsequent frameworks (e.g. Anderson & Krathwohl, 2001; Halpern, 2003a; Marzano, 2001, Moseley et al., 2005; Romiszowski, 1981), it has prompted the development of practical hierarchical inventories of cognitive processes linked to specific learning outcomes in educational settings, as well as operational definitions to facilitate measurement of these learning outcomes. This is particularly useful in the context of educational research as the translation of operational definitions into specific measurement tools has allowed subsequent research to examine student performance across skills in the hierarchy under different experimental conditions (e.g. Berkowitz, 1986; Butchart et al., 2009; Farrand, Hussain & Hennessy, 2002; Meyer, Brandt & Bluth, 1980; Taylor, 1982, van Gelder, Bissett & Cumming, 2004; see Chapter 2).

1.1.2 Romiszowski's (1981) Framework for Knowledge & Skills

Bloom's taxonomy heavily influenced Romiszowski's (1981) investigation of cognitive processes and methods of improving educational instruction. Romiszowski defines instruction as "a goal-directed teaching process which is more or less pre-planned" (p. 4). According to Romiszowski, the relationship between instruction and learning is described as a three-part sequence which consists of *input* (i.e. information), the use of a *system* (i.e. a cognitive process – such as those identified by Bloom) and *output* (i.e. performance or knowledge). During this sequence, one can apply any number of cognitive processes on taught information for the purpose of producing knowledge.²

According to Romiszowski, it is often difficult to define or even differentiate systems. Thus, he metaphorically refers to these systems as acting in a *black box*, wherein the contents of the box (i.e. cognitive processes such as those described by Bloom) cannot be viewed directly. The value of a specific instructional method is assessed via a cost-benefit analysis, in which the benefit of training is directly compared with the cost of the training, in terms of time and effort expended by student and/or teacher. If after the training there is a "deficiency in performance" or the output is not worth the cost, then it may become necessary to "open the black box" in order to see if the cognitive process (i.e. the system) used was the cause of the deficiency (Romiszowski, 1981, p. 253). Ultimately, Romiszowski provides a *skill-cycle* (see Figure 1.2) to describe not only what processes are "inside the black box",

² Romiszowski claims that when designing instruction, it is important that the input, the system, and the output are designed or utilised specifically in relation to one another (e.g. if the input is altered, so too will the system and the output). Training must be designed to foster certain skills in order to produce a desired outcome. This is an important notion to consider for later in this thesis, when training is designed for purposes of enhancing memory and critical thinking performance.

but also the way in which these processes interact; in the event that the black box “needs to be opened” (p. 253).

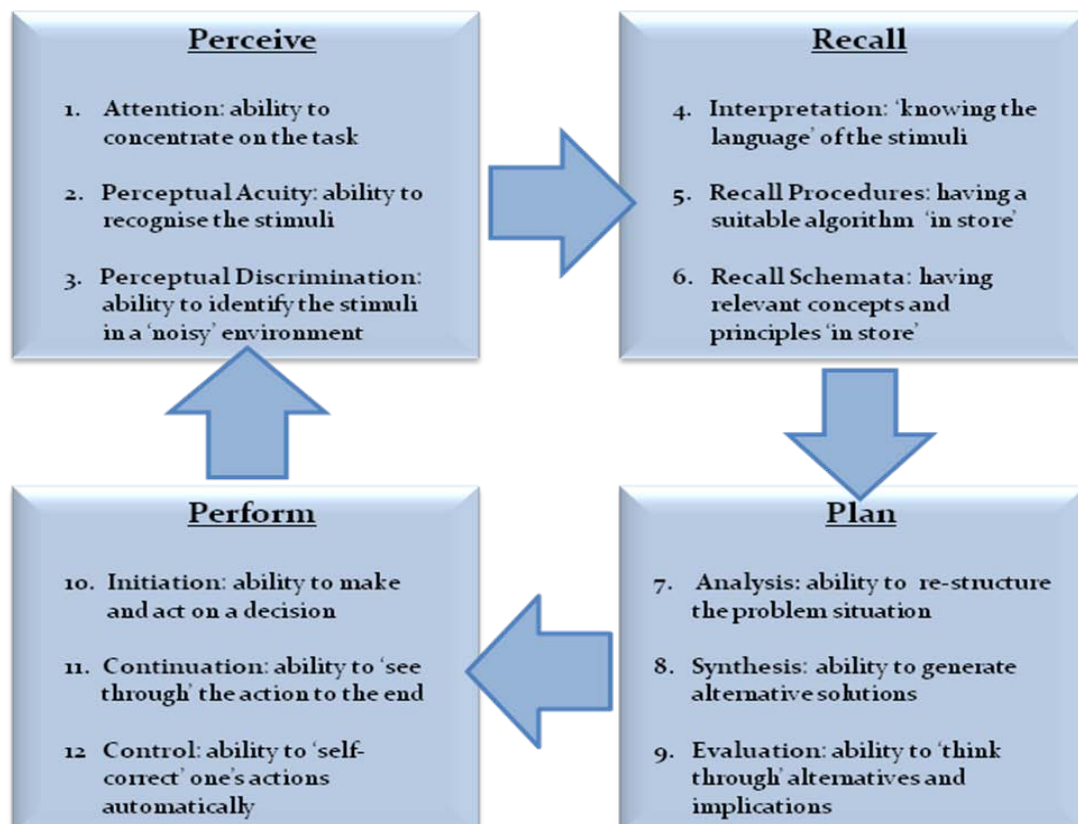


Figure 1.2: Romiszowski's (1981) Skill Cycle (adapted from Moseley et al., 2005)

According to Romiszowski's skill-cycle, skill development is dependent on how often the skill is practiced and on how well skill development is supported by instructional design and an accommodating learning environment. Skills act upon novel, incoming information as well as pre-existing knowledge. Within this framework are four types of *knowledge*. The first type of knowledge is based on knowing *facts* (e.g. “knowing objects, events or people”; p. 242). The second type of knowledge is based on knowing *procedures* (i.e. “knowing what to do in given situations”; p. 242). The third type of knowledge Romiszowski describes is knowledge based around *concepts* (i.e. “knowing specific concepts or groups of concepts, such as being able to give or recognise instances of a given phenomena”,

p. 243). The final type of knowledge is based on knowing *principles* (i.e. “principles which link certain concepts or facts in a specific way, such as being able to explain or predict phenomena”, p. 243). The distinctions Romiszowski makes between different types of knowledge resonate with other theories, including Tulving’s (1984) distinction among semantic memory (e.g. memory for meanings and concepts), episodic memory (e.g. memory for events) and procedural memory (e.g. memory for how to do things). Consistent with Bloom’s scheme, the four types of knowledge laid out by Romiszowski can be recalled and understood via the processes of ‘memory’ and ‘comprehension’. They can also be used in different ways to plan and perform educational tasks.

During the skill-cycle, an individual *perceives* information, *recalls* information, makes *plans* and *performs* based on that information. According to Romiszowski, when perceiving information, the individual concentrates on a task or problem, recognises the relevant stimuli (e.g. visual stimuli, verbal stimuli or an integration of both), and is able to discriminate those stimuli from others. The next step, recall, is where the individual interprets the perceived stimuli from the environment by retrieving the appropriate knowledge (i.e. facts, procedures, concepts or principles) that is necessary to apply in the specific task context or problem situation. The third step involves making plans based on information obtained from the previous steps (i.e. perception and recall), by analysing, synthesising and evaluating this information (i.e. by restructuring the problem-situation, generating alternative solutions and judging these alternatives). The final step in Romiszowski’s skill cycle is performance: in accordance with the plan just devised, the individual makes and acts on a decision, sees that decision through and is able to correct or self-

regulate their own actions, based on the information manipulated in previous steps. As this cycle is reiterated, the skill(s) which is engaged is able to develop.³

Romiszowski's skill-cycle is similar to Bloom's taxonomy in several respects. The second phase of the skill-cycle is broadly congruent with Bloom's lower-order thinking skills (i.e. memory and comprehension), while the third phase of the skill-cycle is congruent with Bloom's higher-order thinking skills (i.e. analysis, evaluation and synthesis). However, unlike Bloom's Taxonomy, Romiszowski's skill-cycle is more closely aligned with cognitive architecture (Anderson, 1990) and neuroscientific accounts of the perception-action cycle (Fuster, 2000), and the two phases of Romiszowski's skill-cycle not directly elaborated upon by Bloom are those of perception and performance. In terms of perception, it may be taken for granted by Bloom that any 'stimuli' subject to thought first requires attention (i.e. in order to think about some specific educational task, one must perceive a stimulus). Furthermore, in relation to performance, it may be that this final phase of Romiszowski's skill-cycle is congruent with Bloom's category of application, given that both 'application' and 'performance' refer to utilisation of knowledge. However, Romiszowski's concept of performance is also distinct from Bloom's notion of application in important ways. For example, while Bloom refers to application in terms of particular uses of knowledge (e.g. applying a theory to a social problem), Romiszowski's concept of performance is much more akin to cognitive notions of

³ Romiszowski's theory resonates with Piaget's (1952) *theory of cognitive development* and Fischer's (1980) *Dynamic Skill Theory*, in that knowledge can be conceptualised by the individual as concrete (i.e. facts and procedures) or abstract (i.e. concepts and principles). Fischer's theory and empirical work highlights the fact that skill development is often domain specific (i.e. skills develop independent of one another and at different rates). Different skills draw upon different knowledge. Furthermore, Fischer argues that skills develop through the hierarchical coordination of lower level action systems into higher-order structures, with abstractions and principles derived from the coordination and mapping of actions and representations. Fischer's conceptualisation of skill development is important to consider here as it is central to our discussion of the development of reflective judgment ability below.

executive control (Fuster, 2000), and metacognitive, self-regulatory processes (Boekaerts & Simons, 1993; Marzano, 2001; Pintrich, 2000); specifically, with its focus on the ability to initiate, persevere, and control actions.

The development of a skill through Romiszowski's cycle involves the planned operation of that skill on some item of knowledge for a particular purpose. According to Romiszowski, the amount of planning required to perform each skill dictates whether that skill is either *reproductive* or *productive* in practice. A *reproductive skill* refers to "skills that are more or less reflexive in nature, that are repetitive and that show little variation in execution from one instance to another" (Romiszowski, 1981, p. 250). Romiszowski's description of reproductive thinking is somewhat consistent with Bloom's conceptualisation of 'knowledge'. For example, a student may draw upon knowledge of a specific fact, convention, pattern, classification, criteria, methodology, principle, or theory in a reflexive manner without much by way of planning, for example, in response to a question posed by a teacher in a classroom setting. Conversely, a *productive skill* refers to "a skilled behaviour that requires a certain amount of planning, that involves the use of some strategy for decision-making and shows substantial variations in execution from one instance to another" (Romiszowski, 1981, p. 251).⁴

An example of how reproductive and productive thinking are applied is provided by Romiszowski in terms of painting and decorating a home. Romiszowski (1981) suggests that a painter requires little knowledge in order to paint a wall, apart from mixing paint and laying it on the wall without leaving brush-marks. This is considered analogous to reproductive thinking because all the painter needs to do is to

⁴ Romiszowski's concept of productive thinking is akin to the description of metacognitive processes provided by others (e.g. Ku & Ho, 2010b; Marzano, 1998; 2001), which is characterised by both the dispositional/self-regulatory functions of thinking; and the strategical planning and application of high-order thinking skills (i.e. analysis, evaluation and inference) when thinking about thinking (Brown, 1987; Flavell, 1979; Ku & Ho, 2010b).

reproduce a procedure. On the other hand, a decorator who is applying wallpaper and seeking to plan out and configure an ideal living space, needs not only knowledge of the *procedure* for measuring the room, cutting the paper and mixing/applying paste, but also the aesthetic *principles* of obtaining a high quality of finish (in order to visualise the finished product), concerning for example, the positioning of patterns and joints in the wallpaper, balance of the patterns in the wallpaper around doors and windows and how to accentuate or conversely hide certain features of the wallpaper, in order to produce a pleasing effect. These principles are derived from a much larger knowledge-base and combine into a more complex set of decision-making strategies that enables the decorator not only to know how to complete each task, but also why it is being done.

Previous research suggests that both reproductive and productive components of thinking can be measured and enhanced. For example, past research conducted on verbal recall performance (i.e. a type of reproductive thinking) suggests that recall can be improved by reading from and constructing organisational representations of text-based information (e.g. Berkowitz, 1986; Farrand, Hussain & Hennessy, 2002; Taylor, 1982; Taylor & Beach, 1984). Past research also suggests that critical thinking (i.e. a type of productive thinking; Ennis, 1998; Halpern, 2003a; Moseley et al., 2005) can be enhanced through critical thinking-focused training interventions (e.g. Abrami et al., 2008; Alvarez-Ortiz, 2007; Hitchcock, 2004; Reed & Kromrey, 2001; van Gelder, 2001; van Gelder, Bissett & Cumming, 2004).

The importance of Romiszowski's framework is that it positions many of the thinking processes identified by Bloom as acting in a cycle (e.g. recall processes, analysis, synthesis and evaluation), describing how and when they are to be utilised in educational settings. In addition to Romiszowski's framework, more recent efforts

have been made to build upon Bloom's taxonomy of educational objectives (e.g. Anderson & Krathwohl, 2001; Marzano, 2001). Anderson and Krathwohl's (2001) taxonomy follows Romiszowski's general path of development (i.e. which placed an action-oriented focus on each thinking process as working in a skill-cycle) by, first, transforming Bloom's hierarchical thinking processes from noun form to verb form (i.e. naming actions instead of 'things') and, second, by placing acts of 'creation' as the pinnacle process in the hierarchy.

1.1.3 Anderson & Krathwohl's (2001) Revision of Bloom's Taxonomy

To reiterate, Anderson and Krathwohl's (2001) revised taxonomy made a number of changes to Bloom's taxonomy, including the ordering of the processes (i.e. the process of evaluation and synthesis [now creation] were exchanged); and the presentation of each process in verb form as opposed to noun form (see Figure 1.3). Another change was that Anderson and Krathwohl (2001) were explicit in proposing that the listed categories of cognitive processes no longer form a unified cumulative hierarchy. That is, one does not need to master lower levels of the taxonomy in order to ascend to higher levels (e.g. working at level six [*Creating*] does not require mastery of level five [*Evaluating*]). However, Anderson and Krathwohl claim that a cumulative, hierarchical interdependence exists between the activities of *understanding*, *applying* and *analysing* (i.e. located in the mid-section of the framework; again, see Figure 1.3); and that all processes are dependent upon *knowledge* as it is not possible for one to use thinking processes (located on the subsequent rungs of the taxonomy) if one does not know, or cannot remember the information one is supposedly thinking about (Krathwohl, 2002). This is an important issue to consider given that it has been argued by researchers in the field of critical thinking that the ability to think critically about specific information (i.e. analyse,

evaluate, and infer reasonable conclusions) is directly affected by one's ability to recall and understand (i.e. lower-order thinking skills) the information one is required to think about (Halpern, 2003a, Maybery, Bain and Halford, 1986).

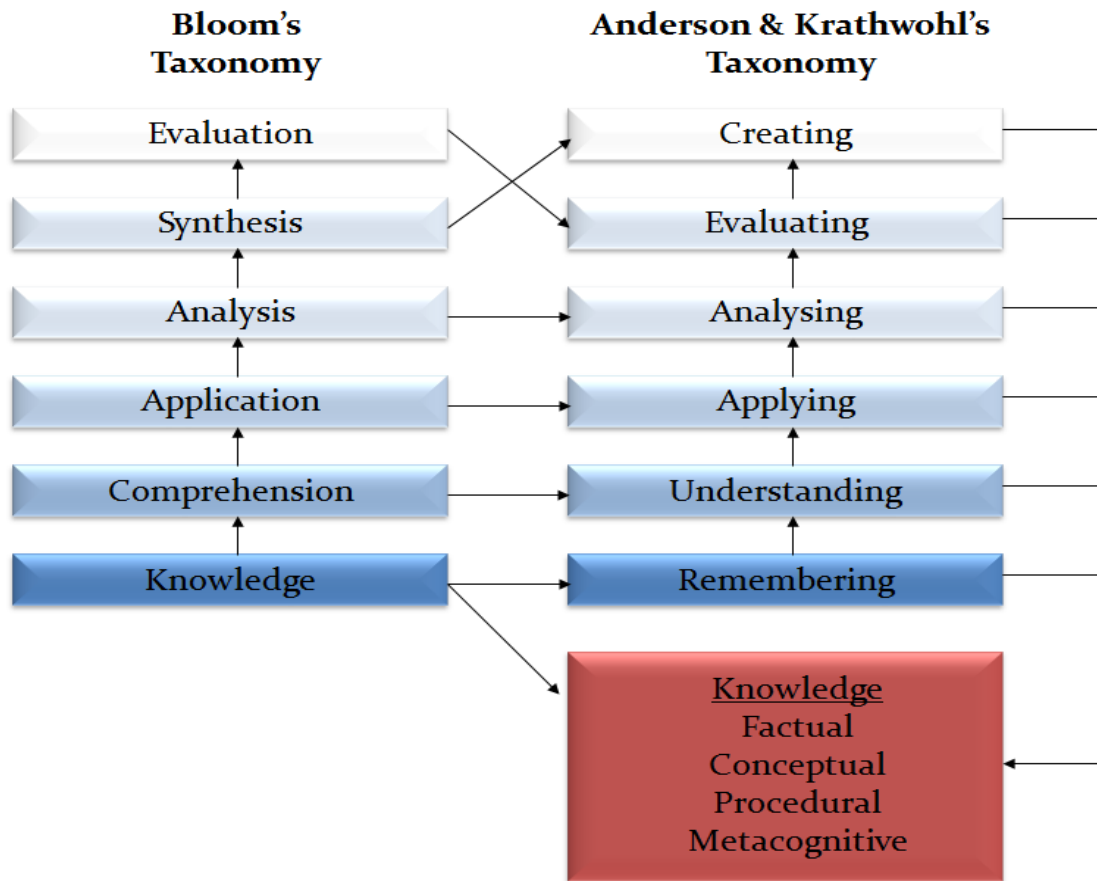


Figure 1.3: Bloom's Taxonomy and Anderson and Krathwohl's (2001) Revision

The implied dependence of the various cognitive processes upon knowledge in the revised taxonomy results from another important difference between these taxonomies. In Bloom's taxonomy, a single, hierarchical arrangement of cognitive processes was explicated (i.e. consisting of knowledge, comprehension, application, analysis, synthesis and evaluation); whereas in Anderson and Krathwohl's (2001) revised taxonomy, two dimensions are described – both a knowledge dimension and a cognitive process dimension. Specifically, in Bloom's taxonomy, knowledge encompassed both *knowledge* of different forms of facts, procedures and abstractions,

as well as the ability to *remember* facts, procedures and abstractions (Krathwohl, 2002). In the revised taxonomy, knowledge is described as a separate dimension (Krathwohl, 2002; Moseley et al, 2005), whereas the ability to remember is described as one of the thinking processes on the other dimension. Anderson and Krathwohl made this decision to highlight their belief that each of the six processes acted upon knowledge in their own right (i.e. remembering, understanding, applying, evaluating and creating knowledge).

Furthermore, the addition of this new dimension of knowledge is an important feature of Anderson and Krathwohl's (2001) revised taxonomy because it also includes an additional knowledge component not included in Bloom's original taxonomy: *metacognitive knowledge*; which in this context refers to strategic knowledge, knowledge about cognitive processes and tasks, and self-knowledge (Anderson & Krathwohl, 2001). However, the presentation of this additional feature in Anderson and Krathwohl's revised taxonomy is not to suggest that Bloom's original taxonomy did not take metacognitive processes into account. In fact, Bloom's higher-order thinking skills can be viewed as metacognitive processes. That is, when thinking about thinking, the processes of analysis, synthesis and evaluation, along with self-regulation, can aid strategical planning (Brown, 1987; Flavell, 1979; Ku & Ho, 2010b). For example, an individual may use metacognition to think about thinking in the context of *analysing*, *synthesising* and *evaluating* their own thinking and/or the thinking of others.

Notwithstanding the fact that Romiszowski's framework also captures these metacognitive processes in his skill-cycle steps of *planning* and *performing*; Anderson and Krathwohl's development of Bloom's taxonomy is important to consider because it is the first framework discussed which *explicitly* includes reference to a distinct

metacognitive component within the thinking processes. Anderson and Krathwohl's taxonomy is further important to consider because it presents knowledge and remembering knowledge as separate functions of information storage and recall, respectively; as opposed to a collation of processes under one heading (as in Bloom's taxonomy). Another recent taxonomy which includes a separate knowledge construct, as well as a metacognitive component is Marzano's (2001) taxonomy of educational objectives, which is heavily informed by a large-scale meta-analysis of educational interventions designed to facilitate a broad range of different learning outcomes (Marzano, 1998).

1.1.4 Marzano's (2001) New Taxonomy of Educational Objectives

Similar to Anderson and Krathwohl (2001), Marzano (2001) also developed a taxonomy of educational objectives (see Figure 1.4) based on Bloom's taxonomy. The importance of Marzano's taxonomy is that it is also based on very specific empirical research - summarized in Marzano's (1998) meta-analysis, which examined the effect of various instructional techniques on academic achievement. Marzano utilised over 4,000 effect sizes involving roughly 1.237 million subjects. Broadly speaking, the results of the meta-analysis revealed that instructional techniques that focused directly on the knowledge domain had an average effect size of .60. Interventions that focused on the cognitive system had an average effect size of .75. Interventions that focused on the metacognitive system had an average effect size of .55; and interventions that focused on the self system had an effect size of .74. Marzano's meta-analysis also revealed significant effects on overall learning of specific instructional techniques, such as: note-taking (ES = .99), advanced organisers (i.e. structured explanations of information; ES = .48), and graphic representation (ES = 1.24). These results are important to consider in the following chapters, where specific organisational and

graphical representation methods of instruction including, but not exclusive to argument mapping, are discussed in terms of their effects on memory, comprehension and critical thinking.

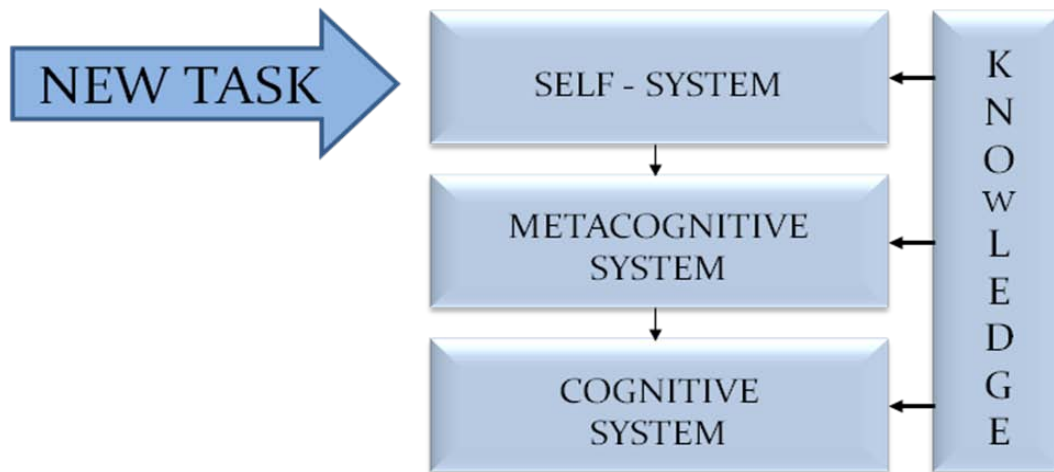


Figure 1.4: Marzano's (2001) 'New' Taxonomy of Educational Objectives

Though the structure of Marzano's taxonomy differs from those developed by Bloom (1956), Romiszowski (1981) and Anderson and Krathwohl (2001), it remains similar to all three in that it includes (under the broad category of the *cognitive system*) the processes of *knowledge retrieval* (i.e. memory/recall), *comprehension* (i.e. knowledge representation), *analysis* (i.e. classifying, identifying errors, generalising, matching and specifying) and *knowledge utilisation* (i.e. decision-making, problem-solving, investigation and experimental enquiry). In addition, Marzano's (2001) taxonomy is also similar to Anderson and Krathwohl's taxonomy in that it explicitly includes a metacognitive component of thought. In Marzano's taxonomy, the *metacognitive system* acts as an executive control of all processes; more specifically, a self-regulatory process with a focus on goal and process specification, as well as process and disposition monitoring (Marzano, 1998). Marzano's taxonomy also presents a self-system in which goals are produced (to be executed by the

metacognitive system) as a result of motivation, attention, beliefs and the interaction of such processes (Marzano, 1998; Moseley et al., 2005). Specifically, it is the self-system that determines whether or not any given task will be undertaken. These three systems (i.e. the cognitive, metacognitive and self-systems) all act upon retrieved content from an individual's *knowledge domain*, the fourth component of Marzano's taxonomy, which consists of stored information as well as knowledge of mental and psychomotor procedures. According to Moseley et al. (2005), this knowledge can be represented verbally, non-verbally, or in an affective manner.

In light of the broader meta-analytical findings, Marzano (1998, p. 121) proposes that when the metacognitive and cognitive systems function together, they can enhance learning, as they “provide individuals with an awareness of the manner in which their minds work” and “requires them to monitor their mental activity”. Marzano (2001) further suggests that recognition by educators of the influence of both the self-system and metacognitive system on the cognitive system and the development of knowledge is of utmost importance in educational settings. Marzano's work represents an important advance upon past frameworks as it provides empirical support for the inclusion of self-regulatory and monitoring processes within frameworks for thinking. Also, based on large effect sizes reported by Marzano in his meta-analysis, it appears that note-taking, organisational strategies and graphic representation of knowledge are three learning and teaching strategies that can be usefully applied in different learning contexts. This theme will be further explored in Chapters 2 and 3.

1.1.5 Summary of the Frameworks for Thinking

In summary, a number of frameworks have identified cognitive processes necessary for thinking in educational settings. Broadly speaking, two major

components of thinking are often distinguished, variously described as lower-order thinking skills and higher-order thinking skills (Bloom, 1956); reproductive and productive thinking skills (Romiszowski, 1981); and a cognitive process dimension and knowledge dimension (Anderson & Krathwohl, 2001; Marzano, 2001). Similarly, for the remainder of the thesis, lower and higher-order thinking skills will be referred to as functioning in a cognitive system (i.e. thinking processes such as memory and comprehension) and a metacognitive system (i.e. regulatory processes used when thinking about thinking, such as analysis, evaluation and synthesis), respectively. The focus of this thesis is on the effects of argument mapping on a number of sub-components of cognition and metacognition that are measurable in specific ways (i.e. memory, comprehension, analysis, evaluation and inference)⁵.

Though the taxonomies presented above are adequately descriptive in terms of identifying thinking processes and the links among them, it is also important to consider the empirical cognitive psychology research which has investigated these processes. In addition, a possible weakness of the frameworks above is that they do not elaborate on the manner in which one *applies* higher-order thinking processes. Bloom (1956) himself admitted that the process of *application* (i.e. the ability to use learned material in new and concrete situations) is the least well elaborated skill presented in his taxonomy. Though Anderson and Krathwohl did elaborate upon application by describing it as involving ‘carrying out or using a procedure in a given

⁵ The conceptualisation of higher-order thinking skills in Bloom’s taxonomy and Anderson & Krathwohl’s revision are analogous to the concept of critical thinking used in this thesis (i.e. Facione, 1990), as it consists of the skills of analysis, evaluation and synthesis. Though Bloom uses the term *synthesis*, his description of it is akin to *inference* as described by others working the field of critical thinking (i.e. the gathering of information used to develop a conclusion based on previous evaluation and analysis; Facione, 1990). Notably, Reeves (1990, p. 6) has amalgamated levels 3 (i.e. application) through 6 (i.e. evaluation) of Bloom’s taxonomy and relabelled the resulting collection of processes as *critical thinking*. Thus, the metacognitive process of critical thinking referred to here is analogous to Bloom’s higher-order thinking skills. Furthermore, the term synthesis will no longer be used in its more colloquial sense of creative synthesis, but will be referred to from here on as inference.

situation [through] execution or implementation (Krathwohl, 2002, p. 215), the concept remains incompletely revealed. Application may instead be reconsidered as a distinct, conjoint process (Moseley et al., 2005), similar to that described by Romiszowski (1981) as *perform* in his model of the skill-cycle, which implies that one must perform (i.e. initiate, continue and control) thinking processes based on the knowledge gained through the preceding processes (e.g. one may apply what was comprehended, what was analysed, or what was evaluated). One feature of *application* that is pertinent in the context of this thesis is the reflective judgment an individual brings to bear in the application of knowledge. Reflective judgment (i.e. the ability to *apply* critical thinking skills of analysis, evaluation, and inference whilst acknowledging uncertainties and limitations within one's knowledge; King & Kitchener, 1994), will be discussed below and elaborated upon as a key feature of higher-order thinking processes that can be the focus of educational interventions.

Based on the discussion of thinking frameworks and the identification of the key outcome variables of interest in this thesis (i.e. memory, comprehension, analysis, evaluation, inference and reflective judgment) the remainder of this chapter will examine these components of thinking in more detail and by reference to research from the field of cognitive psychology. In particular, the remainder of the chapter will discuss working memory, long-term memory and the link between comprehension and memory in light of research and theory in cognitive psychology, and the value of this research and theory for understanding why the use of argument mapping (as discussed in Chapter 3) may enhance recall and comprehension performance (i.e. the first two levels of Bloom's taxonomy). Next, the higher-order, metacognitive skills of analysis, evaluation and inference, as well as reflective judgment, will be discussed in the light of definitions of *critical thinking*, with a more detailed review of critical

thinking intervention studies and argument mapping-infused critical thinking intervention studies postponed to Chapters 2 and 3. The final section in this chapter will focus on the issues of cognitive load and the problematic nature of text-based learning. This section provides further context and an empirical foundation for understanding why argument mapping may facilitate the development of memory, comprehension, critical thinking and reflective judgment skills.

1.2 Memory

In order for an individual to remember information, a number of processes must first take place, such as the active processing, encoding, storage and retrieval of information. More specifically, one must first attend to the information and then organise it in a meaningful way, for the purposes of successful recall. For example, when reading for the purpose of learning, people actively process information. Broadbent (1958) proposed that information is held in limited capacity short-term storage after it is actively attended to or processed; and through manipulation of that information within short-term storage, it can be transferred into permanent storage, where it is presumably represented as a form of knowledge (Atkinson & Shiffrin, 1968; Broadbent, 1958; Baddeley, 2000; Baddeley & Hitch, 1974; Craik and Tulving, 1975). The effective transfer of information from short-term storage to long-term memory dictates what will be remembered. A number of multi-storage models of memory have been developed since Broadbent's proposal more than half a century ago, and the distinction between short-term storage, working memory and long-term memory remain prominent in cognitive psychology literature (Atkinson & Shiffrin, 1968; Baddeley & Hitch, 1974; Baddeley, 2000; Cowan, 2000; 2008; Ericsson and Kintsch, 1995).

1.2.1 Baddeley's (2000) Model of Working Memory

Early studies referred to the storage of newly acquired information for brief periods of time as short-term memory. A large body of empirical research conducted by Alan Baddeley and colleagues (e.g. Baddeley 1986, 2000, 2002; Baddeley, Eldridge & Lewis, 1981; Baddeley & Hitch, 1974; Baddeley & Wilson, 2002), spanning the course of thirty plus years, suggests that this short-term memory storage is better described as a multiple component working memory system. Baddeley and Hitch (1974) argued that short-term memory was in fact a number of cognitive processes that work together to aid the encoding, storage and retrieval of information within brief periods of time (i.e. working memory; Baddeley, 1986, 2000). Baddeley and Hitch (1974) also distinguished working memory from long-term memory, by stating that unlike the latter, working memory does not involve cognitive processes associated with the construction of mental representational frameworks and the enablement of relatively permanent storage.

Baddeley and Hitch (1974) suggested that working memory is a multi-component system which includes two slave systems, referred to as the phonological loop and the visuospatial sketchpad; and, more recently, a storage centre known as the episodic buffer (Baddeley, 2000; 2002). These components of working memory are governed by a central executive⁶, which integrates the information from the slave

⁶ Alternative models describe long-term memory (LTM) as the governing component of working memory, that is, LTM acts as the *central executive* within working memory (Sweller, 2005). To clarify, this is not to claim that a central executive does not exist, but instead that what has previously been accepted as a central executive (e.g. Baddeley, 1986, 2000, 2002; Baddeley & Hitch, 1974) is actually an additional function of LTM (Sweller, 2005). According to Sweller, schemas (as discussed below) govern the way in which information is processed in working memory for the simple reason that they are organised representations of previously encoded information, which directly aid the processing of novel information in working memory. This view is consistent with Baddeley's (2000, 2002) model of how the episodic buffer functions to support working memory. Sweller disputes Baddeley's conception of the central executive because it is "not feasible for any conception of a central executive apart from a *learned* (i.e. schema-based) central executive to function" (Sweller, 2005, p. 25). For example, Sweller argues, "If schemas are not available, as occurs when dealing with new information, there is no alternative central executive to call upon" (Sweller, 2005, p. 25).

systems by acting as a limited capacity attentional system (Baddeley, 2002; see Figure 1.5). The phonological loop encodes phonological, speech-based information (i.e. what is heard and read). The visuospatial sketchpad deals specifically with visual and spatial information (i.e. what is seen).

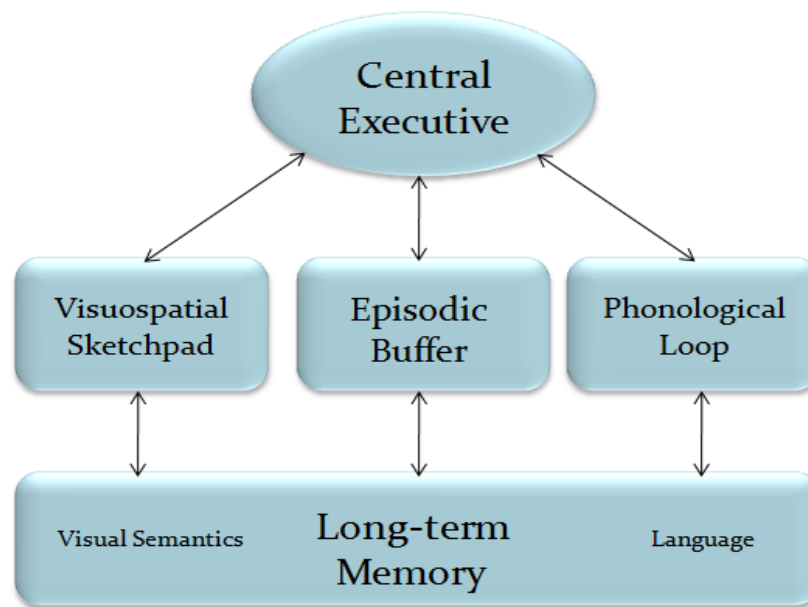


Figure 1.5: Baddeley's (2000) Model of Memory

1.2.1.1 The Slave Systems

Baddeley and Hitch (1974) proposed that both working memory slave systems are critical for the ability to recall information. The visuospatial sketchpad is crucially involved in the memorising of information because it provides one with a work space, or *sketchpad*, in which to manipulate visual stimuli that have been attended to within working memory; and briefly stores this information based on its various visual and spatial characteristics (e.g. colour, shape, orientation and location). At the same time, it is proposed that the phonological loop, which processes auditory and verbal information (Baddeley, 2002), is also crucially involved in the memorisation of information (e.g. while generating a sub-vocal representation of words). Having

access to two slave systems - a visuospatial and a phonological system – provides two routes to potential long-term memory storage and increases the overall capacity base of working memory.

Consistent with the model of working memory proposed by Baddeley and Hitch, research conducted by Paivio (1971; 1986) suggests that visual information and verbal (i.e. phonological) information are processed differently and as a result, separate representations of the information are created. According to Dual-coding Theory, memory for verbal information can be enhanced if a relevant visual aid is simultaneously presented or is imagined (Paivio, 1971; 1986). Likewise, memory for visual information can be enhanced when paired with a relevant verbal or phonological aid.

Building upon Paivio's assertions that separate representations of the same information are created from both coding processes and that their simultaneous use aids memory, Mayer (1997) also suggests that not only do separate visual-spatial and verbal information processing systems exist in short-term memory, but also that the simultaneous use of these systems aid learning. In the context of instructional design, Mayer (2005) refers to the explicit use of these systems for educational purposes as multimedia learning. Specifically, Mayer (1997) proposed that learning is optimised when learners select, organize and integrate verbal and visual information presented to them; and then construct a new internal representation from that information which integrates verbally-based and visually-based models (see Figure 1.6). According to Mayer (1997), in order for integration to take place, both the visual and verbal information must be held in working memory at the same time. However, due to the limited capacity of working memory (discussed below), integration can sometimes be difficult (Chandler & Sweller, 1991; Sweller et al., 1990).

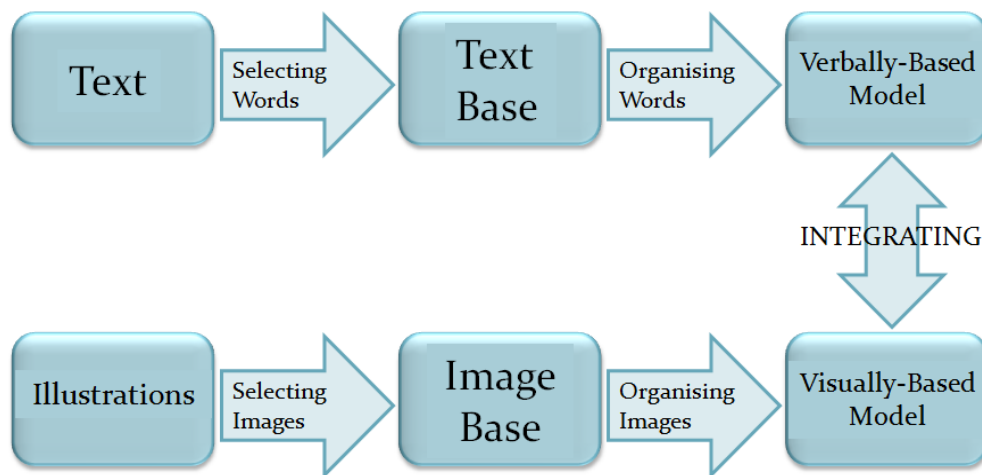


Figure 1.6: Mayer's (1997) Model of the Integration of Visual & Verbal Information

Nevertheless, in a series of eight experiments, Mayer and colleagues (Mayer, 1989; Mayer & Anderson, 1991; 1992; Mayer & Gallini, 1990) highlighted the potential benefits of offering students integrated study materials to work with (i.e. study materials that integrate visual and verbal information into one representation). They compared the problem-solving transfer performance of students who learned about the mechanics of tyre-pumps and braking systems from *integrated* visual and verbal study materials with those who learned from verbal materials *only*. Specifically, these study materials presented students with step-by-step procedures for how tyre-pumps and braking systems work. Problem-solving transfer questions used in this series of studies were, for example, ‘*What can be done to make a pump more reliable, that is, to make sure it would not fail?*’ and ‘*Suppose you push down and pull up the handle of a tyre pump several times but no air comes out. What could have gone wrong?*’

Three of these experiments compared study materials with narration (i.e. verbal) integrated with animation (i.e. visual) and study materials using narration alone (Mayer & Anderson, 1991; 1992); and five experiments compared text integrated with illustrations versus text alone (Mayer, 1989; Mayer & Gallini, 1990).

In all experiments, participants were instructed to study their allocated materials on mechanics and then generate as many solutions as possible to a series of transfer problems. Based on these eight experiments, overall, those who received integrated verbal and visual study materials produced more than 75% more creative solutions to the transfer problems than those who received the learning materials in verbal form only (Mayer, 1997).

Similarly, Mayer and colleagues (Mayer, 1989; Mayer & Anderson, 1991; 1992; Mayer & Sims, 1994; Mayer et al., 1995) conducted a series of 10 experiments comparing the problem-solving transfer performance of students who studied from *integrated* study materials with those who studied the same information from verbal and visual materials *at separate times*. The topics of study were mechanics principles (similar to those described above), the human respiratory system and the development of lightning storms. Based on these 10 experiments, it was observed that, on average, those who received integrated study materials produced more than 50% more creative solutions to transfer problems than those who learned from verbal and visual materials at separate times (Mayer, 1997).

The combined use of both verbal and visual systems of representation to improve learning is supported by numerous research studies (Baddeley, Eldridge & Lewis, 1981; Chandler & Sweller, 1991; Murray, 1968). In addition to Mayer's hypothesis that the combined presentation of visual and verbal information to students facilitates the construction of 'integrated internal representations of information', it has also been proposed that presenting both visual and verbal forms of information to students can help to reduce cognitive load (i.e. demands placed upon an individual in using and distributing working memory resources during cognitive activities; Chandler & Sweller, 1991; see below for a detailed discussion).

1.2.1.2 The Episodic Buffer

Recently, Baddeley (2000) expanded his initial model of working memory to include the episodic buffer. The newly conceptualised episodic buffer is the temporary storage system within working memory, capable of storing information from the visuospatial sketchpad, the phonological loop or integrated information from both slave systems (Baddeley, 2002). This component is episodic because it is assumed to bind information into a unitary episodic representation (e.g. the chronological organisation of events in a narrative, “whereby this information is integrated across (storage) space and potentially extended across time” (i.e. information held in the episodic buffer can be subject to manipulation from long-term memory; Baddeley, 2000, p. 421). Essentially, the episodic buffer is much like the traditional view of short-term memory, as it acts as a limited capacity storage unit for information manipulated by the slave systems.

The episodic buffer also builds upon Ericsson and Kintsch’s (1995) conceptualisation of long-term working memory (LTWM), which is similar to the episodic buffer in that LTWM integrates novel information processed in short-term/working memory with that from long-term storage. Ericsson and Kintsch claim that the existence of LTWM explains remarkably rapid and accurate processing of novel information even when there is a considerable amount of information and/or it is highly complex. They suggest that this accuracy and speed results from the aid of long-term memory in processing novel information. For example, based on the work of Chase & Simon (1973) and de Groot (1965), Ericsson and Kintsch (1995) speculated that some form of established knowledge and expertise (which is represented in long-term memory) must aid chess-masters when assessing an opponent’s move, as the limited capacity of working memory cannot simultaneously

store and process all the information necessary to complete the often complex game assessments that expert chess player engage in.

Baddeley's theory on the episodic buffer is based on empirical research conducted by Wilson and Baddeley (1988) and Baddeley and Wilson (2002), which revealed that amnesiac patients who were not able to encode new information from working memory into long-term storage, were nevertheless able to recall more novel information than what can be stored in short-term storage (Baddeley & Wilson, 2002). This finding suggests that working memory must be in some way aided by information or schemas (see below) in long-term storage, encoded prior to the onset of amnesia. Baddeley (2000, p. 419) further clarifies these results via the following example: "if asked to recall a sequence of unrelated words, subjects typically begin to make errors once the number of words exceeds 5 or 6. However, if the words comprise a meaningful sentence, then a span of 16 or more is possible". According to Baddeley, given that an individual possesses some level of expertise in literacy (i.e. the sentence is meaningful as a result of pre-existing knowledge stored in long-term memory) this 'chunking' of words (Miller, 1956) is thus facilitated by long-term memory and aids in the retention of more information in the episodic buffer.⁷ Nevertheless, in order to *permanently* store information, efforts must be made to transfer information from working memory to long-term memory.

1.2.2 Long-term Memory

Whereas working memory can store a limited amount of information for a limited amount of time, long-term memory (LTM) is a region of memory that enables relatively permanent storage of information, for example, facts in *semantic* LTM,

⁷ Based on Baddeley's research, the episodic buffer may potentially be considered the work space where higher-order cognitive skills (i.e. critical thinking and reflective judgment) engage novel information and draw from pre-existing knowledge in order to draw conclusions, judge situations or solve problems.

events in *episodic* LTM, and procedures in *procedural* LTM (Tulving, 1984). The transfer of information from working memory to LTM depends crucially on the manner in which the information is encoded (Tulving & Thompson, 1973; Tulving, 1984). Encoding refers to the efforts made to organise information in working memory, for example, through rehearsal (Craik & Watkins, 1973) and/or schema construction (Craik, 1983) for purposes of storage in LTM and potential retrieval. The success of encoding depends on the amount, or depth of information processing in working memory. For example, research suggests that more successful encoding results from deep, semantic processing of information, whereas shallow processing of information (e.g. a focus on the colour, as opposed to the meaning of words) may result in poor subsequent recall (Craik, 1983). Furthermore, the likelihood of information being stored in LTM is increased when it is encoded into one's existing schemata/schemas (i.e. representations of knowledge that have been assembled from previous experience, which function as a plan or a set of expectations that guide subsequent information processing).

1.2.2.1 Schemas

Though there is no single definition of the concept of a schema, various descriptions have been offered. Bartlett (1932) described a schema as an active organisation of past reactions or experiences. Another definition is provided by Sweller (1999), who describes a schema as:

“a cognitive construct that permits people to treat multiple elements of information as a single element categorised according to the manner in which it will be used” (Sweller, 1999, p. 10).

In the context of memory research and theory, cognitive psychologists often assume that schemas are large cognitive structures within LTM (Neisser, 1976), which can be

used to facilitate the assimilation of new information. Schemas can also be used to build knowledge structures by organising elements of information. For example, smaller, specific schemas (i.e. lower-order schemas) can be reconstructed into a larger, more comprehensive and complex schema (i.e. a higher-order schema).

The role of schemata in memory is complex. Information that is the focus of active processing in working memory can be processed and organized in many different ways, depending on the manner in which items of information are classified and arranged into systems of representation. These active, ongoing systems of representation may be transformed not only by new incoming information from the environment but also by pre-existing schemas in LTM. For example, to determine the criteria people use for inclusion of information into certain categories and the method of categorisation they use, Chi, Glaser and Rees (1982) asked two groups with different levels of expertise (novice and expert) to categorise 24 physics problems based on their similarities. Though both groups identified approximately the same number of categories, qualitative analysis revealed that novices (i.e. those with insufficient, relevant schemata) categorised the problems according to either the objects referred to in the problem (e.g. a spring), the keywords that have meaning in physics (e.g. friction), or the interaction or configuration of various objects (e.g. a block on an inclined plane). Conversely, it was found that experts (i.e. those with sufficient, relevant schemata) categorised the problems according to the law of physics that governed each problem (i.e. the solution method).

The authors argued that such expert solution methods are *higher-order schemas* because they are coordinated in the context of mathematical formulae and computational systems of relations between abstract and concrete features of the problem. The authors also argued that these solutions were more advanced and

abstract than the *lower-order schemas* possessed by the novices, which largely focused on concrete aspects of the problem. Thus, consistent with many developmental accounts that focus on knowledge growth and the levels of complexity and integration of concrete and abstract forms of representation (e.g. Fischer, 1980; Piaget, 1952), schema theories of memory often assume that schemas can have subordinate schemata embedded in them (e.g. a system of concrete representations embedded in a lower-order schema) and can also be embedded in super-ordinate schemata (e.g. a system of abstractions embedded in a higher-order schema; see Figure 1.7).

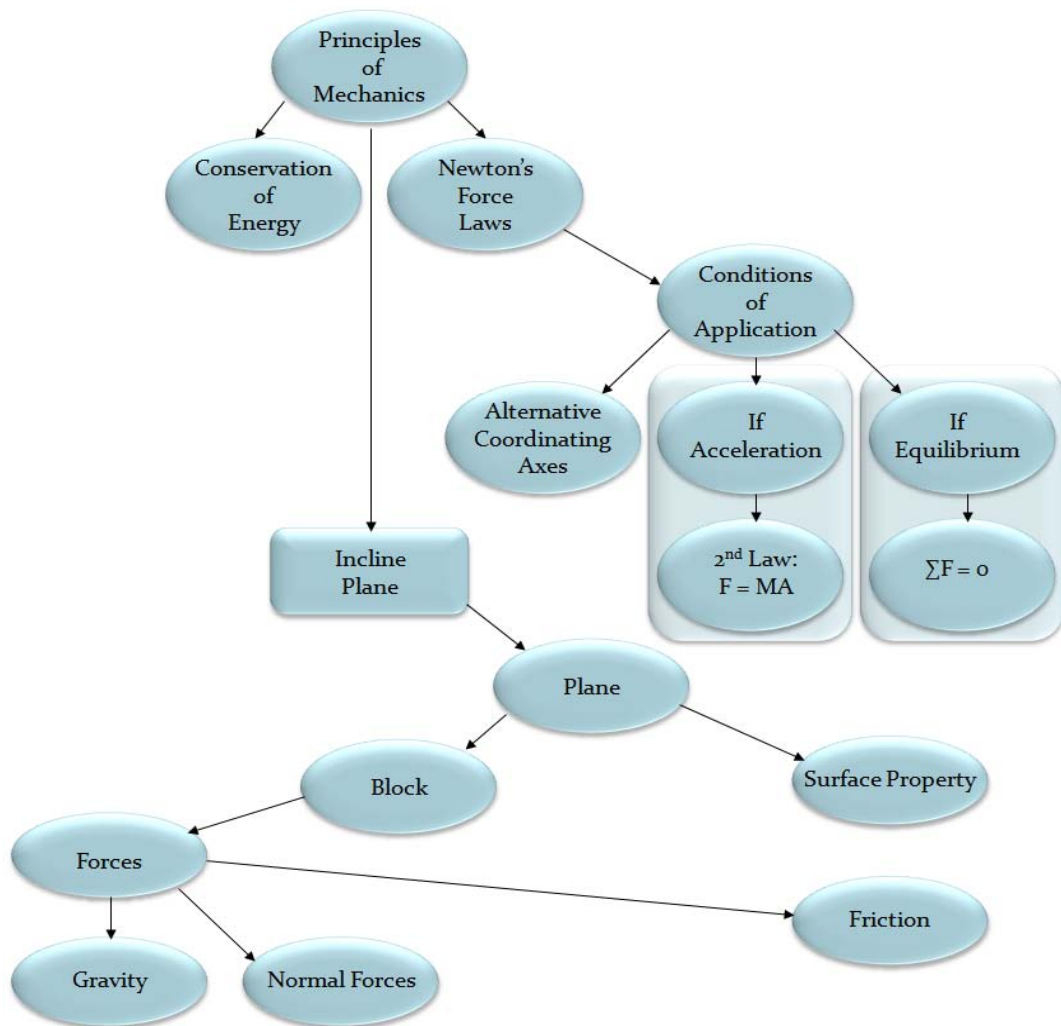


Figure 1.7: Expert Schema for Principles of Mechanics (Chi, Glaser & Rees, 1982)

As part of LTM, schemas aid working memory by providing a system of representations that facilitate the encoding, storage, and retrieval of information. When novel information is processed within working memory, any number of schemas can be used to provide relevant knowledge to assist the processing of the novel information. In Baddeley's model, schemas may act upon information in the episodic buffer and thus facilitate the reconfiguration and transfer of information into LTM. In this way, novel information can be integrated into existing schemas in LTM, thus freeing up space within working memory. The role of the schema, in this context, is very important as processing space within working memory is limited.

The findings of Chi, Glaser and Rees (1982) as well as similar research by Chase and Simon (1973) and Kotovsky, Hayes and Simon (1985) suggest that once information is adequately encoded into a schema, it is treated as stored knowledge (Sweller, 1999). To elaborate, in order to create knowledge, schemas must be constructed in a meaningful way; that is, specifically linked with the context in which they are to be used, in order for them to be retrieved or to aid in the encoding of new information at a later time. For schemas to be constructed in a meaningful way, the information subject to schema construction must be understood, or comprehended (Sweller, 2005; 2010).

1.2.2.2 Comprehension as LTM

In addition to the recall of knowledge, comprehension (i.e. the second outcome variable examined in this thesis) also shares interdependency with schema-construction, given that schema-construction, according to some theorists, is essentially the same as building understanding, or comprehension (Pollock, Chandler & Sweller, 2002; Sweller, 2005). According to Bloom (1956), comprehension is the ability to understand or grasp the meaning of information, which implies the ability to

translate information from one level of abstraction to another, one symbolic form to another, or one verbal form to another. Bloom's taxonomy further describes comprehension as the confirmation of knowledge, in the sense that knowledge can be confirmed via explaining, summarising, paraphrasing, or illustrating information based on prior learning (Huitt, 2011).

A more recent conceptualisation of comprehension is the view developed by Sweller (2005). Broadly speaking, Sweller (2005, p. 21) describes comprehension as "changes in LTM, along with the effect of those changes on working memory. Without changes in LTM, nothing has been understood." The nature of *changes* in LTM that Sweller speaks of refers specifically to schema construction. Sweller further describes comprehension, or understanding, as the ability to integrate schemas from LTM with novel information simultaneously in working memory. See Figure 1.8 for a diagram of the relationships among working memory, LTM and comprehension in this context. Sweller (1994) also claims that the acquisition of knowledge (i.e. in LTM) is dependent upon schema construction, because it is only once a schema (i.e. knowledge) has been constructed that information can be understood or comprehended. Sweller (1999) further simplifies his conceptualisation of comprehension by describing it as the ability to make required connections between novel items of information and/or schemas.

Making the necessary connections between new items of information and pre-existing knowledge is important in educational settings because it allows students to gain understanding of novel information and create new levels of comprehension. Subsequently, such understanding and knowledge is applied by students to answer questions, draw conclusions and solve problems. In order to develop a reasonable answer, conclusion or solution, students must reflect upon their own thinking

processes and often the thinking of others as well. Thus, consistent with Anderson and Krathwohl (2001) and Marzano (2001), the ability to apply knowledge and understanding successfully often depends on an individual's metacognitive abilities.

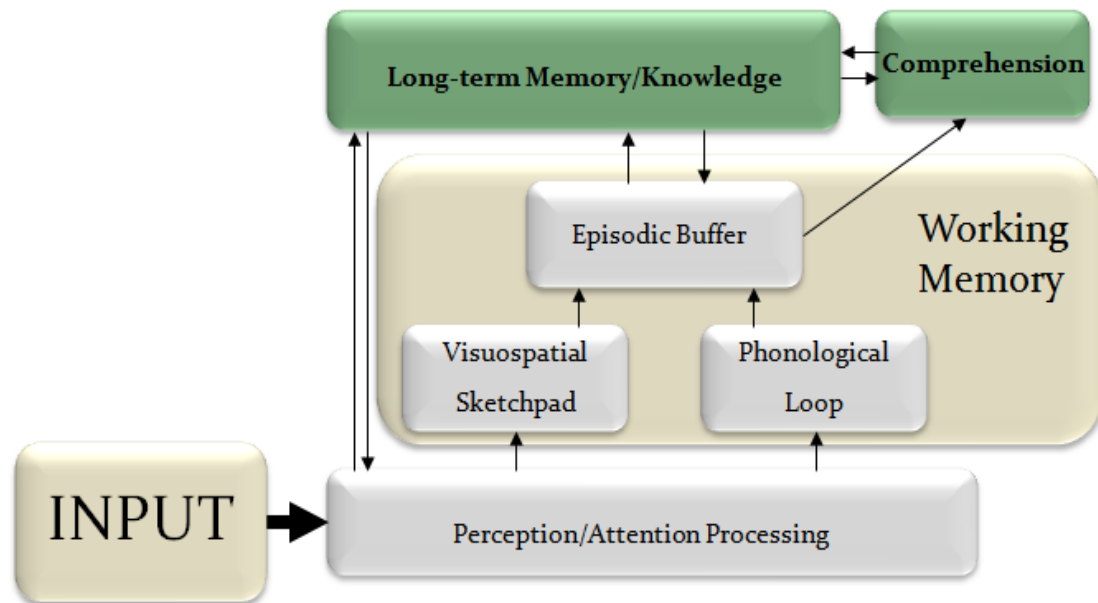


Figure 1.8: Working Memory as a buffer between informational input and storage in LTM as knowledge

1.3 Metacognition

Though the term *metacognition* was not used by Bloom, many modern conceptualizations of metacognition are similar to what he described as higher-order thinking processes. For example, Wegerif (2002, p. 6), has described metacognition as being “another term often used as a synonym for thinking skills or higher-order thinking” which “originates in an information processing model of the mind as something like a computer running both low-level software, to do the basic cognitive processes and high-level software, to monitor and correct the low-level software.” Metacognition was first described by Flavell (1976, p. 232) as “knowledge concerning one’s own cognitive processes and products or anything related to them; and the active monitoring, consequent regulation and orchestration of these

processes". According to Boekaerts and Simons (1993), Brown (1987) and Ku and Ho (2010b), individuals think metacognitively in two ways: first, individuals must be aware of their own cognitive processes (e.g. through self-monitoring or self-regulation); second, individuals must be able to apply available cognitive processes for purposes of learning or devising solutions to problems (e.g. using critical thinking or reflective judgment; Dawson, 2008a; as discussed below). These concepts are also reflected in other definitions of metacognition:

- The higher-order control processing used in executive planning and decision-making (Sternberg, 1985, p. 226).
- The higher-order cognitions that supervise a person's thoughts, knowledge and actions (Weinert, 1987).
- The awareness of one's own knowledge and the ability to understand, control and manipulate individual cognitive processes (Osman & Hannafin, 1992, p. 83).
- One's ability to consciously think about thinking as a self-regulatory function, that is, the monitoring of one's own cognitive activities, as well as the results of those activities (Demetriou, 2000).
- Thinking about thinking, usually conceptualised as an interrelated set of competencies for learning and thinking, and include many of the skills required for active learning, critical thinking, reflective judgment, problem-solving and decision-making (Dawson, 2008a, p. 4).
- Knowing one's cognitive processes and the strategies one applies to control these processes (Ku & Ho, 2010b, p. 263).

Though Marzano (2001) distinguishes the self-regulatory functions of thinking from dispositional factors, as acting in distinct systems (i.e. a *metacognitive system* and a *self-system*, respectively); consistent with definitions of metacognition that have been developed in the literature (e.g. Boekaerts and Simons, 1993; Demetriou, 2000; Ku & Ho, 2010b), the conceptualisation of metacognition used in this thesis is characterised by both the dispositional/self-regulatory functions of thinking, as well as the strategical planning and application of higher-order thinking processes (i.e. analysis, evaluation and inference) when thinking about thinking (e.g. Brown, 1987; Flavell, 1979; Ku & Ho, 2010b).

In the context of education research, the dispositional/self-regulatory functions of thinking can refer to, for example, dispositions towards thinking, epistemological beliefs, motivation to think and learn, and perceived need for cognition. Disposition towards thinking refers to the extent to which an individual is *disposed*, or willing, to perform a given thinking skill (Valenzuela, Nieto & Saiz, 2011) and can include the disposition towards truth-seeking, open-mindedness, analyticity, systematicity, confidence, inquisitiveness and maturity (cf. Facione & Facione, 1992). Epistemological beliefs refer to an individual's beliefs about the nature of knowledge (Schommer-Aikins, 2002), as well as the beliefs about how they justify knowledge (King & Kitchener, 1994). Epistemological beliefs include beliefs in relation to the malleability of knowledge (e.g. 'the belief that knowledge is fixed at birth' or 'the belief that the ability to learn can be improved'), the structure of knowledge (e.g. 'belief that knowledge is best characterized as complex interrelated networks') and the stability of knowledge (e.g. 'the belief that knowledge is unchanging' or the 'belief that knowledge is evolving'). Motivation towards thinking and learning includes, for example, the motivation to regulate effort, thinking processes and

learning beliefs (Pintrich et al., 1991). Finally, perceived need for cognition refers to the willingness to explore and engage in relatively complex cognitive activities (Cacioppo, Petty & Kao, 1984). Though these dispositional, self-regulatory functions of thinking are important to consider as part of any effort to understand how an individual applies their cognitive and metacognitive skills, the discussion below now turns to the higher-order thinking abilities of analysis, evaluation, inference and reflective judgment, which are central to definitions of *critical thinking*. The dispositional/self-regulatory processes of metacognition are discussed in more detail in Chapters 5 and 6, where they are introduced as empirically analysed variables.

1.3.1 Critical Thinking

Critical thinking is a metacognitive process that consists of a number of sub-skills (i.e. analysis, evaluation and inference) that, when used appropriately, increases the chances of producing a logical solution to a problem or a valid conclusion to an argument. On one hand, the development of critical thinking skills in educational settings is often endorsed because it is believed that these skills allow students to transcend lower-order, memorisation-based learning strategies to gain a more complex understanding of the information or problems they encounter (Halpern, 2003a; Pollock, Chandler & Sweller, 2002). On the other hand, in the past century, there has been little agreement on how to define critical thinking (CT), resulting in the existence of many diverse conceptualisations of CT (e.g. Bensley, 1998; Dewey, 1910; 1933; Ennis, 1987; Glaser, 1941; Halpern, 2003a; Paul, 1993).

Historically, Dewey's (1933, p. 8) conceptualisation of reflective thought as "active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it and the further conclusions to which it tends" helped give birth to the concept of CT (Moseley et al., 2005; Paul,

Elder & Bartell, 1997). Recognition of the importance of CT in education also followed the growth of interest in informal logic initiated in part by the work of Stephen Toulmin in the late 1950's (Allen, Feezel & Kauffeld, 1967; Toulmin, 1958). Informal logic (see Chapter 3) is a type of logic that emphasises the justificatory function of argumentation, namely that a good argument requires sufficient support (e.g. reliable and valid empirical evidence). Deliberations in relation to CT skills grew in part from the notion of informal logic, in that, claims could only be made (and justified) after a sufficient amount of analysis and evaluation had been conducted on propositions and their logical interdependencies. Since then, many definitions of CT have been offered (see Table 1.1).

Though debate is ongoing over the definition of CT and the core skills necessary to think critically, one definition and list of skills stands out as a reasonable consensus conceptualisation of CT. In 1988, a committee of 46 experts in the field of CT gathered to discuss both a definition and the skills necessary to think critically. The report of the findings of this meeting, known as *The Delphi Report*, defined CT as:

“...purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based” (Facione, 1990b, p. 3).

Furthermore, the Delphi panel overwhelmingly agreed (i.e. 95% consensus) that *analysis, evaluation and inference* were the core skills necessary for CT (Facione, 1990b; see Table 1.2 for the description of each skill provided by the Delphi Report).

Table 1.1: Past Definitions and Descriptions of Critical Thinking

Author	Definition/Description
Glaser (1941)	Critical thinking is: an attitude of being disposed to consider, in a thoughtful way, problems and subjects that come within the range of one's experience; knowledge of the methods of logical enquiry and reasoning; and some skills in applying those methods. Critical thinking calls for a persistent effort to examine any belief or supposed form of knowledge in the light of the evidence that supports it and the further conclusions to which it tends.
Ennis (1987)	Critical thinking is reasonable, reflective thinking focused on deciding what to believe or do.
Kurfiss (1988)	Critical thinking is the ability to detect and avoid fallacious reasoning and to analyse deductive and inductive arguments.
Allegretti & Frederick (1995)	Critical thinking involves evaluating the arguments of others, evaluating one's own arguments, resolving conflicts and understanding the source of conflicts in argumentation; thus coming to a resolution regarding complex problems and gaining confidence in one's own thinking processes.
Paul (1993)	A unique kind of purposeful thinking, in which the thinker systematically and habitually imposes criteria and intellectual standards upon thinking, taking charge of the construction of thinking, guiding the construction of the thinking according to the standards, assessing the effectiveness of the thinking according to the purpose, the criteria, and the standards.
Wilkinson (1996)	Critical thinking is goal-oriented, purposeful thinking that involves a number of mental skills, such as determining what data is relevant, evaluating the credibility of sources and making inferences.
Bensley (1998)	Critical thinking is reflective thinking in which a person evaluates relevant evidence and works to draw a sound or good conclusion.
Halpern (2003a)	Critical thinking is purposeful, reasoned and goal-directed thinking – the kind of thinking involved in solving problems, formulating inferences, calculating likelihoods and making decisions.
Thomson (2009)	Critical thinking involves the identification and evaluation of reasons and conclusions within an argument, the ability to draw one's own conclusions and the use of appropriate language in order to communicate and construct one's own arguments.

1.3.1.1 Analysis

The Delphi report describes *analysis* as a CT skill that is used in the context of argumentation to detect, examine and identify the propositions within an argument and the role they play; for example, the main conclusion, the premises and reasons provided to support the conclusion, objections to the conclusion and inferential relationships among propositions (Facione, 1990b). Notably, at the core of the Delphi definition of analysis is the ability of an individual to analyse the structure of an argument, which depends not only on their knowledge and skill as a reader/listener, but also on the way in which the author of the argument uses relational cues, or signals, that guide the reader/listener (Meyer, Brandt & Bluth, 1980). For example, words like *but*, *because* and *however* can be used by the author to indicate that propositions that follow are *objections*, *reasons*, or *rebuttals* for propositions that have come before.

In the context of reading, these relational cues also shape the organisation of paragraphs within an argument, as well as the various paths of reasoning an argument may take based on the evidence presented within. For example, an author may decide to present all their reasons in support of a particular claim in the first few paragraphs of their argument and follow this with a series of paragraphs presenting all the objections in relation to a particular claim, followed by justification for these objections, and/or rebuttals. Alternatively, an author may choose to mix reasons, objections, and rebuttals throughout all of these paragraphs and thus sequence and organise their argument in different ways. The organisation and identification of propositions within an argument is critical for the reader, as the structure of propositions has been found to affect the reader's ability to comprehend the

Table 1.2: Core CT Skills According to the Delphi Report (adapted from Facione, 1990b)

Skill	Description
<u>Analysis</u>	<p>To identify the intended and actual inferential relationships among statements, questions, concepts, descriptions or other forms of representation intended to express beliefs, judgments, experiences, reasons, information or opinions.</p>
	<p>Examining ideas: to determine the role various expressions play or are intended to play in the context of argument, reasoning or persuasion; to compare or contrast ideas, concepts, or statements; to identify issues or problems and determine their component parts, and also to identify the conceptual relationships of those parts to each other and to the whole.</p>
	<p>Detecting arguments given a set of statements or other forms of representation, to determine whether or not the set expresses, or is intended to express, a reason or reasons in support of or contesting some claim, opinion or point of view.</p>
	<p>Analysing arguments: given the expression of a reason or reasons intended to support or contest some claim, opinion or point of view, to identify and differentiate: (a) the intended main conclusion, (b) the premises and reasons advanced in support of the main conclusion, (c) further premises and reasons advanced as backup or support for those premises and reasons intended as supporting the main conclusion, (d) additional unexpressed elements of that reasoning, such as intermediary conclusions, non-stated assumptions or presuppositions, (e) the overall structure of the argument or intended chain of reasoning, and (f) any items contained in the body of expressions being examined which are not intended to be taken as part of the reasoning being expressed or its intended background.</p>
<u>Evaluation</u>	<p>To assess the credibility of statements or other representations which are accounts or descriptions of a person's perception, experience, situation, judgment, belief or opinion; and to assess the logical strength of the actual or intended inferential relationships among statements, descriptions, questions or other forms of representation.</p>
	<p>Assessing claims: to recognize the factors relevant to assessing the degree of credibility to ascribe to a source of information or opinion; to assess the contextual relevance of questions, information, principles, rules or procedural directions; to assess the acceptability, the level of confidence to place in the probability or truth of any given representation of an experience, situation, judgment, belief or opinion.</p>
	<p>Assessing arguments: to judge whether the assumed acceptability of the premises of an argument justify one's accepting as true (deductively certain), or very probably true (inductively justified), the expressed conclusion of that argument; to anticipate or to raise questions or objections, and to assess whether these point to significant weakness in the argument being evaluated; to determine whether an argument relies on false or doubtful assumptions or presuppositions and then to determine how crucially these affect its strength; to judge between reasonable and fallacious inferences; to judge the probative strength of an argument's premises and assumptions with a view toward determining the acceptability of the argument; to determine and judge the probative strength of an argument's intended or unintended consequences with a view toward judging the acceptability of the argument; to determine the extent to which possible additional information might strengthen or weaken an argument.</p>

Table 1.2: Core CT Skills According to the Delphi Report (continued)

Skill	Description
<u>Inference</u>	<p>To identify and secure elements needed to draw reasonable conclusions; to form conjectures and hypotheses; to consider relevant information and to deduce the consequences flowing from data, statements, principles, evidence, judgments, beliefs, opinions, concepts, descriptions, questions or other forms of representation.</p> <p>Querying evidence: in particular, to recognize premises which require support and to formulate a strategy for seeking and gathering information which might supply that support; in general, to judge that information relevant to deciding the acceptability, plausibility or relative merits of a given alternative, question, issue, theory, hypothesis, or statement is required, and to determine plausible investigatory strategies for acquiring that information.</p> <p>Conjecturing alternatives: to formulate multiple alternatives for resolving a problem, to postulate a series of suppositions regarding a question, to project alternative hypotheses regarding an event, to develop a variety of different plans to achieve some goal; to draw out presuppositions and project the range of possible consequences of decisions, positions, policies, theories, or beliefs.</p> <p>Drawing conclusions: to apply appropriate modes of inference in determining what position, opinion or point of view one should take on a given matter or issue; given a set of statements, descriptions, questions or other forms of representation, to deduce, with the proper level of logical strength, their inferential relationships and the consequences or the presuppositions which they support, warrant, imply or entail; to employ successfully various sub-species of reasoning, as for example to reason analogically, arithmetically, dialectically, scientifically, etc; to determine which of several possible conclusions is most strongly warranted or supported by the evidence at hand, or which should be rejected or regarded as less plausible by the information given.</p>

information within the argument (e.g. Meyer, Brandt & Bluth, 1980; Munch, Boller & Swasy, 1993, Myers, 1974).

1.3.1.2 Evaluation

The Delphi report describes *Evaluation* as a CT skill that is used in the assessment of propositions and claims in terms of their credibility, relevance and the logical strength of their relationships with other propositions; thus deciding the overall strength or weakness of an argument (Facione, 1990b). Evaluating the credibility of claims and arguments involves progressing beyond merely identifying the source of propositions in an argument, to actually examining the credibility of

those identified sources (e.g. personal experiences, common beliefs/opinions, expert/authority opinion and scientific evidence). Evaluation also implies deep consideration of the relevance of claims within an argument, which is accomplished by assessing the contextual relevance of claims and premises (i.e. the pertinence or applicability of one proposition to another). Finally, evaluating the logical strength of an argument is accomplished by monitoring both the logical relationships amongst propositions and the claims they infer.

1.3.1.3 Inference

Like Bloom's (1956) conceptualisation of synthesis, the final CT skill, *inference*, involves the "gathering" of credible, relevant and logical evidence based on the previous analysis and evaluation of available evidence, for the purposes of "drawing a reasonable conclusion" (Facione, 1990b, p.9). Drawing a conclusion always implies some act of synthesis. However, inference is a unique form of synthesis in that it involves the formulation of a set of consequences and conclusions that are derived from a set of arguments or a body of evidence. This may imply accepting a conclusion pointed to by an author in light of the evidence they present, or "conjecturing an alternative", equally logical, conclusion or argument based on the available evidence (Facione, 1990b, p.9). According to the Delphi definition, another important aspect of inference is "querying the evidence" available, for example, by recognising the need for additional information or justification and by being able to gather such additional information or justification to draw a conclusion; and to judge the plausibility of utilising such additional information or justification for purposes of conjecturing an alternative conclusion. Notably, in the context of querying evidence and conjecturing alternative conclusions, inference overlaps with evaluation to a certain degree in that both skills are used to judge the relevance and acceptability of a

claim or argument. Nevertheless, it remains necessary to query and judge the inclusion of propositions within an argument, before gathering them to draw a conclusion.

The definition of CT provided by the Delphi Report was adopted by the American Philosophical Association and as a result, became the accepted definition for good CT (Beckie, Lowry & Barnett, 2001). The Delphi definition of CT also inspired the creation of the California Critical Thinking Skills Test (Facione, 1990a), which is a commonly used assessment of CT performance that measures CT according to test-takers' ability to use analysis, evaluation and inference skills. Though the Delphi Report has shed some light on what CT is, how to conceptualise it and how to measure it, at the same time, it is often acknowledged that CT skills take time to develop (Dawson, 2008a; Halpern, 2003a; King & Kitchener, 1994; Kuhn, 1999). However, there is no consensus as to how best develop CT skills. In order for CT to develop to a high level, related metacognitive processes may be needed to support CT skill development and to aid in the successful application of CT to real-world problems. Reflective judgment is one such metacognitive process that can aid in the support, development and application of CT, particularly in the context of real-world problems.

1.3.1.4 Reflective Judgment

The ability to think about thinking (Flavell, 1976; Ku & Ho, 2010b) and the ability to apply CT skills to a particular problem implies a reflective sensibility and the capacity for reflective judgment (King & Kitchener, 1994). Like CT, reflective judgment is an important skill for students to acquire and practice, because it may facilitate their ongoing acquisition and application of knowledge both inside and outside of school and university (Folsom-Kovarik et al., 2010; Huffaker & Calvert,

2003; U.S. National Research Council, 2002). According to King and Kitchener (1994), reflective judgment (RJ) is an individuals' understanding of the nature, limits, and certainty of *knowing* and how this can affect how they defend their judgments and reasoning in context. Moreover, RJ involves the ability of an individual to acknowledge that their views might be falsified by additional evidence obtained at a later time (King & Kitchener, 1994).

The ability to acknowledge levels of certainty and uncertainty when engaging in CT is important because sometimes the information a person is presented with (along with that person's pre-existing knowledge) provides only a limited source of information from which to draw a conclusion. This is often the case when a person is presented with an ill-structured problem (King, Wood & Mines, 1990), that is, a problem that cannot be solved with absolute certainty (Wood, 1993). Specific thinking skills are necessary when people realise that some problems cannot be solved with certainty (Dewey, 1933; King & Kitchener, 2004; Wood, 1993); that is, in the context of uncertainty, a combination of CT skills (as defined by the Delphi committee) and RJ (as defined by King & Kitchener) is necessary in situations where one seeks to arrive at a reasonable conclusion or decide upon a reasonable course of action.

RJ is often used when an ill-structured problem is encountered, where the uncertainty associated with the problem indicates that multiple paths of reasoning and action are possible (e.g. "*What is the best way of decreasing global warming?*"). Such encounters often lead thinkers to reasonably consider multiple, alternative solutions (e.g. "*Make everyone drive electric cars*", or, "*Cut down on cattle farming in order to lower methane emissions*"). However, some solutions are deemed better than others based on the organisation, complexity and careful consideration of the

propositions within an argument (e.g. in comparison with the unsupported singular claims above, a more complex and better considered response might propose that *‘Although research is still ongoing in this area, mathematical models based on existing research findings suggest that by making small decreases in emissions in all walks of life, whether it be travel, farming, industry or energy production, emissions around the globe will decrease substantially’*). Therefore, it is not only the conclusion one reaches, or the inference one draws, correct or otherwise; but also the manner in which one *arrives* at the conclusion which is important in RJ and CT. This description of RJ, as involving inferential, CT processes, further suggests that there is an interdependence between RJ and CT.

RJ is often considered as a component of CT (Baril et al., 1998; Huffman et al., 1991), because RJ allows one to acknowledge that epistemic assumptions (i.e. assumptions about one’s knowledge) are vital to recognising and judging a situation in which CT may be required (King & Kitchener, 1994). RJ may also influence how well an individual applies each CT skill (King, Wood & Mines, 1990). This interdependence between CT and RJ is also consistent with Kitchener and King’s (1981) Reflective Judgment Model (RJM; see Table 1.3), in which CT is embedded explicitly.

Research suggests that like CT skills, child and adult development may see a progressive development of RJ ability toward greater levels of complexity and skill. Kitchener and King (1981) created the RJM in order to characterise the development of people’s RJ ability. The RJM describes changes in the thinker’s recognition of limited knowledge (i.e. uncertainty) and how these changes influence other thinking skills, such as analysis, evaluation and inference. Within the RJM, a series of developmental changes occur in the way people come to understand the process of

knowing and reasoning. More importantly, research supports a developmental trajectory of RJ along the lines described by King and Kitchener (1994). However, RJ development is not a simple function of age or time, but more so a function of the amount of interaction, or active engagement an individual has in the context of working on ill-structured problems, such that the development of higher levels of reasoning and RJ ability can emerge (Brabeck, 1981; Dawson, 2008a; Fischer & Bidell, 2006).

Table 1.3: The Reflective Judgment Model (Adapted from King & Kitchener, 2002)

Period	Stage	Description
<u>Pre-reflective thought</u>	1	Knowledge is assumed to exist absolutely and concretely; it is not understood as an abstraction. It can be obtained with certainty by direct observation. Beliefs need no justification since there is assumed to be an absolute correspondence between what is believed to be true and what is true. Alternate beliefs are not perceived.
	2	Knowledge is assumed to be absolutely certain or certain but not immediately available. Knowledge can be obtained through direct observation or via authority figures. Beliefs are unexamined and unjustified or justified by their correspondence with the beliefs of an authority figure. Most issues are assumed to have a right answer, so there is little or no conflict in making decisions about disputed issues.
	3	Knowledge is assumed to be absolutely certain or temporarily uncertain. In areas of temporary uncertainty, only personal beliefs can be known until absolute knowledge is obtained. In areas of absolute certainty, knowledge is obtained from authorities. In areas in which certain answers exist, beliefs are justified by reference to authorities' views. In areas in which answers do not exist, beliefs are defended as personal opinion since the link between evidence and beliefs is unclear.
<i>Quasi-reflective thought</i>	4	Knowledge is uncertain and knowledge claims are idiosyncratic to the individual since situational variables dictate that knowing always involves an element of ambiguity. Beliefs are justified by giving reasons and using evidence, but the arguments and choice of evidence are idiosyncratic.
	5	Knowledge is contextual and subjective since it is filtered through a person's perceptions and criteria for judgment. Only interpretations of evidence, events, or issues may be known. Beliefs are justified within a particular context by means of the rules of inquiry for that context and by the context-specific interpretations as evidence. Specific beliefs are assumed to be context specific or are balance against other interpretations, which complicates conclusions.

Table 1.3: The Reflective Judgment Model (continued)

Period	Stage	Description
<i>Reflective thought</i>	6	Knowledge is constructed into individual conclusions about ill-structured problems on the basis of information from a variety of sources. Interpretations that are based on evaluations of evidence across contexts and on the evaluated opinions of reputable others can be known. Beliefs are justified by comparing evidence and opinion from different perspectives on an issue or across different contexts and by constructing solutions that are evaluated by criteria such as the weight of the evidence, the utility of the solution, and the pragmatic need for action.
	7	Knowledge is the outcome of a process of reasonable inquiry in which solutions to ill-structured problems are constructed. The adequacy of those solutions is evaluated in terms of what is most reasonable or probable according to the current evidence, and it is re-evaluated when relevant new evidence, perspectives, or tools of inquiry become available. Beliefs are justified probabilistically on the basis of a variety of interpretive considerations, such as the weight of the evidence, the explanatory value of the interpretations, the risk of erroneous conclusions, consequences of alternative judgments, and the interrelationships of these factors. Conclusions are defended as representing the most complete, plausible, or compelling understanding of an issue on the basis of the available evidence.

Kitchener and King's RJM is a seven stage model that is broken down into three periods of development. Progress on the RJM (from one stage to another; and from one period to another) is a type of evolution of RJ, in which each progression marks the increasing complexity of the thinking required to justify a belief. The more developed one's RJ, the better able one is to present "a more complex and effective form of justification, providing more inclusive and better integrated assumptions for evaluating and defending a point of view" (King & Kitchener, 1994, p. 13). Notably, King and Kitchener describe reflective judgment here as a collation of the sub-components of evaluation and inference (i.e. CT skills; again, see Table 1.2) - once again indicating the importance of the interdependency between CT and RJ.

The relationship between CT and RJ has been confirmed in a number of research studies and the suggestion has been made that CT and RJ develop in an interdependent, cyclical manner (Brabeck, 1981; Dawson, 2008a; King & Kitchener, 1994; King, Wood & Mines, 1990; see Figure 1.9). For example, Brabeck (1981) assessed 119 university students on both CT and RJ ability. CT was measured using the Watson-Glaser Critical Thinking Appraisal (WGCTA, Watson & Glaser, 1980) and RJ was measured using the Reflective Judgment Interview (Kitchener & King, 1981). Results revealed a positive correlation between both measures ($r = .40, p < .001$). There was also a significant difference between high-scoring and low-scoring critical thinkers on RJ performance.

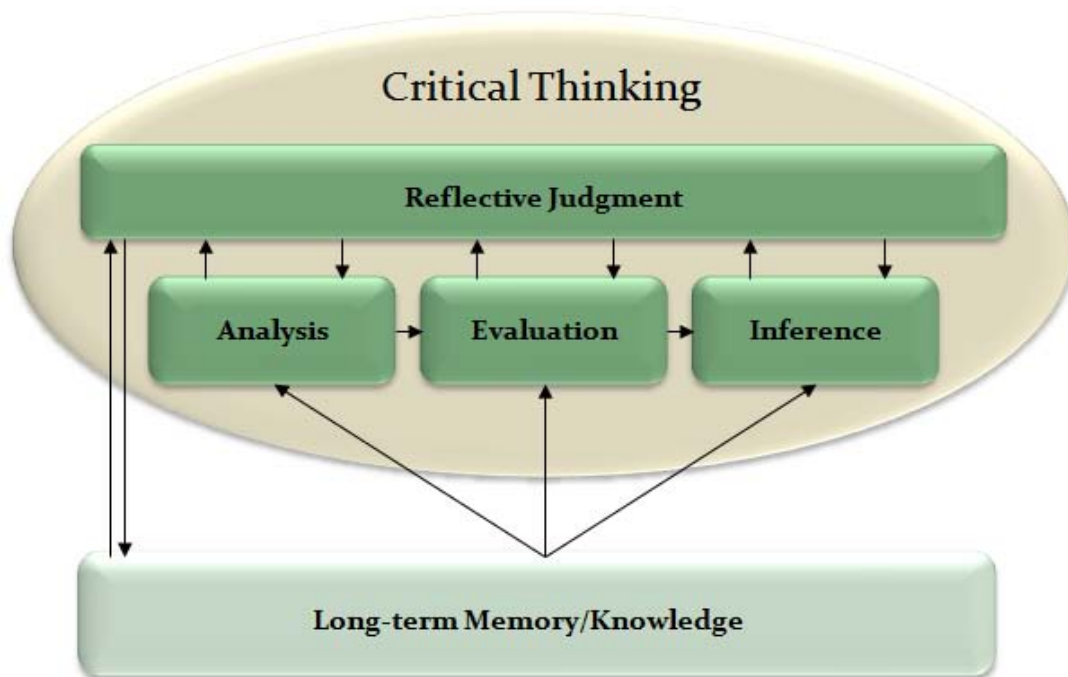


Figure 1.9: The Interdependencies among Critical Thinking Skills and Reflective Judgment

Based on these findings, Brabeck (1981) suggested that there is an inextricable link between CT and RJ. This link was confirmed by further research conducted by King, Wood and Mines (1990), which examined the RJ and CT performance of both

university undergraduate and graduate students. Results revealed a significant correlation between RJ (as measured by the Reflective Judgment Interview) and CT, measured using both the WGCTA and the Cornell Critical Thinking Test (Ennis, Millman & Tomko, 1985; $r = .46$, $p < .01$ for both).

Though this research does indicate that RJ and CT are significantly correlated, it is less clear from the literature how the development of RJ might facilitate the development of CT and vice versa. While RJ models are traditionally developmental (e.g. Kitchener & King, 1981; King & Kitchener, 1994; 2002), models of CT do not provide a detailed account of how specific CT sub-skills of analysis, evaluation and inference develop (e.g. Ennis, 1998; Halpern, 2003a; Paul, 1993). For example, it may be that a certain level of RJ ability is needed before a student can begin to understand and apply certain CT sub-skills (e.g. evaluating logical strength and inferring plausible conclusions or alternatives). Thus, further research is necessary to provide more than correlational evidence alone in support of the link between RJ and CT development. Nevertheless, it is reasonable to assume that through the acknowledgment of uncertainty in decision-making and problem-solving, an individual with good RJ skills will be able to apply CT skills with caution and awareness of the alternative conclusions and/or solutions that may be drawn. Consistent with this view, it has been noted that those who show good RJ are more likely to exhibit greater care when *applying* CT skills (King & Kitchener, 2004).

Cumulatively, the models of RJ, CT and memory described above are integrated below into a cognitive framework for purposes of describing the focus of experimental work presented in this thesis (see Figure 1.10). Given that the cognitive and metacognitive processes under investigation in this thesis have been identified, discussed and structured into a cognitive system, it becomes important to also explore

some of the factors that may negatively impact these processes, such as cognitive load and the problematic nature of text-based reading. The following discussion of impediments to learning will pave the way for a closer analysis of argument mapping (see Chapter 3) and the ways in which argument mapping can potentially overcome such learning impediments and also facilitate memory, comprehension, CT and RJ ability (i.e. in empirical Chapters 4-6).

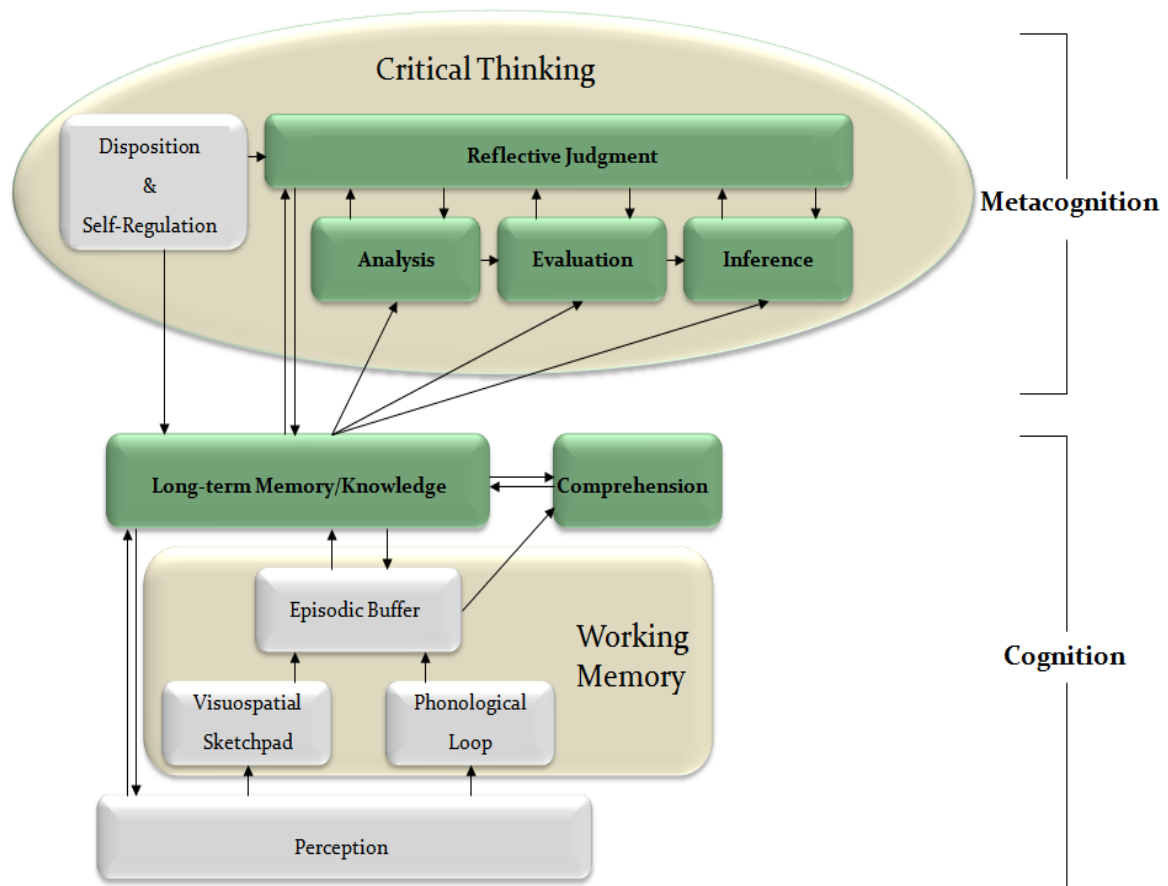


Figure 1.10: The proposed system of cognitive processes for educational objectives used in this research

1.4 Cognitive Load

Cognitive load refers to the cognitive demands put upon an individual in using and distributing working memory resources during cognitive activities such as learning and problem-solving (Sweller, 1999). Based on the review of working memory and taxonomies for thinking above, excess demands placed on working

memory resources during cognitive activities are speculated to impede performance of all higher-order thinking processes (see Figure 1.11), given that subsequent thinking processes are dependent upon one's ability to process and remember individual propositions and interdependent proposition sets (Halpern, 2003a; Maybery, Bain and Halford, 1986). For example, in support of the link between memory and deductive reasoning ability, Maybery, Bain and Halford (1986) found that students often have difficulty in inferring conclusions in deductive problem-solving situations problems because of the necessity to simultaneously recall and consider all elements in a premise or set of premises.

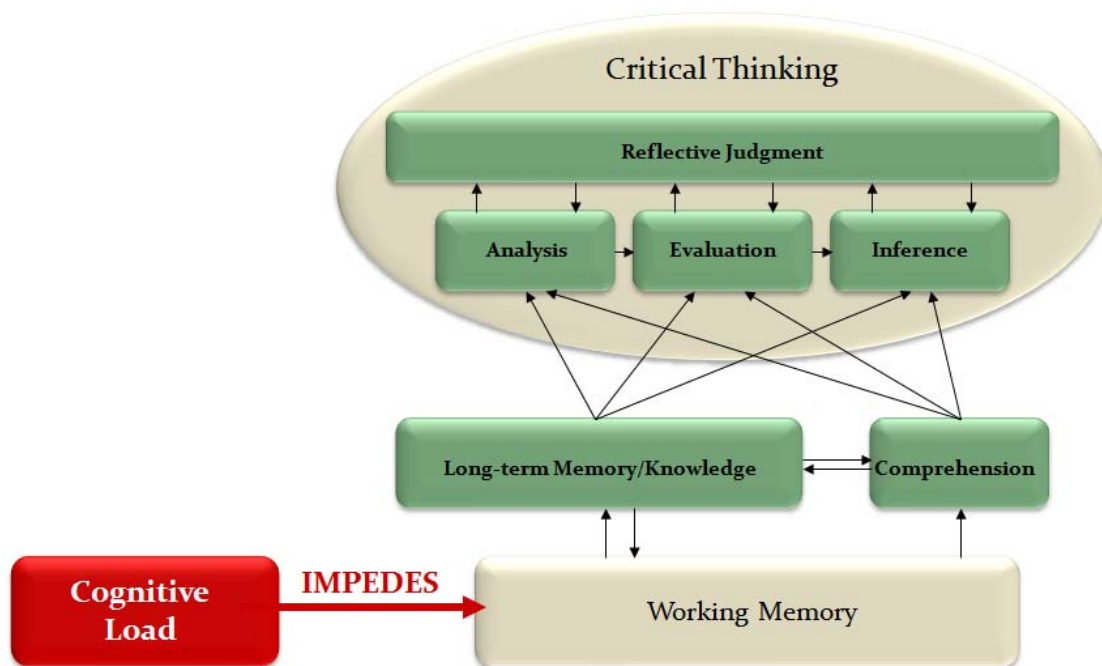


Figure 1.11: The effect of Cognitive Load on Thinking Processes

Cognitive load also refers in part to the amount of information that needs to be processed within an already limited capacity working memory (Baddeley, 2000; 2002; Cowan, 2000; Gobet & Clarkson, 2004; Miller, 1956; Pollock, Chandler & Sweller, 2002; Sweller, 1999). For example, George Miller (1956) postulated that

only seven (plus-or-minus two) items of information can be processed in short-term/working memory at any given time. More recent research disputes Miller's 'magical number' of seven items, or chunks⁸, that can be processed, claiming that the number is actually closer to four items of information (Cowan, 2000), or possibly less (Gobet & Clarkson, 2004). According to Sweller (1994), there are two types of cognitive load that constrict working memory capacity, impede learning and together determine total cognitive load - *intrinsic* and *extraneous* cognitive load.

1.4.1 Intrinsic Cognitive Load

Intrinsic cognitive load refers to the cognitive demands "imposed by the basic characteristics of information" (Sweller, 1994, p.6). The amount of intrinsic cognitive load placed on an individual during a cognitive task depends on the element interactivity, which refers to the number of elements, or items of information that must be simultaneously processed in working memory in order to learn and understand the information. "All of the learning elements interact and unless all are considered simultaneously in working memory, the problem and its solution will not be understood" (Sweller, 2010, p. 41). Simply, intrinsic cognitive load is caused by the necessity to simultaneously assimilate a relatively high number of elements, and their interactivity, in order to solve a problem.

For example, Pollock, Chandler and Sweller (2002) found that working memory capacity is often at its peak when trying to assimilate complex, or too much information (i.e. where there are many elements of information to assimilate and high levels of element interactivity). In their research, 22 first-year industrial trade students

⁸ According to Miller, this limited capacity can be increased by *chunking*, or organising items of information together into *chunks*, each of which possesses its own discrete meaning and acts as a new, higher-order (i.e. more complex) item of information. Since it is not necessarily five to nine discrete items of information that a person can recall, but five to nine *chunks*, chunking allows one to increase the amount of information that can be encoded, stored and retrieved in working memory.

were asked to study materials designed to facilitate learning of electrical tests. Two study formats were available to students. One group of students studied from an *interacting elements* format and the other group studied from what the authors called an *isolated-interacting elements* format. The interacting elements group were provided with a diagram of *'the insulation resistance test'* and *'the Earth continuity test'* along with a set of instructions that explained the theory behind the steps they were to follow. Those asked to learn in this condition had to simultaneously consider multiple elements, such as the aim of the test; the required setting of the voltmeter; the required setting of the appliance switch; the location of the earth lead and the line lead during both phases of the test; and the criteria by which to judge if the voltmeter readings were safe. Interacting elements are “characterised by the inclusion of all elements required for understanding but at the cost of an impossibly high working memory load” (Pollock, Chandler & Sweller, 2002, p. 66). Those in the isolated-interacting elements group were provided the same diagrams as the other group and only brief instructions of exactly what to do for each step. As these isolated elements required less integrated understanding, but rather sequential understanding and application of rules, they also placed less of a load on working memory.

Participants were asked to rate the mental effort needed to study and were then tested for their knowledge based on their study materials. Forty-eight hours later, both groups were again asked to study, but this time both groups received the interacting elements format. Results revealed that those who had initially studied isolated-interacting elements performed significantly better on subsequent high-element interactivity questions and better on average on practical tests, than those in the interacting elements group. Those in the isolated-interacting elements group also

found their learning condition to be significantly less demanding than those in the interacting elements group.

Furthermore, in a follow-up experiment by Pollock, Chandler and Sweller (2002), no differences in learning and self-reported difficulty of learning conditions were found between novices who studied via the isolated-interacting elements approach and experts who studied via the interacting elements approach. According to the authors, findings indicated that expertise can also reduce intrinsic cognitive load on working memory and allows for less stressful learning (i.e. in terms of perceived difficulty), due to the experts already possessing the schemas necessary to learn and succeed on the tests. Thus, the authors proposed that complex information cannot be simultaneously processed for purposes of understanding and memorisation unless the information has been previously organised into a schema. The authors further proposed that the promotion of schema construction would lead to an increase in the learner's understanding, which in turn would facilitate subsequent memorisation (Pollock, Chandler & Sweller, 2002, p. 82).

In summary, Pollock, Chandler and Sweller (2002) argue that a high level of element interactivity can cause cognitive load and that readers perform better when they learn by first processing information in an organised, sequential manner, possibly by integrating this new information into existing schemas or by building new schemas via sequential addition of elements; and then by re-reading the text, in order to commit the information to memory. Notably, an individual's level of expertise or prior knowledge of a subject area may influence the experience of intrinsic cognitive load (Sweller, 1999, 2010). That is, even if information possesses a high level of element interactivity, individuals with relevant, pre-existing knowledge may not

experience as much intrinsic cognitive load as those with less knowledge (Chi, Glaser & Rees, 1982; Kotovsky, Hayes and Simon, 1985; Larkin et al., 1980).

As such, although intrinsic cognitive load is fixed by virtue of the number of elements that need to be coordinated to solve a problem, if working memory has access to established schemas from LTM, then the demand associated with element interactivity in working memory may be decreased as a result of schema-driven chunking, thus reducing overall cognitive load. Though intrinsic load cannot be reduced by the format of instructional materials (i.e. regardless of format, a fixed body of information contains a fixed number of elements; Moreno & Park, 2010; Paas, Renkl & Sweller, 2003), research suggests that efforts to promote schema-construction through training can aid in the reduction of intrinsic load (Pollock, Chandler & Sweller, 2002; van Merriënboer, Kirschner & Kester, 2003). Those who are trained in a specific domain are provided the opportunity to develop expertise in that domain through the construction of relevant schemas during training (Chi, Glaser, Rees, 1982; Kotovsky, Hayes & Simon, 1985); and thus, are better equipped to assimilate information with high element interactivity than are those who do not possess the relevant pre-existing knowledge (Pollock, Chandler & Sweller, 2002; Sweller, 2010). Notably, a vast body of research (e.g. Gadzella, Ginther & Bryant, 1996; Hitchcock, 2004; Reed and Kromrey, 2001; Rimiene, 2002; Solon, 2007; discussed in Chapter 2) indicates that training in CT yields better CT performance than performance prior to training. These findings suggest that CT, which may be applied to problematic situations with high levels of element interactivity, can be trained; and subsequently, it is possible that schemas for CT strategies can reduce the cognitive load associated with high element interactivity by offering the critical thinker a set of cognitive, or metacognitive, strategies that allow them to navigate

problematic situations and infer reasonable solutions to problems. Though domain-specific knowledge training and the training of CT may facilitate schema construction - which may in turn aid in overcoming intrinsic cognitive load, there still remains the issue of overcoming extraneous cognitive load.

1.4.2 Extraneous Cognitive Load

Extraneous cognitive load refers to the cognitive demands imposed by instructional design (van Merriënboer & Ayres, 2005) and can be caused by a number of features of instructional materials that impose different demands on working memory. Sweller and colleagues have argued that extraneous cognitive load is caused by the demands placed on working memory associated with, for example, the need to switch attention during the assimilation of information and assimilating redundant information (Ayres & Sweller, 2005; Chandler & Sweller, 1991; Sweller et al., 1990; Sweller, 1999; Tindall-Ford, Chandler & Sweller, 1997). In the context of extraneous cognitive load, redundancy refers to the presence of information that does not contribute to schema acquisition or interferes with learning (e.g. irrelevant information and information presented more than once; Sweller, 1999; 2010). Attention-switching demands refer to the demands placed on working memory by switching between multiple sources of information that are unintelligible in isolation and results in less learning than if the multiple sources were presented in an integrated format (Sweller, 1999; 2010).

While intrinsic cognitive load is fixed by virtue of the number of elements one must coordinate to solve a problem or assimilate new knowledge, extraneous cognitive load can be reduced via the manipulation of instructional materials. For example, Chandler and Sweller (1991) conducted multiple experiments and found that when two sources of study information are integrated (e.g. a diagram and text),

students recall more information than when asked to study from separated texts and diagrams. In one of their experiments, Chandler and Sweller (1991) compared two groups of students who studied an electrical wiring installation procedure from either conventional study materials (i.e. separate diagram and text) or integrated study materials (i.e. text appropriately placed within the diagram). Those in the integrated group spent significantly less time processing the electrical wiring installation material than those in the conventional group. In addition, those in the integrated group scored significantly higher than the conventional group on subsequent tests of their knowledge for electrical wiring installation. These results suggested that switching attention from one source of information to another, while studying the conventional materials, was a source of cognitive load that impeded learning. Similar results were reported by Sweller et al. (1990) and Tindall-Ford, Chandler and Sweller (1997).

Like Chandler and Sweller (1991), Tindall-Ford, Chandler and Sweller (1997) also found that learning is impeded when instructional materials require a high degree of attention switching. Their research examined participants' ability to recall a list of instructions for an electrical test. Two groups were examined: a group who were given a diagram with the instructions integrated into the diagram and a group who were given the instructions separate from the diagram. The group who used integrated diagrams and instructions performed significantly better than the group with separate diagram and instructions on a test of transfer knowledge (i.e. the ability to apply the principles of electrical testing, which they studied, to other electrical systems). Tindall-Ford, Chandler and Sweller (1997) concluded, in conjunction with research by Sweller et al. (1990), and Chandler and Sweller (1991), that encoding environments that increase the cognitive load placed on the learner (e.g. via attention

switching) tend not only to slow the learning process, but also reduce overall levels of learning (Sweller et al., 1990; Sweller & Chandler, 1991; Tindall-Ford, Chandler & Sweller, 1997).

1.4.2.1 The Problematic Nature of Text-Based Reading

Interestingly, one of the most traditional methods of assimilating information - text-based reading, is often associated with a high degree of attention-switching demands. When reading text, attentive and motivated students may strive to remember as much information as they can. However, learning by reading and assimilating information from text can be problematic. For example, Harrell (2005) notes that students often fail to understand the 'gist' (Kintsch & van Dijk, 1978) of text-based information. More often, students fail to adequately 'follow' an argument's chain of reasoning or justification of claims within text. Harrell (2005) speculates that this is due to the fact that most students do not even acknowledge that information within a text presents an argument and instead read it as if it were a story. Conversely, authors who do understand the nature of argumentation often construct verbose, 'maze-like' arguments that consist of large amounts of text (Monk, 2001). Students who are presented with these texts may thus find it very difficult to capture anything more than the 'gist' of the argument. For example, text-based arguments often contain many more sentences than just the propositions that are part of the argument and these sentences may obscure the intention of the piece and the inferential structure of the argument (Harrell, 2004).

Some of the problems with text-based learning are derived from the linear structure of text itself, which does not allow one to readily connect statements that support or dispute specific propositions (i.e. other reasons and objections) and this can make it difficult for one to assimilate the information within a text-based argument

(van Gelder, 2003). The learner must link propositions that are often located in different paragraphs, different pages or different chapters. For example, when reading text, a person may read a statement on page four and not read any relevant support (or objection) to this claim until they reach page twelve. Between pages four and twelve, it could be that a variety of other propositions, tables or figures are presented, which places additional cognitive load on the reader. As a result, when reading a text-based argument, the reader must mentally construct the argument, which may require that they switch attention away from the information presented in the text. Based on previous research findings (e.g. Chandler & Sweller, 1991; Pollock, Chandler & Sweller, 2002; Sweller, 1999; Sweller et al., 1990; Tindall-Ford, Chandler & Sweller, 1997), presenting educational materials in a way that reduces the level of attention switching may minimize cognitive load and improve memory, comprehension and critical thinking performance. This suggestion implies that due to the linear nature of text, it may be important for textual information to be manipulated in a manner similar to the way in which Sweller and colleagues manipulated their educational materials (e.g. Chandler & Sweller, 1991, Pollock Chandler & Sweller, 2002; Tindall-Ford, Chandler and Sweller, 1997); specifically, by integrating text and diagrammatic representations in order to devise a study format that can potentially improve learning.

1.5 Summary

A number of frameworks have identified cognitive and metacognitive processes necessary for thinking in educational settings, such as memory, comprehension, analysis, evaluation, inference and reflective judgment. There are a number of factors that may negatively impact the operation of these cognitive and metacognitive processes in educational settings, namely intrinsic and extraneous cognitive load. Given that all cognitive activities impose at least some cognitive load

on the learner, it is often important for students to perform such activities in a manner which allows them to process information with as little load as possible. Of particular interest here is the potentially problematic nature of text-based reading in cases where the goal of the learner is to facilitate their memory, comprehension, analysis, evaluation, inference and reflective judgment ability by using text-based learning materials.

In the current research, it is hypothesised that argument mapping (i.e. an instructional/learning strategy) can potentially reduce extraneous cognitive load and enhance memory, comprehension and CT (i.e. analysis, evaluation, inference and RJ). This is because argument maps present information via the integration of textual and diagrammatic representations. Argument maps can also simplify the reading of an argument structure by reducing the potential demands associated with high levels of attention switching, specifically, by keeping related arguments together in the diagrammatic representation of the argument structure. Finally, argument maps facilitate the assimilation of core statements and relations that are central to an argument via the use of hierarchical organisation and visual-spatial cues (i.e. colour and proximity). However, before fully describing argument mapping and a detailed rationale for why it is hypothesised to enhance memory, comprehension and CT, it is important to first discuss past research which has examined alternative methods of enhancing these thinking processes. More specifically, strategies discussed above as potential methods of reducing extraneous cognitive load (e.g. well-organised and integrated educational materials) and intrinsic cognitive load (e.g. training) will be presented; as will their effects on memory, comprehension and CT performance. Thus, in the following chapter, a literature review of previous research on methods of enhancing memory, comprehension and CT is presented. This literature review will

help pave the way for the empirical focus of this thesis, which examines the impact of argument mapping on memory, comprehension and CT.

Chapter 2

Review of the Literature Examining Strategies Aimed at Improving Memory, Comprehension & Critical Thinking

In this chapter, a selective review of previous research conducted on strategies designed to improve memory, comprehension and critical thinking ability is presented. First, a review of research on the effects of organisational strategies on memory and comprehension of text-based information is presented. Second, a review of research on the effects of various training interventions on critical thinking performance is presented. The overall pattern of findings in the literature and some of the gaps in the existing literature help to situate the empirical aims of this thesis in the wider empirical literature and help us to further understand why the strategic use of argument mapping in an educational context may potentially enhance students' memory, comprehension and critical thinking ability.

2.1 Improving Memory and Comprehension Ability

Teachers and students can use a variety of different strategies to facilitate memory and comprehension of text-based information. Prominent amongst the strategies aimed at improving memory and comprehension ability are methods of organizing information for the purpose of easy assimilation (e.g. Berkowitz, 1986; Farrand, Hussain & Hennessy, 2002; Meyer, Brandt & Bluth, 1980; Taylor, 1982). For example, research suggests that when *to-be-remembered* information is presented in a well-organized manner, free recall is better than when information is presented in a random order (Bower et al., 1969; Chan, 2009; Myers, 1974). Myers (1974) conducted research in which three groups of participants were asked to read and memorise five paragraphs concerning five fictional countries. The sentences in the

five paragraphs were organised in a different manner for each group: (1) according to country (e.g. all sentences in paragraph 1 presented information on *Melin*, followed by a paragraph on *Pemol*, then on *Tupel*, *Gamba* and *Sayon* – all fictional countries), (2) according to specific attributes of all five countries (e.g. all sentences in paragraph 1 presented information on climate, followed by a paragraph on language, then on agricultural produce, industrial produce, and geography), and (3) randomly (e.g. sentences in relation to the specific attributes of different countries were randomly presented throughout the five paragraphs). Results suggested that the two groups who read from organised texts (i.e. those who read texts organised according to country or specific attributes) showed more accurate recall than those who read disorganised (random) texts.

Similarly, Armbruster, Anderson and Ostertag (1987) and Meyer, Brandt, and Bluth, (1980) both found that students who recognise the structure of text (i.e. the various components of an argument: central claims, supporting claims, objections) comprehend and recall more information than readers who do not recognise or differentiate these distinct structural components of an argument. Recognition and awareness of text structure aids in the identification of super-ordinate propositions, that is, the 'top-level' propositions, or main ideas/central claims within the text (Meyer, Brandt, & Bluth, 1980). The recognition of such top-level propositions is important because due to the limited capacity of working memory (Baddeley, 2000), not every piece of information presented can be memorised and thus, some pieces of information may become more important to encode than others (Kintsch & van Dijk, 1978). However, this is not to say that lower-level propositions are not as important as top-level propositions, considering that the former often provide the necessary justification for the latter. Rather, the importance of recognising top-level propositions

is that such recognition may act as a grouping strategy - an organisational cue that allows the reader to group lower-level propositions according to each top-level proposition.

For example, Meyer, Brandt and Bluth (1980) asked three groups of students (i.e. students with good, average or poor comprehension skills; as assessed by the Stanford Achievement Test comprehension scale and a reading achievement test) to read two text-based passages on *dehydration* and *super-tankers*. After reading each passage, students were asked to write down all they could remember from the passage in their own words as a recall assessment. A week later, students were again asked to write down all they could remember from the passage. Students who were previously shown to possess good comprehension skills tended to use the same type of top-level structure for organising their writing as the original author of the passage had used. Specifically, students with good comprehension skills organized their writing according to the same main ideas expressed by the authors in the original text. Conversely, students with poor comprehension skills were more disorganized in their writing style – they did not organise their writing by reference to the main ideas expressed by the authors in the original text. In addition, students who organized their writing by reference to the main ideas in the text recalled significantly more propositions for both passages of text, at both testing times, than those who did not.

According to Meyer, Brandt and Bluth (1980), the analysis of text structure, through the identification of super-ordinate propositions and the assimilation of the organisational structure of text, helps “determine what is important to remember” (p. 72). That is, recall can be aided through the organisation of target propositions during the reading, comprehension and analysis of those propositions. Furthermore, they speculate that, “additional processing of super-ordinate propositions *and* their

interrelationships increases the depth with which they are processed”, as the structure of text indicates the logical connections among these super-ordinate propositions, along with connections with sub-ordinate propositions (Craik & Lockhart, 1972; in Meyer, Brandt & Bluth, 1980, p.78).

Similarly, Yin-Kum (1995) found that recall ability is better when studying from structured text in comparison with unstructured text. In addition to facilitating ‘deep’ processing and focus on what is ‘important’ to remember, the strategy used by students of selecting top-level propositions around which to structure their recall behaviour may reflect an underlying capacity to organize information and use these organizational strategies to facilitate memory performance. Consistent with the ‘top-level’ strategy used by students in research by Meyer Brandt and Bluth (1980), explicit teaching and learning strategies such as *hierarchical summarisation* and *graphic organisation* have also been used to facilitate the organisation of text-based information according to the structure of the argument provided within the text.

2.1.1 Hierarchical Summarisation

Hierarchical summarisation of text is a more explicit, active and organisational strategy used to facilitate recall than simply summarising or learning the ‘gist’ (Kintsch & van Dijk, 1978) of a text. Hierarchical summarisation (sometimes described by educators as ‘creating an outline’) is a method of extracting and summarising both the super-ordinate and sub-ordinate themes from a text and then arranging these themes in a hierarchically organised list of super-ordinate and sub-ordinate propositions, respectively (see Figure 2.1 for an outline used in Taylor & Beach, 1984).

Outlines present the significant information from a text in a hierarchically organised manner and have been commonly used as a method for educators to present

notes and for students to take notes (Robinson & Kiewra, 1995). In addition, the hierarchical organisation of text-based information has been shown to facilitate recall and comprehension (Bower et al., 1969; Taylor, 1982; Taylor & Beach, 1984). For example, Taylor and colleagues (Taylor, 1982; Taylor & Beach, 1984) found that the use of hierarchical summarisation, or outlining, increased the recall of text in students who were trained in the use of the technique.

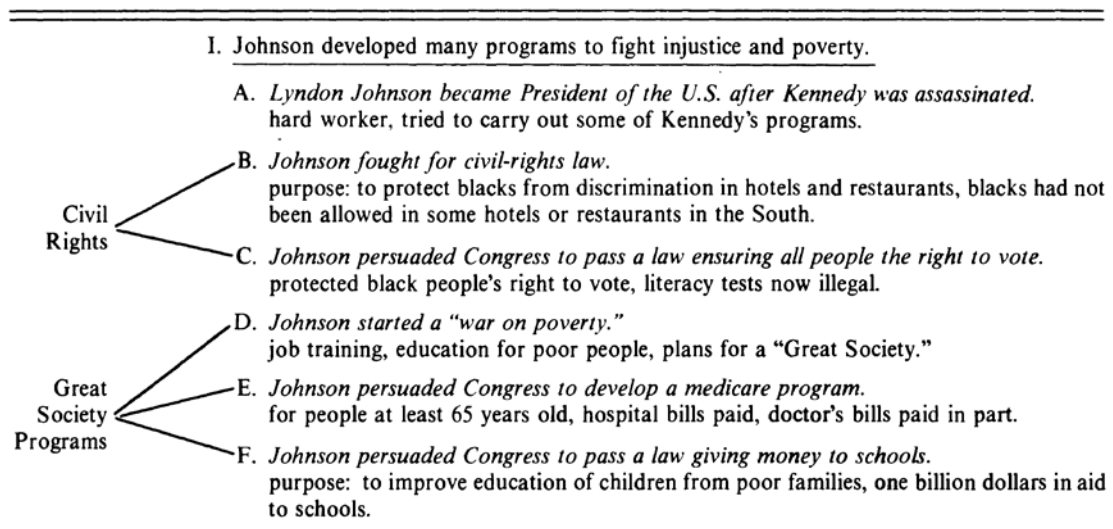


Figure 2.1: An example of an outline (hierarchical summary) as used in Taylor & Beach (1984, p. 139)

In research by Taylor (1982), 48 fifth-grade students were randomly assigned to one of two groups: an experimental outlining group or a conventional study group. Once a week for seven weeks, the students assembled in their assigned group for an hour and practiced reading and studying selected passages using either an outlining method or conventional methods (i.e. note-taking and question answering). During the testing phase, students were asked to read an 861 word piece of text. Those in the experimental group were asked to construct an outline of the information from the text and those in the conventional group were asked to answer questions regarding the text. Both groups were given 45 minutes to read the text and use their respective study

techniques, after which all study materials were collected. The next day students answered 20 short-answer questions based on the reading material, which probed students to recall a number of different propositions from the original text passage. Results revealed that those who outlined the text after reading it scored significantly higher on memory performance than those who answered questions after reading the text. According to Taylor (1982), these results indicate that the organisation of text through outlining enhances memory performance. Notably, however, it may not have been the organisational features of outlining alone that facilitated enhanced memory performance - it may also have been that outlining provided students with a more active learning strategy than simple question answering; and, active learning strategies have been shown to improve learning (e.g. Hake, 1998; Laws, Sokoloff & Thornton, 1999; Redish, Saul & Steinberg, 1997).

This experiment was replicated by Taylor (1982), using a different text, which contained 819 words. However, in this second experiment, no differences between hierarchical summarisation and conventional reading groups were observed. Taylor found that, after inspecting the completed outlines, students who outlined in Experiment 1 scored 60% on accuracy on an assessment of logical organisation of propositions, whereas students in Experiment 2 only scored 43% correct on a similar test, which suggests that students who participated in the first experiment produced better, more accurate outlines than those in the second experiment. From this, Taylor concluded that recall can only be improved by outlining information from text when the outline is hierarchically summarised *correctly* or appropriately (i.e. better organised outlines result in higher recall). Taylor's findings imply that memory performance may be contingent upon 'comprehension' during the study phase. That is, students who do not fully comprehend the text cannot construct a good (i.e.

accurate) outline. Furthermore, Taylor and Beach (1984) found significant variations in performance as a function of the topic studied, in students who were using hierarchical summarisation as a study method. More specifically, when students were more familiar with a certain topic (i.e. as assessed via a self-reported 5-point likert scale), those who studied using hierarchical summarisation performed no better than those in the conventional study condition. Nevertheless, the results suggest that, under certain circumstances, hierarchical summarisation is a learning method that may improve recall.

2.1.2 Graphic Organisers

Similar to outlines, graphic organisers hierarchically structure information, yet do so in a visual, graphical manner so that relationships among propositions are made explicit. Graphic organisers can also allow for flexible manipulation of the location of concepts or propositions within the structure and can be arranged in multiple configurations, unlike both text and outlines which are generally structured in a linear format. Furthermore, graphic organisers (e.g. pros-and-cons lists, cause-and-effect lists, pie-charts, flow-charts, spider-diagrams, tree-diagrams, Venn-diagrams and various mapping strategies) are often recommended for those who struggle to derive meaning from text (Oliver, 2009; Zmach et al., 2007).

Oliver (2009) argues that graphic organisers facilitate comprehension because they provide visual scaffolds that encourage students to extract and represent key details in text (e.g. reasons and objections for specific claims) and provide relationships between linked propositions. Oliver recommends the use of graphic organizers because they can employ the principles of hierarchical organisation (Taylor, 1982) and can provide visual diagrammatic representations that facilitate the

extraction of main ideas, relevant information (Larkin & Simon, 1987) and structural details within text (Meyer, Brandt & Bluth, 1980).

In two experiments by Robinson and Kiewra (1995), students' ability to comprehend material presented to them from text only, text and outlines, or text and graphic organisers was examined. The method of graphic organisation used in both experiments is presented in Figure 2.2. In Robinson and Kiewra's first experiment, students in the graphic organiser group (i.e. those who studied from both graphic organiser and text) and the outline group (i.e. those who studied from both outline and text) were given 45 minutes to study their materials. After 45 minutes of text study, the text was collected and the students were provided with another 15 minutes to study either the graphic organiser or the outline (i.e. depending on their group). Students in the text-only group were given one full hour to study their text. Students were examined on hierarchical relations (i.e. as assessed via cued-recall questions which required knowledge of the text structure), on facts (i.e. as assessed using multiple choice questions), and on conceptual relationships (i.e. comprehension was assessed through the administration of an essay test) either immediately after studying or 2 days later.

The researchers found that those who were tested immediately after studying performed better than those in the delayed testing group on all outcome measures; and that those who studied from text and graphic organisers comprehended more conceptual relationships (as assessed by essay testing) than those in the outlining or text-only groups. In their second experiment, the same procedure was followed, except all students were tested a day after studying. On the day of testing, students were given an additional 15 minutes to study prior to assessment. Results from this experiment revealed that those who studied from text and graphic organisers learned

more hierarchical relations and conceptual relationships than those studying from either outlines and text or text alone.

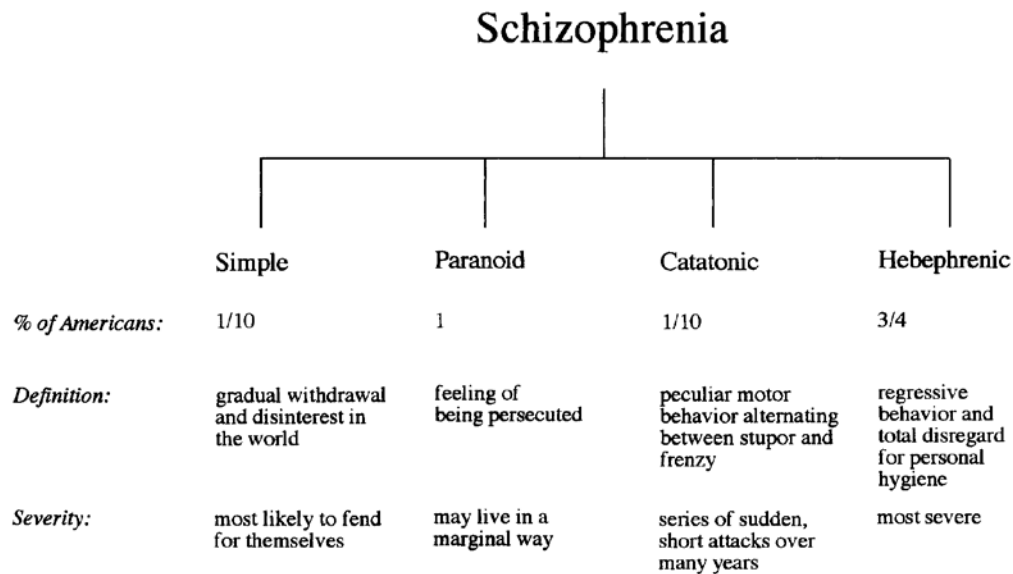


Figure 2.2: An example of a graphic organiser as used in Robinson & Kiewra (1995)

2.1.2.1 Mapping Strategies

Mapping strategies are one type of graphic organisation technique designed to help students map the associations, relationships, or inferential connections between different concepts or propositions and contentions in a text (Davies, 2010). Various research studies examining these mapping techniques (e.g. mind-mapping and concept mapping) have revealed beneficial effects on comprehension and recall performance (Berkowitz, 1986; Farrand, Hussain & Hennessy, 2002; Oliver, 2009). For example, Farrand, Hussain and Hennessy (2002) found that the use of *mind-maps* (Buzan & Buzan, 1997) can improve memory. A mind-map is a mapping strategy in which the major concepts related to a topic extend from a central node with a series of smaller branches representing sub-concepts extending from larger concept branches (Budd, 2004). In the mind-map, concepts are presented as a single word or phrase, a collection of which are ultimately used to illustrate the associations amongst the subject area that is mapped (a detailed account of the features of mind-mapping is

provided in Chapter 3). In the research conducted by Farrand, Hussain and Hennessy (2002), fifty under-graduate medical students were assigned to either a mind-map condition or a self-selected study condition. Students were asked to read an article and study it using either the mind-mapping technique or a self-selected study technique. Results revealed that those in the mind-map group scored significantly higher on both the immediate and delayed recall tests.

Berkowitz (1986) devised a mapping technique in which the main ideas from texts were summarised in separate boxes and supporting claims were listed beneath each of the main ideas (see Figure 2.3). The boxes were organized in a radial structure (i.e. around a *main idea*). Berkowitz compared variations of this technique (map construction; map study) with alternative techniques (question answering; text re-reading) as methods for learning information as assessed using a post-study memory test. The map-construction group was provided with training in how to construct maps and also how to study them. Those in the map-study group received prepared maps and were trained in how to study them. In the question-answering group, students were instructed to read text and study it using a set of questions. In the re-reading group, students were instructed to read the text twice. A free recall test and short-answer test were then administered both immediately after study and again two weeks later. Results revealed that the map-construction group scored higher on average than those in other groups on both immediate and delayed memory tests.

In summary, research has shown that when information is presented in a well-organized manner, recall is better for that information than from less organised information (e.g. Myers, 1974). Similarly, students who demonstrate awareness of the structure of text comprehend and recall more information than readers who do not (Armbruster, Anderson & Ostertag, 1987; Meyer, Brandt, & Bluth, 1980). Research

has also shown that the active organising of text-based information into a hierarchy can facilitate improved recall and comprehension (Oliver, 2009; Robinson & Kiewra, 1995). Research studies using various different methods of graphic organisation, such as mapping techniques, have also revealed beneficial effects on comprehension and recall performance (Berkowitz, 1986; Farrand, Hussain & Hennessy, 2002; Oliver, 2009). Like hierarchical organisation, graphic organisers structure text-based information hierarchically, but do so via the provision of visual scaffolds. Argument mapping, as discussed in the next chapter, is another form of graphic organisation which is hypothesised in the current research as an optimal method of organising information and decreasing cognitive load.

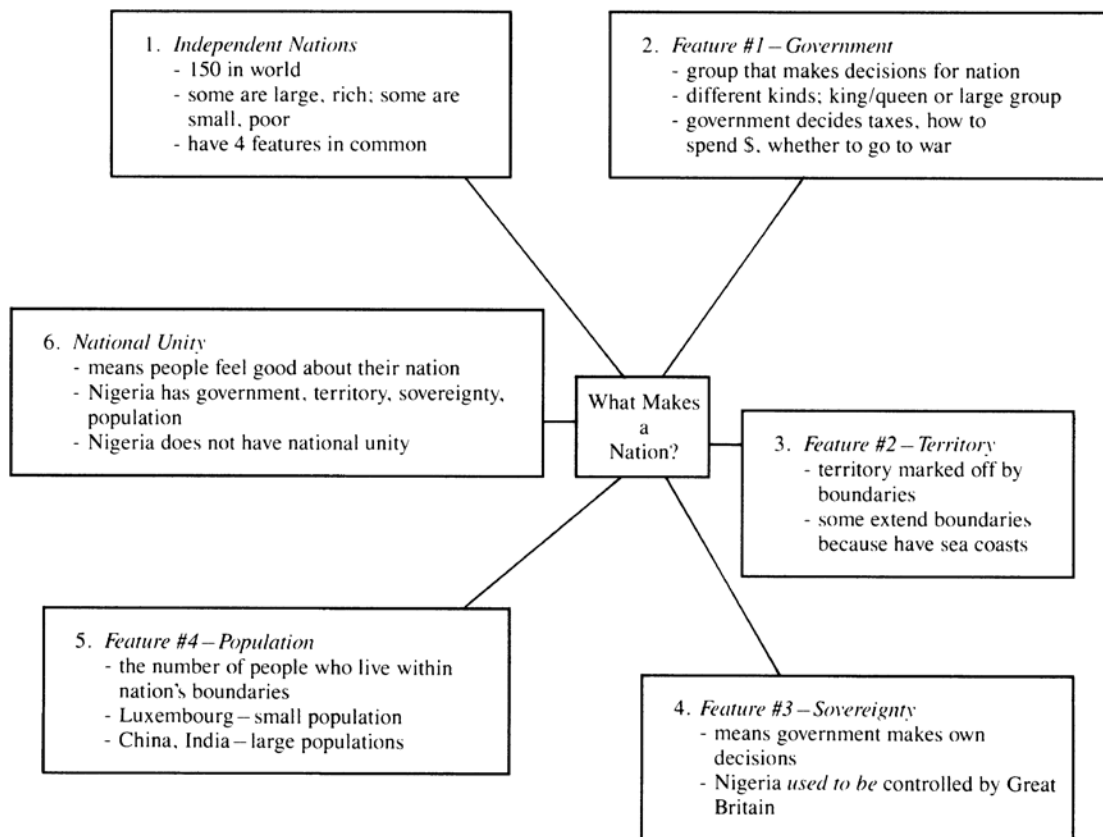


Figure 2.3: An example of a mapping strategy, as used in Berkowitz (1986, p. 165)

However, while organisational strategies, such as outlining and graphical organisers have largely been used to support learning objectives at the bottom of Bloom's taxonomy (i.e. knowledge/memory and comprehension), the majority of critical thinking (CT) training research studies have not focused exclusively on training in the use of organisational strategies as a means of accelerating CT skill development. Given that CT takes time to develop (Halpern, 2003a; King, Wood & Mines, 1990), much of the research on CT has examined the effects of various CT training courses on CT performance, some of which have used argument mapping in the process of training (Alvarez-Ortiz, 2007).

2.2 Improving Critical Thinking Performance

Critical thinking (CT) courses have been taught at University in varying academic domains including law, philosophy, psychology, sociology and nursing, all with the goal of improving CT performance. Such CT courses have also been informed by varying conceptualisations of CT (e.g. Ennis, 1987; Facione, 1990b; Halpern, 2003a; Paul, 1993). These varying conceptualisations can make it difficult for researchers and teachers to understand or agree on the key components of good CT. These difficulties may impede the ability of researchers and teachers to construct an integrated theoretical account of not only how best to *train* CT skills⁹ but also how best to *measure* CT skills. As a result, researchers and educators must consider the wide array of CT measures available, in order to identify the best and the most appropriate measures, based on the CT conceptualisation used for training. Popular measures of CT performance include the California Critical Thinking Skills Test (CCTST; Facione, 1990a), the Cornell Critical Thinking Test (CCTT; Ennis, Millman

⁹ In the absence of greater clarity in relation to the components of CT skills and the way these components work together in the context of solving critical thinking problems, it can be difficult to design critical thinking training programs.

& Tomko, 1985), the Watson-Glaser Critical Thinking Assessment (WGCTA; Watson & Glaser, 1980) and the Ennis-Weir Critical Thinking Essay Test (EWCTET; Ennis & Weir, 1985).

It has been noted by some commentators that these different measures of CT ability may not be directly comparable (Abrami et al., 2008). For example, the CCTST consists of 34 multiple-choice questions (MCQs) and measures CT according to the same core skills as identified by the Delphi Report (Facione, 1990b): analysis, evaluation and inference, as well as inductive and deductive reasoning. The CCTT consists of 52 MCQs which measure skills of critical thinking associated with induction; deduction; observation and credibility; definition and assumption identification; and meaning and fallacies. The WGCTA consists of 80 MCQs that measure the ability to draw inferences; recognise assumptions; evaluate arguments; use logical interpretation and deductive reasoning (Watson & Glaser, 1980). The EWCTET is an essay-based assessment of the test-taker's ability to analyse, evaluate and respond to arguments and debates in real-world situations (Ennis & Weir, 1985; see Ku, 2009 for a discussion). The authors of the EWCET provide what they call a "rough, somewhat overlapping list of areas of critical thinking competence"¹⁰ which is measured by their test (Ennis & Weir, 1985, p. 1). This list can be seen in Table 2.1.

Furthermore, the reported reliability and validity of different CT measures also varies, which has led Abrami and colleagues (2008, p. 1104) to ask: "How will we know if one intervention is more beneficial than another if we are uncertain about the validity and reliability of the outcome measures?" Abrami and colleagues add that, even when researchers explicitly declare that they are assessing CT, there still remains the major challenge of ensuring that measured outcomes are related, in some

¹⁰ It is important to note, that the EWCTET has been criticised for its domain-specific nature (Taube, 1997), the subjective nature in which the essays are scored and potential biases in favour of test-takers who are more proficient in writing (Adams, Whitlow, Stover & Johnson, 1996).

meaningful way, to the conceptualisation and operational definition of CT that informed the teaching practice. Often, the relationship between the concepts of CT that are taught and those that are assessed is unclear; and a large majority of studies in this area include no theory to help elucidate these relationships.

Table 2.1: Topics Assessed by the Ennis-Weir Critical Thinking Essay Test (adapted from Ennis & Weir, 1985).

Getting the point
Seeing the reasons and assumptions
Stating one's point
Offering good reasons
Seeing other possibilities (Including other possible explanations)
Responding appropriately to and/or avoiding:
Equivocation
Irrelevance
Circularity
Reversal of an <i>if-then</i> (or other conditional) relationship
The Straw Person Fallacy
Overgeneralization
Excessive scepticism
Credibility Problems
The use of emotive language to persuade

Nevertheless, researchers have attempted to group CT intervention studies in an effort to examine whether or not CT can be improved via explicit instruction and how it is best improved. For example, a recent meta-analysis by Alvarez-Ortiz (2007) examined 52 studies which investigated a wide range of teaching strategies designed to improve CT. The meta-analysis was specifically conducted in order to answer the question as to whether or not participation in philosophy courses improved CT ability. The criteria for inclusion in the meta-analysis were that studies must have had: an intervention designed to enhance CT; a sample consisting of undergraduate students;

CT as a dependent variable, measured by an objective MCQ test; conducted quantitative analysis (i.e. including data for an overall effect size, means, standard deviations and sample sizes); and a pre-post-test design. Notably, of the 52 studies, only 29 used an intervention that explicitly taught at least some CT and only 12 of those used dedicated CT instruction (Alvarez-Ortiz, 2007). The remaining studies used interventions that taught other subjects, including philosophy, nursing, classics and history, psychology, politics and sociology and mathematics; but did not teach CT. Results of the meta-analysis revealed that participation in courses that taught at least some CT (effect size of .30 SD, CI [.16, .43]) yielded better CT performance than courses that did not explicitly teach CT in some form (effect size of .12 SD, CI [.08, .17]). Results from this meta-analysis also suggested that philosophy courses yielded a mean effect size of .26 SD, CI [.12, .39], with little evidence (i.e. $p > .05$) to suggest that participation in a philosophy course had any greater effect on CT performance than any other academic course (.16 SD, CI [.11, .21]).

In another recent meta-analysis of CT interventions, conducted by Abrami et al. (2008), 3,720 studies that focused on CT were identified and of that, 117 were deemed suitable for inclusion. Criteria for inclusion were: “(1) accessibility - the study must be publicly available or archived; (2) relevancy - the study addresses the issue of CT development, improvement, and/or active use; (3) presence of intervention - the study presents some kind of instructional intervention; (4) comparison - the study compares outcomes that resulted from different types or levels of treatment (e.g. control group and experimental group or pre-test and post-test); (5) quantitative data sufficiency - measures of relevant dependent variables are reported in a way that enables effect size extraction or estimation; (6) duration - the treatment in total lasted at least 3 hr; and (7) age - participants were no younger than 6 years

old” (Abrami et al., 2008, p. 1108). The 117 studies yielded 161 independent effect sizes, from which Abrami and colleagues (2008) reported a significant effect ($g+ = .34$) of all CT courses included in the meta-analysis on CT performance. However, of these 161 effects, only 91 were measured using standardised CT assessments (which yielded an average effect size of $g+ = .24$), only 101 were from true ($g+ = .34$) or quasi-experiments ($g+ = .36$) and only 16 were derived from the effects of instructional training ($g+ = 1.00$).

Abrami and colleagues used Ennis’ (1989) typology of four CT courses (i.e. general, infusion, immersion and mixed) to differentiate CT training methods. In the *general* approach to CT training, actual CT skills and dispositions “are learning objectives, without specific subject matter content” (Abrami et al., 2008, p. 1105). The *infusion* of CT into a course requires specific subject matter content upon which CT skills are practiced. In the infusion approach, the objective of teaching CT within the course content is made explicit. In the *immersion* approach, like the infusion approach, specific course content upon which critical thinking skills are practiced is required. However, CT objectives in the immersed approach are *not* made explicit. Finally, in the *mixed* approach, critical thinking is taught independently of the specific subject matter content of the course.

Comparing the four CT course types, results of the meta-analysis revealed that courses using the mixed approach had the largest effect on CT performance ($g+ = .94$), followed by the infusion approach ($g+ = .54$), the general approach ($g+ = .38$) and the immersion approach ($g+ = .09$), respectively. It is important to note that the immersion approach (which had the smallest effect) is the only approach that does not make CT objectives explicit to students. This finding suggests that making CT objectives and requirements clear to students may be an important part of any course

design aimed at increasing CT ability (Abrami et al., 2008). Furthermore, Abrami and colleagues concluded that the enhancement of CT ability is greatly dependent upon how CT is taught and more specifically, that the *mixed* and *infusion* approaches to teaching CT worked best as students were explicitly taught how to use and apply CT skills to other course content.

Though these meta-analyses have provided many interesting results, the issue of the inclusion of non-controlled studies should be considered. For example, though both meta-analyses discussed above included the criterion that studies incorporated into the meta-analysis must have compared outcomes that resulted from different types or levels of treatment (e.g. control group/experimental group design *or* pre-to-post-testing), this criterion did not ensure that all studies contained a control group (i.e. many studies used a pre-to-post-testing design). As a result, the effect sizes reported in both meta-analyses are derived from a mixed sample of controlled and non-controlled studies and should therefore be interpreted with caution. Notably, the inclusion of a control group or a similar experimental group with which to compare a CT intervention is necessary in order to confidently conclude that gains in CT observed are a result of the CT intervention and not simply the result of practice effects or maturation effects. For example, previous research has shown that CT skills can increase over time, without the aid of a CT intervention, as a result of maturation or experience (Pascarella & Terenzini, 1991).

Following an in-depth review of the available literature on CT interventions, five research studies have been selected for more detailed discussion below (i.e. Gadzella, Ginther & Bryant 1996; Hitchcock, 2004; Solon, 2007; Reed and Kromrey, 2001; Rimiene, 2002), because they all (1) examined CT performance as a dependent variable; (2) provided participants with explicit instruction in CT; (3) assessed a

sample consisting of undergraduate students; (4) conducted quantitative analysis of the results (i.e. including data for an overall effect size, means, standard deviations and sample sizes); and importantly, (5) to some degree described the conceptualisation of CT used as the basis to instruct and train CT skills within the intervention. The last criterion for inclusion in this selective review is particularly important to consider because, due to the many varying conceptualisations of CT available, there is no definitive method of how best to teach and measure CT. Having said that, only a handful of studies that examined the effects of CT instruction/training on CT performance have described in any detail the conceptualisation of CT used as the basis to instruct and train CT skills. Quite often, the detail that is provided in this context is very limited. In addition, it is important to note that though all five studies reviewed below also compared CT performance outcomes that resulted from different types or levels of treatment (e.g. pre-and-post-test design); this does not mean that they all included a control group, which is a common problem in CT intervention research. Although the importance of including a control group was discussed above and will continue to be an important criterion of experimental rigour in CT intervention research, the main purpose of this selective review is to demonstrate the variety of different conceptualizations of CT used in past CT training research and to critically examine the issues of alignment between conceptualization, training and assessment of CT.

Gadzella, Ginther and Bryant (1996) investigated the effects of a 14-week long CT course on the CT performances of 113 first year university students. Students in the intervention were assessed prior to commencement of the CT course via the administration of the Watson-Glaser Critical Thinking Appraisal, Form A (WGCTA; Watson & Glaser, 1980). During the course, students “were taught CT skills and

given problems (and brought some of their own) which were analysed by the whole class, small groups and individual students” (Gadzella, Ginther & Bryant, 1996, p. 5). According to the Gadzella and colleagues, the main focus of CT instruction was the analysis and solving of problems. Though the authors do not describe their conceptualization of CT in great detail, they do provide two definitions of CT - those by Chaffee (1988) and Dressel & Mayhew (1954), which they claim were the basis for the conceptualisation of CT taught to students. Chaffee (1988, p. 29) defined CT as “active, purposeful and organised efforts to make sense of our world by carefully examining our thinking and the thinking of others, in order to clarify and improve our understanding.” Dressel and Mayhew (1954) defined CT as “exclusively linked with abilities that are needed for solving problems, selecting pertinent information for problem-solving, recognising assumptions, formulating hypotheses, drawing valid conclusions and judging the validity of inferences” (Gadzella, Ginther & Bryant, 1996, p. 3).

As addressed by Gadzella, Ginther and Bryant (1996), as well as Watson and Glaser (1980; 1994), the WGCTA was developed in light of both research conducted by Glaser (1941) and Dressel and Mayhew’s definition of CT. This influence is evident as there is a great deal of congruence between Dressel and Mayhew’s definition and what the WGCTA measures (e.g. they both highlight the skills of recognising assumptions, evaluation and inference). After completion of the course, students completed the WGCTA, Form B. Results revealed a significant gain in CT ability from the pre-test assessment ($M = 47.69$, $SD = 10.03$) to post-test assessment ($M = 51$, $SD = 7.88$) with an effect size of $d = .37$. Notably, gains in CT were not compared with those of a control group.

Research by Hitchcock (2004) examined the effects of a computer-assisted CT course on CT performance (i.e. computer-assisted instruction was provided to students as a series of tutorials). Four-hundred and two undergraduate students participated in the study, of which 278 completed both pre-and-post-testing. Students in the intervention were assessed prior to commencement of the CT course using either Form A or Form B of the California Critical Thinking Skills Test (CCTST).

After completion of the pre-test, students engaged in 15.8 hours of computer-assisted CT training over the duration of 13 weeks. The components of CT trained in this study were based on LeBlanc's (1998) book, *Thinking Clearly* (see Table 2.2). There were two class-based MCQ assignments on the nature of argumentation, which were completed by the students in groups. Students also completed exercises derived from LeBlanc's (1998) *Thinking Clearly*, as well as MCQs and exercises from the computer-software *LEMUR*, which was used as a means of presenting computer-assisted CT instruction. There was also a course web-site available to students, in which answers to the text-book exercises were posted, as were additional MCQ exercises and past exams with answers. However, the extent to which students used the software or web-site was not monitored. As this was a compulsory course, students were also assessed via mid-term and final examinations; in which the former covered up to and including *accepting premises*, and the latter covered all topics, as listed in Table 2.2. Although the skills identified in Hitchcock's conceptualisation and those measured by the CCTST are not entirely congruent, some of the components of CT identified by Hitchcock (e.g. identifying arguments, accepting premises, accepting relevance and arguments from experience) are measured by the CCTST as sub-components of the skills of analysis and evaluation (see Chapter 1).

Table 2.2: Critical Thinking skills taught to students in Hitchcock (2004).

1. Identifying arguments
 2. Standardizing arguments
 3. Necessary and sufficient conditions
 4. Language (definitions and fallacies of language)
 5. Accepting premises
 6. Accepting relevance
 7. Arguments from analogy
 8. Arguments from experience
 9. Causal arguments
-

After completion of the course, students completed the CCTST post-test (i.e. either Form A or B - the opposite of what was completed at pre-testing). Analysis of change over time revealed that the students showed a significant gain in critical thinking performance of 6.44% percentage points on the CCTST - an effect size of .49 SD. However, it is important to note that much like Gadzella, Ginther and Bryant (1996), a control group was not included for comparison purposes in Hitchcock's study.

Solon (2007) examined the effects of infusing CT into an introductory psychology course (i.e. teaching an introductory psychology course by means of having students critically think about the topics encountered). Fifty-one students participated. Twenty-six students were allocated to a control group condition (i.e. students attended an introductory psychology course without the infusion of CT). Twenty-five students were allocated to the experimental group, in which over the duration of the course, students were required to think critically about a variety of psychology topics. Solon's list of CT skills applied to each topic is presented in Table 2.3. Students in the experimental group also completed 10 reading and writing homework assignments, derived from Chapters 4-7 of Diane Halpern's (2003a)

Thought & Knowledge: An Introduction to Critical Thinking; and Part II of Meltzoff's (1998) *Critical Thinking about Research: Psychology and Related Fields*. More specifically, these homework assignments involved deductive reasoning, argument analysis, thinking as hypothesis testing and understanding probabilities, likelihood and uncertainty (Halpern, 2003a). Ten hours of in-class activity were devoted to a review of Halpern, Meltzoff and related exercises.

Both the experimental and control groups were administered the Cornell Critical Thinking Test (CCTT Form Z) as pre-and-post-tests. Solon acknowledged that the issue of compatibility between what is taught as part of CT training and what is measured is important to consider, for purposes of clarifying whether or not it is the training of certain skills that is responsible for their improvement. Solon also noted that the conceptualisation of CT skills taught to students in his study and the manner in which CT skills were assessed was compatible, as the CT abilities laid out in this conceptualisation of CT were “observable, measurable and readily lend themselves to objective standardised testing, as in the Cornell” critical thinking test (Solon, 2007; p. 96). The level of compatibility between what was taught and what was measured can be seen through comparison of Table 2.3 (i.e. what was taught in the study) and what the CCTT measures (i.e. induction; deduction; observation and credibility; definition and assumption identification; and meaning and fallacies).

Notably, the results of Solon's research revealed that those who participated in the CT-infused psychology course performed significantly better on CT ability than the control group on post-testing with an effect size of .66 *SD*, CI [.08, 1.24]. Results also revealed that students in the experimental group showed a significant gain in CT ability with an effect size of .87 *SD*, CI [.28, 1.47]. As the control group did not demonstrate a significant gain in CT ability ($p = .49$), the results suggest that the gain

observed in CT ability in the experimental group was not a result of a practice effect (i.e. the same form of the Cornell Critical Thinking Test was used at both pre-and-post-testing). It is worth noting that the research conducted by Solon (2007) was one of the richest and most informative CT intervention studies reviewed, not only because it was a controlled study, but also because a more detailed account, relative to the other studies, was provided for how the CT conceptualization used in the intervention aligned with both the training and assessment.

Table 2.3: Solon's Conceptualisation of Critical Thinking

Critical thinking as a set of basic and generic reasoning skills, including the ability to identify and/or distinguish between:

1. Inferences and non-inferences
 2. Assumptions (covert as well as overt) and conclusions
 3. Consistent and inconsistent statement sets
 4. Deductive and inductive reasoning
 5. Valid and invalid arguments
 6. Credible versus seriously questionable claims and courses
 7. Meaningful versus vague ambiguous, and/or meaningless language
 8. Relevant versus irrelevant evidence
 9. Scientific versus pseudo-scientific procedures
-

Reed and Kromrey (2001) also reported improvements in CT ability, as a result of infusing CT into a U.S. history course. Students met for three hours per week for 15 weeks. Following pre-testing on the Ennis-Weir Critical Thinking Essay Test (EWCTET; Ennis & Weir, 1985) and the California Critical Thinking Dispositions Inventory (CCTDI; Facione & Facione, 1992), 29 students in the experimental group were: (a) explicitly taught Richard Paul's (1993) model of CT (i.e. *Elements of Reasoning*, which focuses on elements of good, quality thinking, reasoning and disposition; see Table 2.4); and (b) trained to use Paul's elements of reasoning to analyse primary source documents and historical problems.

Table 2.4: Paul's Elements of Reasoning (adapted from Paul, 1993)

Elements of Reasoning	Description
Purpose of Thinking	<p>Take time to state purpose clearly.</p> <p>Distinguish purpose from other related purposes.</p> <p>Choose significant and realistic purposes.</p>
Question at Issue	<p>Take time to clearly state the question at issue and express it in several ways to clarify its meaning and scope.</p> <p>Break the question into sub- questions.</p> <p>Identify if the question has one correct answer or requires reasoning from more than one point of view.</p>
Information	<p>Gather sufficient information and restrict claims to those supported by the data readily available.</p> <p>Search for information that both refutes and supports the information.</p> <p>Information used should be clear, accurate, and relevant to the question at issue.</p>
Interpretation & Inference	<p>Infer only what the evidence implies.</p> <p>Check inferences for their consistency with each other.</p> <p>Identify assumptions that lead to the conclusions made.</p>
Concepts	<p>Identify key concepts and provide clear explanations; and consider alternative concepts or alternative definitions to concepts.</p>
Assumptions	<p>Clearly identify assumptions and determine whether they are justifiable and consider how assumptions shape the bias/ balance of the reasoning, or point of view.</p>
Implications & Consequences	<p>Trace and consider all the implications and consequences that follow from the reasoning.</p>
Points of View	<p>Identify the point of view; seek others and identify their strengths as well as weaknesses.</p> <p>Strive to be fair-minded in evaluating all points of view.</p>

Students completed homework assignments that required the use of Paul's model and were provided with a packet of handouts that graphically displayed and further explained the model. Finally, students participated in classroom discussions which focused on the elements and standards set forth within Paul's (1993) model. Twenty-three participants in the control group completed the same assignments as part of their history course, but did not receive any CT instruction. Upon completion of the course, both groups again completed the EWCET and the CCTDI.

Results revealed no differences between the CT and control groups on the EWCET or the CCTDI at pre-testing, suggesting that both groups were adequately matched prior to the intervention. Results from post-testing revealed that those who received the CT training scored significantly higher than the control group on the Ennis-Weir test with an effect size of $f = .83$; and that there was no difference between the groups on CCTDI scores. The null-finding on the CCTDI suggests that the improvement in CT ability made by the CT training group was not a result of change in students' disposition toward thinking.

Rimiene (2002) investigated the effects of a 3-month long CT course on the CT performances of 77 university students, in comparison with a control group of 150 students who did not attend the CT course. Students were randomly allocated to their respective groups. All students were assessed prior to commencement of the CT course on both the CCTST and the CCTDI. There were no differences between groups on pre-test performance, suggesting that both groups were appropriately matched. The CT course "was based on the precepts of humanistic psychology and meaningful learning, and the aims were to introduce students to CT theory, to develop CT skills and to strengthen motivation for CT" (Rimiene, 2002, p. 18). Both Facione, Facione and Giancarlo (1997) and the Delphi Report (Facione, 1990b) were cited in

Rimieni's conceptualisation of skills necessary for CT - both of which name analysis, evaluation and inference as core CT skills. During the course, "students learned the principles and stages of CT, the main criteria of successful thinking and solved different problems" through various forms of active learning including "brainstorming, problem-solving, reflexive writing, active listening, purposeful research, co-operative learning, conversations, discussions, debates, projects and demonstration" (Rimieni, 2002, pp. 18-19). After completion of the course, all students were again assessed on both the CCTST and the CCTDI.

Results revealed that though there was no difference between groups at pre-testing, students in the CT group scored significantly higher than those in the control group on overall CT ($d = 1.57$), analysis ($d = .85$), evaluation ($d = 1.13$), inference ($d = 1.30$), inductive reasoning ($d = 1.10$), deductive reasoning ($d = 1.30$) and overall disposition towards thinking ($d = .62$) at post-testing. Results also revealed that students in the CT group scored significantly higher on overall CT ($d = 1.09$), analysis ($d = .37$), evaluation ($d = .82$), inference ($d = 1.10$), inductive reasoning ($d = .64$), deductive reasoning ($d = 1.17$) and overall disposition towards thinking ($d = .49$) at post-testing when compared with pre-testing. There were no differences between pre-and-post-test scores for the control group.

The findings from this study are interesting to consider given that both CT ability and disposition of those who participated in the CT course increased from pre-to-post-testing. Though CT ability may have developed over time as a result of instruction, it is also possible that improved dispositions towards thinking may have been the catalyst for growth in CT ability. The latter possibility is consistent with both Marzano's (1998; 2001) view on the powerful role of positive self-system changes on educational outcomes as well as previous research and theory that suggests that not

only are CT dispositions and abilities correlated, but that these metacognitive processes are dependent on one another in order to conduct CT (Ennis, 1996; 1998; Halpern, 2003a; 2006; Ku, 2009; Ku & Ho, 2010a; Perkins & Ritchhart, 2004; Valenzuela, Nieto & Saiz, 2011). Though the relationship between changes in CT dispositions and changes in CT ability is not yet well understood, based on the results of past research (e.g. Reed & Kromrey, 2001), it seems likely that beneficial effects of training on CT ability can be observed in absence of any changes in self-reported thinking dispositions. Nevertheless, the impact that disposition towards thinking may have on CT ability should not be overlooked, given that both theory and research suggest that the two are intimately related (Ennis, 1998; Facione et al., 2002; Halpern, 2003a, 2006; Ku & Ho, 2010a; Marzano, 1998).

On balance, it is clear that a variety of different conceptualisations of CT have been used as the basis to instruct and train CT skills. However, the degree of alignment between conceptualisations of CT, training methods and measures of CT used can vary considerably from one study to the next; and it is unclear precisely what impact these variations have on the quality of interventions and their overall impact on students. What is clear is that there is not a very strong emphasis on alignment in the CT intervention literature in general. Similarly, there is not a very strong focus on explicit experimental manipulation of instructional tools within CT intervention study trials. Across studies, it may be possible to broadly evaluate whether or not some instructional tools are better than others. This issue is considered again in the next chapter where previous studies on argument mapping-infused CT training are compared with other CT intervention methodologies.

2.3 Summary of the Previous Research

In summary, research has shown that when *to-be-remembered* information is presented in a well-organized manner, the ability to comprehend and recall that information is enhanced. More specifically, hierarchically structured, graphic organisers can facilitate improved recall and comprehension. In terms of metacognitive, CT skills, Alvarez-Ortiz's meta-analysis (2007) suggests that the crucial factor in improving CT is the explicit teaching of CT itself. Abrami and colleagues' (2008) meta-analysis also suggests that making CT learning objectives explicit to students is vital to improving CT ability. However, as Abrami and colleagues' meta-analysis was conducted in light of very broad distinctions between different types of CT training courses, less is known about how more specific instructional methods, including different methods of *presenting information* to students (e.g. text-based summaries, outlines, graphic organisers, etc.), impact overall training benefits. Such instructional methods are important to consider in CT training interventions because teaching strategies that facilitate the assimilation of argument structures (i.e. analysis of argument structures) and the quality of the logical relationships among propositions within arguments (i.e. evaluation of the argument structures) may in turn facilitate significant growth in critical thinking abilities (van Gelder, 2003). Thus, while researchers (e.g. Solon, 2007) have begun to broach the critical issue of the alignment of CT theory, training and measurement in intervention studies, less research has focused on the manipulation of specific instructional design strategies for training CT.

Central to the goals of the current thesis is an examination of the potentially beneficial effects of argument mapping on both lower-order (memory and comprehension) and higher-order (analysis, evaluation, inference, and reflective

judgment) thinking processes. Argument mapping is a form of graphic organisation that may facilitate the assimilation of argument structures for purposes of developing CT and enhancing memory and comprehension. Thus in the next chapter, a detailed discussion of both argument mapping and previous research which has examined the effects of argument mapping training is presented. Extant research in this area suggests that training in argument mapping may promote skill development across the full gamut of thinking processes identified in Chapter 1, from memory to critical thinking ability.

Chapter 3

Argument Mapping

This chapter defines argument mapping and provides a brief history of its use. A selective focus on certain aspects of argumentation is also presented in order to discuss the principles of argument mapping in terms of the nature of argumentation; specifically, the nature of informal logic, the criteria for evaluating informal logic, and the use of informal logic in the context of ill-structured problems. Previous research conducted on the effects of argument mapping interventions on critical thinking performance is also presented, as is the general rationale for why argument mapping is examined in the current research as a method of enhancing memory, comprehension and critical thinking.

3.1 Argumentation

Argumentation is “a verbal and social activity of reason aimed at increasing (or decreasing) the acceptability of a controversial standpoint...by putting forward a constellation of propositions intended to justify (or refute) the standpoint” (van Eemeren et al., 1996, p. 5). In argumentation, one must work with some form of representation of the argument, in order to identify, analyse and evaluate the logical relationships amongst propositions within the argument (Walton, 2006). This idea of representing the logical relationships amongst propositions within argumentation was presented by Stephen Toulmin in his influential book, *The Uses of Argument* (1958). In his book, Toulmin argued against traditional, formal logic (e.g. syllogistic reasoning) as the sole strategy of inference, as he believed reasoning should not be concerned solely with inferential relationships, but also with the justificatory function

of argumentation (i.e. establishing a claim and then supporting it by a statement of justification).

Although common in scientific thinking, the provision of justification for propositions was used less by philosophers when building arguments using traditional strategies, which tended to emphasise training in the use of syllogisms - the core of formal logic. Toulmin argued in favour of informal logic, or informal reasoning, which operates at the heart of scientific thinking. He found that this type of logic required a grounded, practical strategy which stressed the provision of justification for every reason and objection related to a core claim. According to Toulmin's approach, each claim within an argument must be supported by a satisfactory warrant (i.e. a generally accepted belief or value which is taken for granted by both the writer and reader), which either explicitly or implicitly bridges both datum and conclusion. Through his investigation of argumentation, Toulmin derived what many perceive as the first example of modern argument mapping (see Figure 3.1). Thus, argument mapping is grounded in informal reasoning; where the strength of any claim presented is dependent solely on the strength of its justification, in terms of logical strength (i.e. the strength of the relationship amongst propositions used in order to infer a conclusion), credibility (i.e. trustworthiness or reliability of the information's source) and relevance (i.e. the pertinence or applicability of one proposition to another).

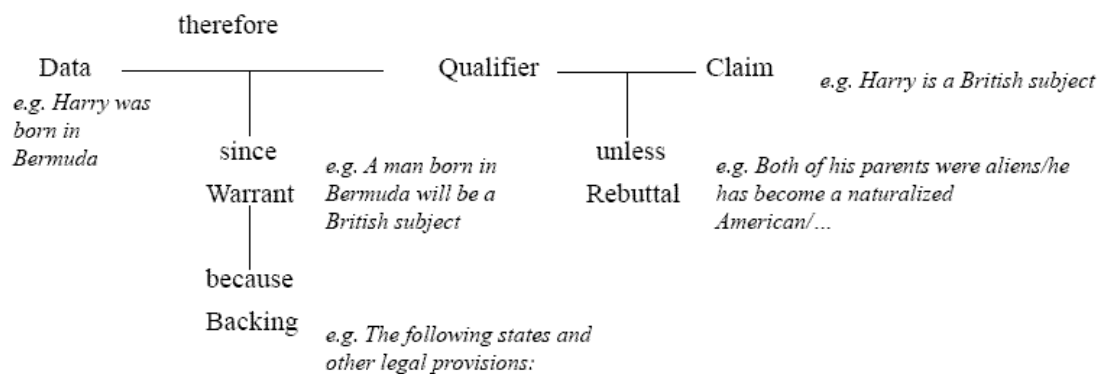


Figure 3.1: An example of Toulmin's Model of Mapping

Informal reasoning is a useful strategy of deliberation for when relevant information to support or refute a claim is scarce, when an argument is open-ended, or when a problem to which the argument pertains is ill-structured (Means & Voss, 1996); in which case, multiple, alternative, solution-methods can be (and often are) constructed to attempt to solve the problem. Ill-structured problems (as discussed in Chapter 1) often refer to everyday, real-world dilemmas where there is conflicting information; and where those involved in the argument disagree about the potential solutions to the problem. While training in formal logic is invaluable in many situations (e.g. for philosophical deliberation, following or constructing a line of deductive reasoning, identifying logical fallacies and rhetorical strategies in political discourse), informal reasoning is essential because it is very often necessary to provide justification for propositions, claims, or decisions when grappling with everyday, real-world, ill-structured problems. This applies to both local and personal problems and questions (e.g. “*Is social networking technology a waste of my time or a useful time investment?*”), as well as more complex, global problems and questions (e.g. “*How can global warming be stopped?*”).

On the other hand, many well-structured problems, such as physics and mathematics problems are open to resolution using formal or mathematical logic and may have singular solutions that are either right or wrong. For example, when presented with the question: “*If a car has been travelling at 60 mph for 6 hours, how far has the car travelled?*”, mathematical computation implies that the only correct answer is 360 miles, and all other solutions are incorrect. However, the problem with relying exclusively on formal logic for the purpose of problem-solving is that for many real-world problems, arriving at solutions - in terms of simply being ‘*right-or-wrong*’ solutions - is not always necessarily feasible.

As arguments can possess a variety of claims, supports and objections, it becomes increasingly important to be certain of the linkage among propositions within the argument, especially in terms of evaluating their relevance to one another and their logical strength of interdependence. However, in the context of reading, writing or deliberation, this ‘inter-linking’, or integration, of propositions within an argument can be difficult and may lead to a cohesion deficit (Duchastel, 1990). For example, when faced with reading about a newly encountered topic, a student may experience a cohesion deficit if they are unable to integrate arguments within the text itself, or link these novel arguments with knowledge they may have previously assimilated. Thus, the relations among propositions in an argument must be made clear to the reader; otherwise, the argument may appear incoherent and may cause a cohesion deficit in the mind of the reader. In terms of the structure of the argument on the page, any potential incoherence in the logical interdependencies between propositions may require that the reader engage in additional cognitive processing such that they can coherently tie the arguments together – thus increasing cognitive load and potentially having a negative impact on ongoing memory, comprehension, and critical thinking processes. In addition, a lack of cohesion between or among propositions in the learning context can increase the likelihood of a student failing to achieve their goal of constructing a good, logical argument under examination (Fox, Grunst & Quast, 1994; Duchastel, 1990).

3.2 What is an Argument Map?

All arguments share the characteristics of being composed of a network of propositions, prose-based or otherwise, that are structured via logical, inferential relationships. An argument map is a visual representation of that logically structured network of reasoning, in which the argument (often extracted from text) is made

unambiguous and explicit (i.e. with no need for attention switching from paragraph to paragraph or from page to page in a linear text, in search for reasons and objections to the central claim around which the argument map is constructed). The argument map (see Figure 3.2 for an example) uses a ‘box and arrow’ design in which the boxes represent propositions (i.e. the central claim, reasons, objections and rebuttals) and the ‘arrows’ among propositions indicate the inferential relationships linking the propositions together (van Gelder, 2002). Thus, the provision of an arrow between two propositions indicates that one is evidence for or against another.

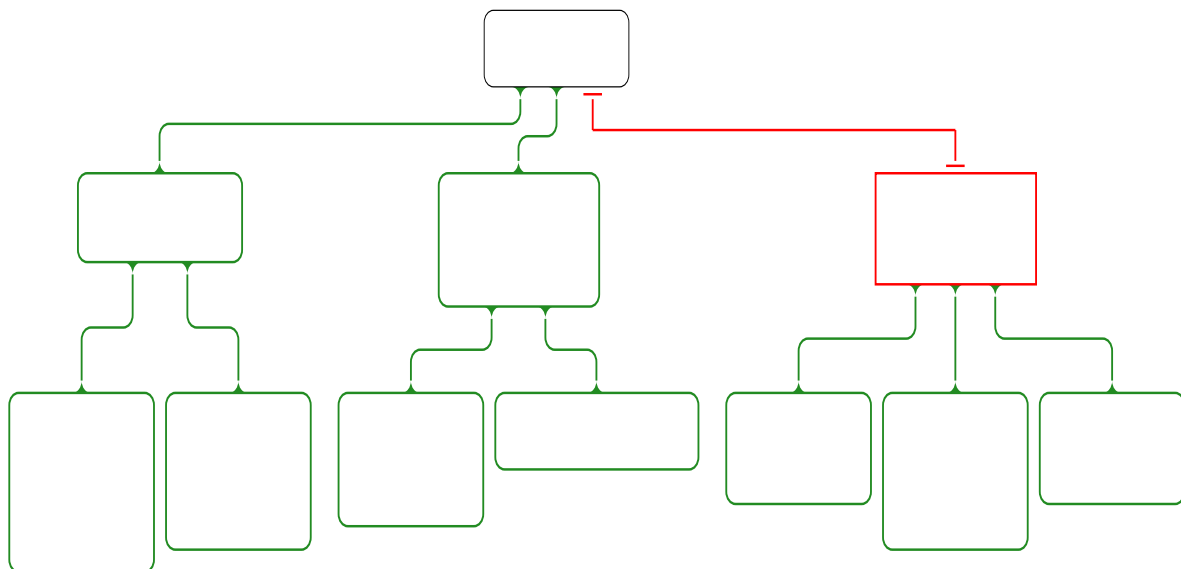


Figure 3.2: An example of an argument map using *Rationale*TM (van Gelder, 2007)

Similarly, colour can be used in an argument map to distinguish evidence for a claim from evidence against a claim. For example, in the software package *Rationale*TM (van Gelder, 2007), green represents a support and red represents an objection to the claim above. More generally, a *good* argument map is designed in such a way that if one proposition is evidence for another, the two will be appropriately juxtaposed (van Gelder, 2001) and the link explained via a relational cue, such as *because*, *but* and *however*.

Modern argument mapping software, such as *Rationale*TM (van Gelder, 2007), allows for the creation of one's own argument map, by means of typing text into blank boxes and dragging these newly created propositions to their appropriate locations on the map. Single propositions, or entire branches of the argument, can be removed or dragged to another location, and edited in the process, in order to facilitate the reconstruction and easy manipulation of an argument map. This aspect of argument mapping is also useful when analysing and evaluating arguments. For example, if an individual observes an error in reasoning within an argument map, they can edit or delete propositions, add propositions, and edit or remove an entire chain of reasoning. Similarly, they can relocate propositions or chains of propositions to a new location on the map, and thus deepen their analysis and evaluation of propositions and argument structures in the process. In this sense, the manner in which propositions and chains of reasoning can be manipulated within an argument map may encourage deeper analysis and evaluation of the argument, as well as further refinements of its inferential structure.

3.3 A Brief History of Argument Maps

Toulmin was not the first person to use argument maps. Argument maps have existed for well over a century (Buckingham-Shum, 2003). In fact, the style of argument mapping endorsed by Toulmin (1958), van Gelder (2000, 2001, 2003, 2007) and others dates back to 1826, when the practice was conducted by the logician Richard Whately, in his book *Elements of Logic* (Reed, Walton & Macagno, 2007). Whately (1826) presented a hierarchical chain of arguments (see Figure 3.3) which sought to reduce the structure of an argument to the necessary, relevant propositions such that logic and reasoning could be applied by the reader (Reed, Walton & Macagno, 2007).

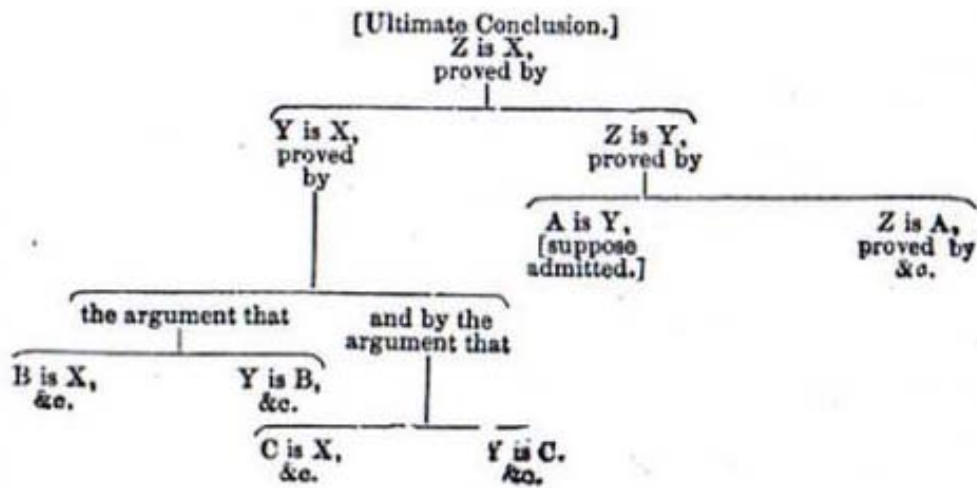


Figure 3.3: An example of Whately's Model of Mapping

As Toulmin's ideas spread among those in the field of argumentation during the late 1950's, the value of argument maps was soon recognised, resulting in the regular appearance of argument mapping in textbooks (e.g. Scriven, 1976; Fisher, 1988). Though the argument maps Toulmin developed were relatively simple, more recent attempts have been made to create maps of very complex arguments. For example, Robert Horn (1999) devised a set of seven maps that present arguments in relation to the question, *Can Computers Think?* Horn's argument mapping strategy organised and presented arguments in a graphical array (see Figure 3.4) to facilitate easier assimilation of the structure of arguments, for such purposes as improving teaching and learning (Horn, 2003; Monk, 2001). Horn's work demonstrates that increasingly complex, ill-structured arguments (which may be difficult for a majority of the population to comprehend), can be translated into a manageable representation. Notably, Horn (2007) has also applied his mapping strategy to a range of social problems, in an effort to facilitate the reasoning and consensus of groups working to solve complex social problems.

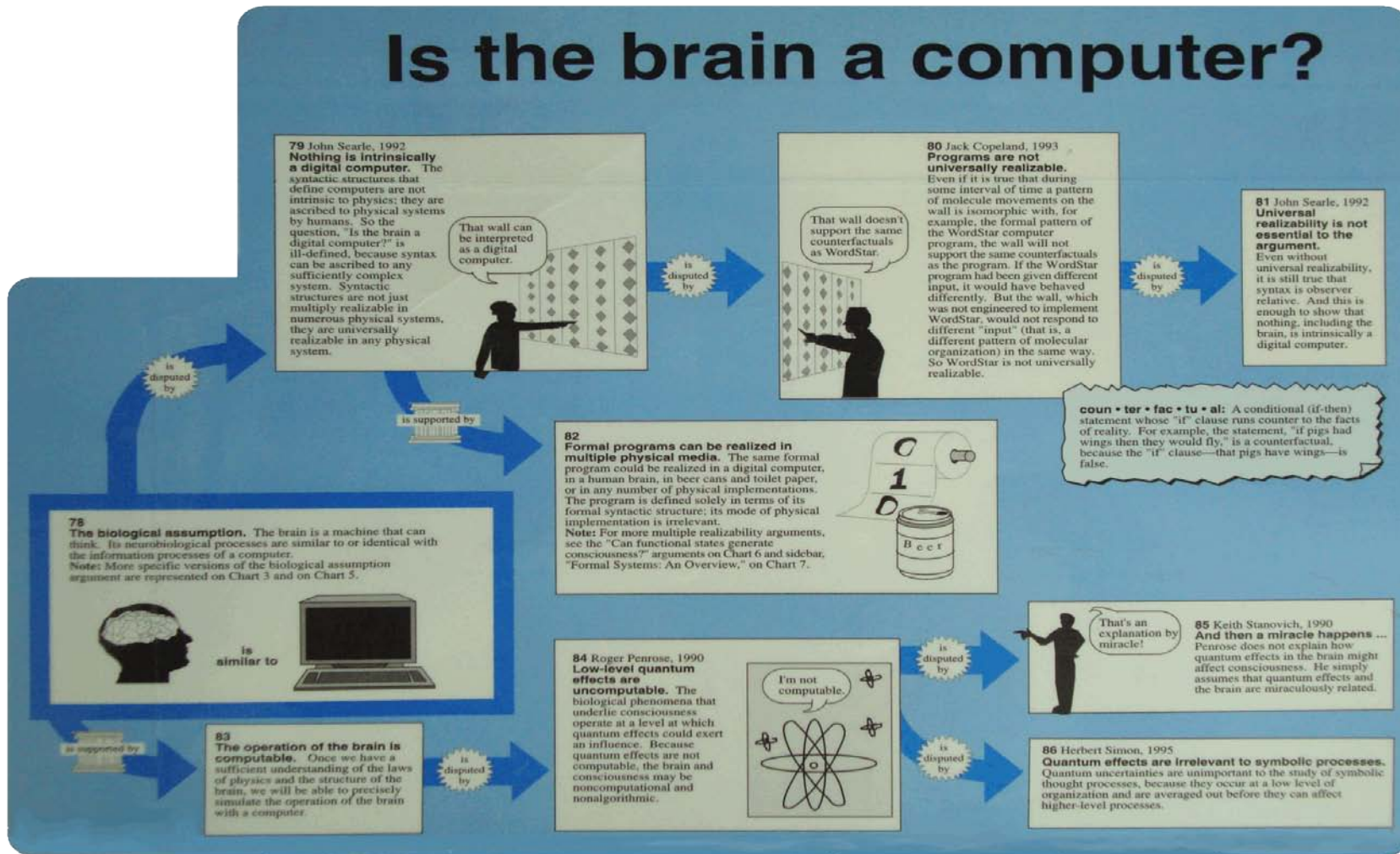


Figure 3.4: An excerpt from Horn's Map "Can Computers Think?"

Although argument maps have been in existence for almost 200 years (Buckingham-Shum, 2003; Reed, Walton & Macagno, 2007; van Gelder, 2007), only recently has argument mapping technology become user friendly. In the past, argument mapping was a slow and tedious task of drawing out measured boxes, filling them in with the appropriate text and accurately connecting them with arrows using only pen and paper. As a result, they have not been widely used as learning tools, despite the possibility that they may provide considerable advantages over standard prose as a medium for presenting reasoning. With the advent of various argument mapping software programmes, such as *Rationale*TM (van Gelder, 2007), the time required to construct an argument map has been substantially reduced, as the construction of an argument using this software needs only the choice of an appropriate box and associated relational cue, the typing of text into the box, and the selection of an appropriate location for the box in the argument structure (i.e. in relation to other propositions).

Perhaps as a result of the relatively recent advancements in argument mapping software, little research has been conducted to test its effects on learning. Though some research has found that argument mapping can offer considerable advantages over traditional text as a method of presenting information (e.g. Butchart et al., 2009; van Gelder, 2001; van Gelder, Bissett & Cumming, 2004), further possible benefits remain largely untested. Thus, further examination of argument mapping as a learning tool is required.

3.4 Why use Argument Maps?

An argument map is the diagrammatic version of any prose-based argument made in speech or in text form. Previous research on the use of diagrams (Larkin & Simon, 1987) and other representational aids in educational settings have shown

positive effects on comprehension and memory (e.g. Berkowitz, 1986; Oliver 2009; Robinson & Kiewra, 1995; see Chapter 2). Previous research has also shown that the integration of text-based information into diagrammatic form can also decrease cognitive load (e.g. Chandler & Sweller, 1991; Pollock, Chandler & Sweller, 2002; Tindall-Ford, Chandler & Sweller, 1997; see Chapter 1).

Diagramming can offer advantages over traditional text-based presentation of information because the indexing and structuring of information can potentially support essential computational processes (i.e. comprehending propositional relationships). In addition, as diagrams can group all the necessary information together based on relatedness, it makes the search for specific, relevant information more efficient, which in turn supports perceptual inferences (i.e. “seeing” the conclusion; Larkin & Simon, 1987, p. 98). Thus, assimilating information and inferences from an argument map is believed to be substantially easier than assimilating information and inferences within potentially unaccommodating linear text. In addition, asking students to produce diagrams such as argument maps can provide teachers with valuable insights into a student’s ‘mental model of the argument in question’ (Butchart et al., 2009). Furthermore, such information can be used to support teachers in offering feedback to students or scaffolding student learning from simple to complex levels of argument comprehension, analysis, and evaluation. Logically, as expertise in argument mapping grows, so does the ability to present a well structured argument, which allows for improvement in writing ability as well.

Though other forms of argument diagramming exist, such as *concept mapping* and *mind-mapping* (Buzan & Buzan, 1997), they differ from argument mapping based on the manner in which they are organised and the way in which each ‘proposition’ is presented. For example, concept mapping is a mapping method in which a set of

terms, sentence fragments or full sentences (which represent propositions), are joined by a set of lines and arrows, based on some relation (Reed, Walton and Macagno, 2007); and refer to sentences within a text (see Figure 3.5). In some cases, the ‘propositions’ (i.e. boxes), within concept maps do not contain words, but a number which refers to a proposition from a separate text or from a list of propositions (for example, see Figure 3.6). Similarly, mind-mapping (see Figure 3.7) is a method of concept mapping in which words or phrases are integrated with coloured (or non-coloured) pictures (i.e. used to represent a concept), in which the coupled pictures and text are connected via lines or arrows to demonstrate some type of relationship. In the context of mind-mapping, the representation of the relationships among combinations of pictures and text is believed to aid memory storage and retrieval (Buzan & Buzan, 1997).

The problem with many concept mapping techniques is that they do not present an argument per se. Instead, they present a graphical structure that acts as a representation of a separate text, which might be used to diagram: the links among concepts, decision-making schemes, a set of plans or instructions, or at best, act as an argument overview – which does not represent the argument in full. Thus, because the text of the argument and the diagram may often be separate entities, concept mapping may become more cognitively demanding by adding the necessity of switching attention from text to diagram and vice versa (e.g. Chandler & Sweller, 1991; Pollock, Chandler & Sweller, 2002; Tindall-Ford, Chandler & Sweller, 1997; see Chapter 1). In addition, if the reader of a concept map is not familiar with the information from the text that the map is derived, then the map itself becomes meaningless. For example, in Figure 3.7, the mind-map presents a relational link between ‘*Dry Wall*’ and ‘*Interest Rate*’ in the context of ‘*Supply & Demand*’ (Budd, 2004); yet presents

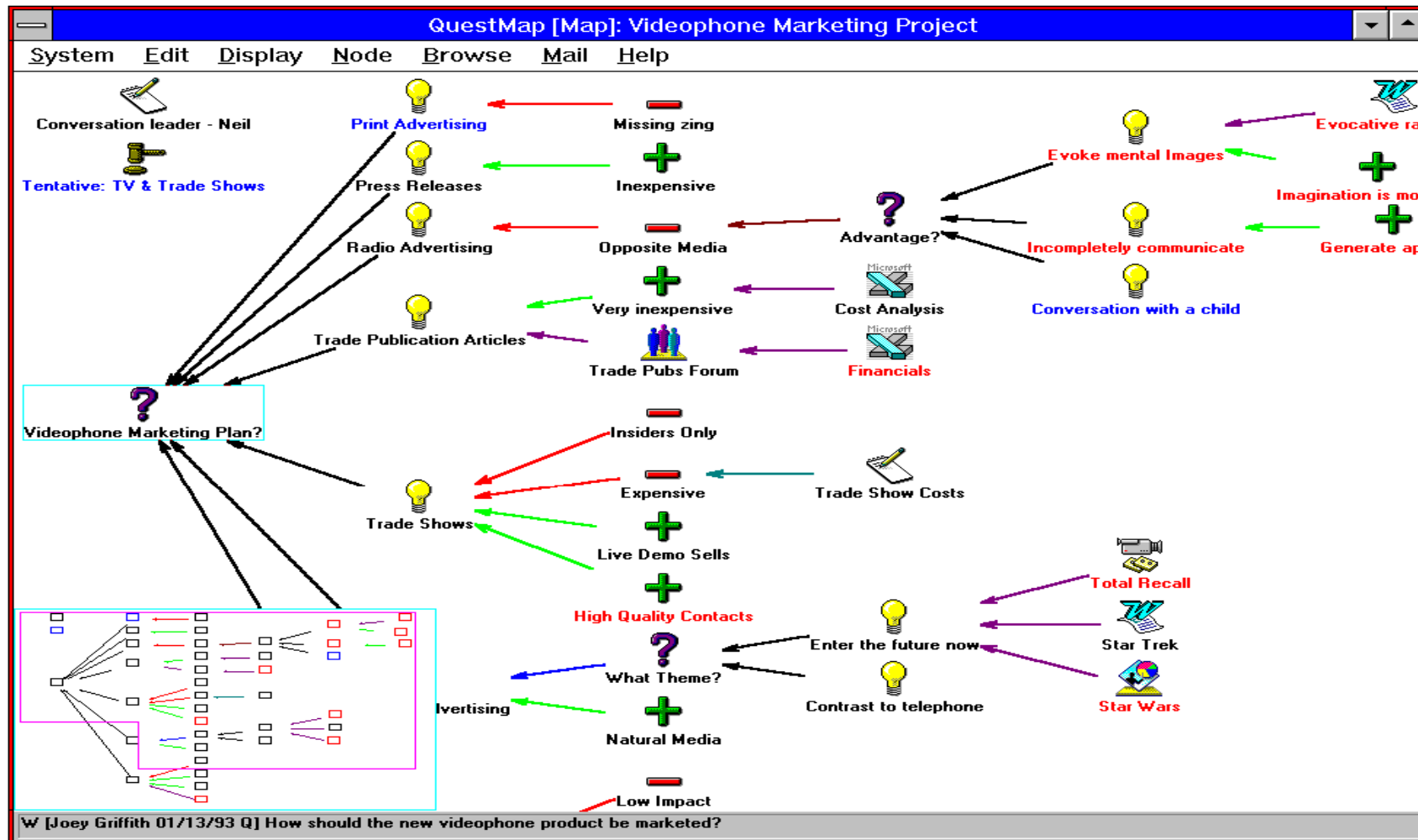
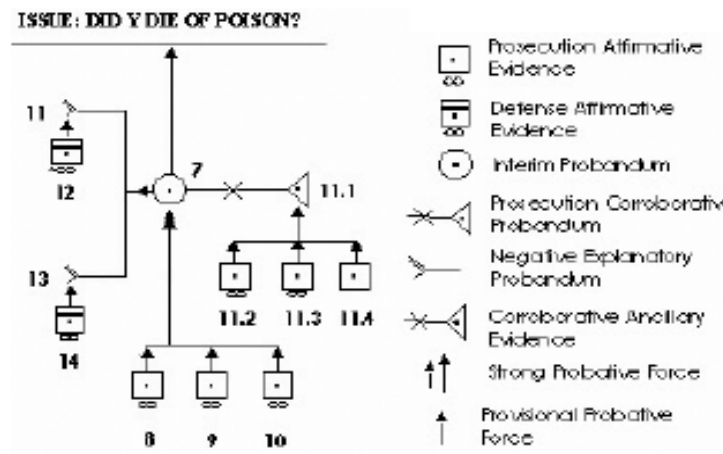


Figure 3.5: An example of a Concept Map created using QuestMap™ (Carr, 2003) in Reed, Walton & Macagno (2007).



Key List is as follows:

- 7 Y died, being apparently in health, within three hours after the drink of whiskey.
- 8–10 Y’s Wife and the Northingtons witness to 7.
- 11. Y might have died by colic from which he had often suffered.
- 11.1 Colic would not have had as symptoms the leg cramps and teeth-clenching; only strychnine could produce these ones.
- 11.2 Y’s wife and the Northingtons witness to Y’s cramps and teeth-clenching.
- 11.3 Expert witness to significance of symptoms.
- 11.4 No testimony as to strychnine traces in the body by post-mortem.
- 12. Anon witness to his former attacks.
- 13. Y might have died from the former injury to his side.
- 14. Anon witness to that injury.

Figure 3.6: John H. Wigmore’s (1913) Classically used Concept Mapping Strategy – An example of a concept map using numbers as reference points.

neither explanation, nor a relational cue (e.g. *because*, *next*, *causes*, *but*, *alternatively*, etc.) as to the nature of the link. There are neither sentences, nor any inferential structures to facilitate comprehension. Thus in this context, concept mapping strategies may not necessarily be useful pedagogical aids that are open to analysis by everyone.

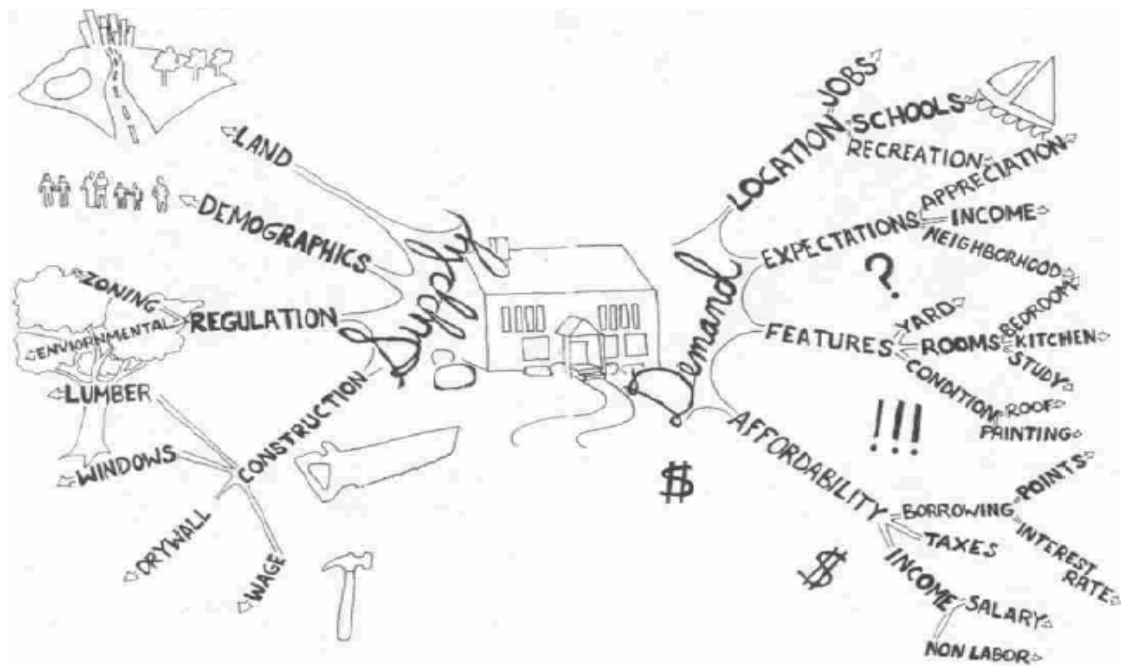


Figure 3.7: An example of a Mind Map (Budd, 2004).

Furthermore, though these mapping strategies are used primarily to represent concepts, decision-making schemes and the organisation of plans (which they may do adequately), they are not specifically designed to present arguments. That is, there is no strict usage of relational cues within these concept mapping strategies because of their many applications, such as creating instructions, procedures, heuristics, or simply linking concepts; rather than exclusively linking propositions within an argument. For example, in a concept map, relational cues such as: *next* may be used to represent the next step in a set of instructions and *alternatively* may be used to refer to an alternative step in a problem-solving heuristic. However, in argument mapping,

relational cues such as *because*, *but* and *however* are necessary and must be used consistently, as they are what ‘glue’ the structure of the argument together. That is, these relational cues directly link one proposition to another. These relational cues further suggest the path in which the argument is moving in terms of either broadly supporting or disputing a core claim. Pedagogically, the existence of these cues within the argument map is important as adequate use of relational cues may help students and teachers to avoid cohesion deficits (Dee-Lucas & Larkin, 1995; Duchastel, 1990). Thus, while argument mapping is similar to other mapping strategies in some respects, due to the nature of argument mapping (i.e. strictly dealing with argument structures), it specifically focuses on the logical and evidential relationships among propositions.

3.5 Previous Research Conducted on Argument Mapping

Given that computer-based argument mapping (AM) is a relatively recent learning strategy (van Gelder, 2000; 2007), little research has examined its potentially beneficial effects on learning. Though there has been no research conducted on AM’s ability to enhance memory or comprehension (to this researcher’s knowledge), there has been some research conducted which suggests that AM training can enhance critical thinking (Alvarez-Ortiz, 2007; Butchart et al., 2009; van Gelder, 2001; van Gelder, Bissett & Cumming, 2004). For example, Alvarez-Ortiz’s (2007) meta-analysis (discussed in Chapter 2) found that academic courses which provided at least some critical thinking (CT) instruction facilitated CT ability better than academic courses that did not provide CT instruction. It was also found in the meta-analysis that students who participated in CT courses that used at least some AM within the course achieved gains in CT ability with an effect size of .68 SD, CI [.51, .86]. In courses where there was “lots of argument mapping practice” (LAMP), there was also a

significant gain in students' CT performance, with an effect size of .78 SD, CI [.67, .89]. The effect sizes for LAMP studies compared favourably with other CT intervention strategies reviewed in Chapter 2, including teaching CT through philosophy (.26 SD) and computer-supported CT training (.49 SD).

The effect size for LAMP studies was derived by Alvarez-Ortiz (2007) from a number of separate studies (e.g. Donohue et al., 2002; Butchart et al., 2009; Twardy, 2004; van Gelder 2000; 2001). However, each of these studies had certain methodological flaws that make it difficult to draw substantive conclusions in relation to the beneficial effects of AM training on CT performance outcomes. Nevertheless, the findings from these studies, reviewed below in more detail, are suggestive and point to the need for more research on the potential benefits of AM.

Donohue, van Gelder, Bisset and Cumming (2002) presented the findings from multiple studies conducted on AM's effect on CT, including those by van Gelder (2000; 2001). Van Gelder (2000) provided 85 undergraduate philosophy students with a 16-week *reasoning* course over two semesters. In Semester 1, 32 students were taught in a traditional 'chalk and talk' manner through lectures, tutorials, homework and exams. In Semester 2, a different cohort of 53 students was presented the same course, but taught instead through AM. Students in the AM group "were officially expected to devote around 10 hours per week" on AM-based homework exercises. Both groups (i.e. the AM group and the no AM group) completed the same amount of class work and were both expected to complete the same amount of homework - only the format of the work was different between the two groups. However, van Gelder reports that it was "impossible to be very exact about how many hours they actually did" (van Gelder, 2000, p. 6). Students were assessed both before and after the course via the WGCTA (Watson & Glaser, 1980).

Results revealed that though there was no significant gain made by the traditional group on CT performance from pre-to-post-testing ($p = .44$, effect size = .02), there was a significant gain made by the AM group ($p < .001$, effect size = .41). Though the results from this study suggest that the use of AM can significantly enhance CT performance, the design of the study makes it difficult to interpret the results. Specifically, with a full semester of study already under their belt, the second cohort of students who received the AM training may have been at an advantage relative to the control group, given the potential for extra knowledge, experience, or dispositional changes from thinking critically in other courses. In other words, students were not randomly assigned to experimental conditions and the students who received the AM training may have differed from the control group in a number of important respects. In addition, the amount of practice completed by the groups was not adequately measured or controlled (i.e. total hours of engagement was not measured, nor was the time it took students to complete individual homework exercises).

Van Gelder (2001) replicated his previous study (i.e. van Gelder, 2000), by providing undergraduate philosophy students with a semester-long (i.e. 12-week) critical thinking course, in which students were “intensively trained” in CT through the use of AM (van Gelder, 2001, p.7). Students’ CT ability was tested both before and after the college semester using alternate forms of the CCTST (Facione, 1990a). Results revealed an improvement in CT ability, yielding an effect size of .84; which accounted for a gain of almost one standard deviation in critical thinking over the course of the semester. It is important to note that, though van Gelder credits much of this gain to the inclusion of AM to the course, he also admits that this gain could also be due to other underlying factors within the course, such as the practice regime.

Notably, this study included neither a control group, nor an alternative CT training regime with which to compare AM training.

Like van Gelder (2000; 2001), Twardy (2004) also provided undergraduate philosophy students with a semester-long CT course, taught through AM. Students were both pre-and-post-tested via alternative forms of the CCTST. Results revealed a gain in students' CT performance from pre-to-post-testing with an effect size of .72 SD. Though research by Twardy (2004) also suggests that AM can enhance CT ability, like van Gelder (2001) and many other CT intervention studies, this study included neither a control group, nor an alternative CT training regime with which to compare AM training. As a result, the lack of either a comparison or a control condition makes it difficult to interpret the observed gains in CT.

Though not included in Alvarez-Ortiz's (2007) meta-analysis; van Gelder, Bissett and Cumming (2004) also conducted research which examined the effects of AM on CT. Fifty-one undergraduate philosophy students were provided with a 12-week semester-long CT course taught through the use of AM. Students were tested prior to commencement of the course using the CCTST. Roughly half of the class completed Form A, while the other half completed Form B. During the course, students were provided with both homework exercises and the opportunity to complete as many additional practice exercises in AM as they wished. Students also attended one tutorial per week in which they had access to both AM software and to direct personal guidance from their tutors. After completion of the course, students were post-tested using the CCTST. Those who completed Form A during pre-testing completed Form B during post-testing and vice versa. Results revealed that CT scores increased significantly from pre-to-post-testing with a large effect size of .8 SD, CI [.66, .94].

More recently, research by Butchart et al. (2009) also investigated the effect of a CT course taught through AM on students' CT performance. Two groups of students who attended a CT course taught through the use of AM (i.e. a module with online automated feedback for AM exercises and a module that contained AM exercises only; N = 43 and N = 41, respectively) were compared with each other as well as a 'standard' CT module (i.e. no AM; N = 65).¹¹ Students participated in a semester-long CT course which heavily concentrated on two skills: analysis (i.e. identifying conclusions, premises, unstated assumptions and the structure of arguments) and evaluation (i.e. evaluating arguments, criticising arguments and identifying fallacies). Notably, though Butchart and colleagues did not explicitly mention the *Delphi Report* in their research, the skills of analysis and evaluation have been identified by the *Delphi Report* as core CT skills (see Chapter 1). The course outline used in this study is presented in Table 3.1. Prior to commencement of the course, students completed the CCTST Form A as a pre-test. During the course, students were provided with eight homework assignments and 10 sets of exercises. Automated feedback was provided to students in the automated feedback AM group while they performed their AM exercises. After completion of the course, students were post-tested using Form B of the CCTST.

Results revealed that those who received automated feedback for their AM exercises showed a significant gain in CT ability with a medium effect size of .45 SD. Students who completed the AM exercises without automated feedback showed a gain with an effect size .22 SD; and those who participated in a standard CT module showed a gain with an effect size of .19 SD. However, statistical differences among

¹¹ Those in the 'standard' CT training group were provided with the same course structure and taught the same content, "as far as possible", as those in both the AM groups (Butchart et al., 2009, p. 278). However, it unclear from the empirical report how this worked in practice.

the three groups in this study were not reported. Furthermore, as admitted by the authors, participants in the automated feedback group could have been provided with more informative feedback, as opposed to receiving only automated notice of a 'correct' or 'incorrect' response to their labelling of a proposition (i.e. the function of each proposition, for example, as a support/premise or a conclusion). One could argue that this automatic 'correction' of AM exercises is a very shallow form of feedback, as an explanation as to *why* the answer is 'correct' or 'incorrect' would be more informative to the student.

Though the results from these studies seem to suggest that CT courses taught through the use of AM may improve CT ability, there have been a number of methodological problems in the research conducted to date. For example, three of the five studies reviewed above (i.e. Twardy, 2004; van Gelder, 2001; van Gelder, Bissett and Cumming, 2004), did not use a control or alternative CT group with which to compare results or gains in CT of those in the AM groups, thus compromising internal validity (van den Braak et al., 2006). In addition, although Alvarez-Ortiz's (2007) meta-analysis suggests that semester-long training courses in AM produce greater gains in CT skills (when compared with standard semester-long courses in introductory philosophy), AM has not been directly compared with other methods of teaching CT, apart from a standard (no AM) CT course used in the research by van Gelder (2000) and Butchart et al (2009). However, even with the inclusion of these comparison groups, there were still methodological concerns. For example, though both van Gelder and Butchart and colleagues compared their AM groups with other groups (i.e. a control group, and both a control group and alternative AM group, respectively); groups were not adequately matched on baseline CT ability, cognitive ability, or disposition towards thinking.

Table 3.1: Course outline used in Butchart et al. (2009)

Week	Lecture	Tutorial Exercises
1	Why study argumentation?	Pre-test
2	Introduction to reason and argument	Identify arguments: Distinguish arguments from non-arguments. 10 multiple choice questions (no argument mapping exercises).
3	Argument analysis 1	Identify conclusions: Identify the main conclusion of the argument. 12 argument mapping exercises.
4	Argument analysis 2	Map simple arguments: Create a map of the argument. 11 exercises.
5	Argument evaluation: Truth, justification	Map complex arguments: Create a map of the argument. 14 exercises.
6	Argument evaluation: Clarity, relevance, strength	Argument structure: Map the argument and identify the role played by a particular statement or the argumentative strategy used. 14 exercises with multiple choice questions.
7	Criticism: Objections and replies	Argument structure: As above. 14 exercises with multiple choice questions.
8	Criticism: Assumptions	Assumptions: Map the argument, selecting the correct assumption from a list. 8 exercises.
9	Fallacies (ambiguity)	Fallacies of ambiguity: Map the given argument and then answer the question to identify the flaw. 8 exercises with multiple choice questions.
10	Fallacies (relevance)	Fallacies of relevance: Map the given argument and then answer the question to identify the flaw. 8 exercises with multiple choice questions.
11	Fallacies (truth, ambiguity)	Fallacies of truth: Map the given argument and then answer the question to identify the flaw. 8 exercises with multiple choice questions.
12	Fallacies (strength)	Fallacies of strength: Map the given argument then answer the question to identify the flaw. 8 exercises with multiple choice questions.
13	Reason and happiness	Post-test

In addition, in neither study were groups randomly assigned to experimental conditions. Rather, groups were assigned to conditions based on the semester (i.e. 1 or 2) they participated in the intervention. Furthermore, the pre-and-post-test scores (and the resultant gains from pre-to-post-testing) of the groups in both studies were not compared. As a result, it is difficult to assess whether or not the groups possessed similar or statistically different CT abilities prior to their participation in the course; and whether or not gains across conditions are statistically different from one another. Thus, without randomly assigning students to different conditions, statistically comparing results of the different conditions, or matching participants on baseline CT ability, cognitive ability, or even disposition towards thinking, the potential benefits of AM training are also questionable.

In summary, though evidence suggests that CT can be taught and enhanced through the use of AM, the studies examining this claim are methodologically flawed in a number of different ways. Due to the various shortcomings in past research described above, it is clear that more carefully controlled research is necessary before any solid conclusions regarding AM's effects on CT can be drawn with confidence.

3.6 Rationale for the Current Research

All of the research which has examined the effects of AM on learning outcomes has focused exclusively on CT ability. This thesis aims to examine the effect of AM on a variety of cognitive and metacognitive processes including memory, comprehension, CT, and the CT sub-skills of analysis, evaluation, inference and reflective judgement. It is hypothesised that learning through AM can help students to overcome cognitive load and significantly improve memory, comprehension and CT. There are three substantive reasons for proposing this hypothesis.

First, unlike standard text, AMs represent arguments through dual modalities (visual-spatial/diagrammatic and verbal/propositional), thus facilitating the latent information processing capacity of individual learners. Second, AMs utilise Gestalt grouping principles that facilitate the organisation of information in working memory and long-term memory, which in turn facilitates ongoing comprehension, analysis, evaluation and inference activities necessary to further promote the development of CT and reflective judgment skills in the classroom. Third, AMs present information in a hierarchical manner which also facilitates the organisation of information in working memory and long-term memory for purposes of enhancing comprehension and promoting CT skills of analysis, evaluation, inference and reflective judgment.

In relation to the first reason, dual-coding theory and research (Paivio, 1971; 1986), Mayer's (1997) conceptualisation and empirical analysis of multimedia learning, and Sweller and colleagues' research on cognitive load (Sweller, 2010; see Chapter 1), suggests that learning can be enhanced and cognitive load decreased by the presentation of information in a visual-verbal dual-modality format (e.g. diagram and text), provided that both visual and verbal forms of representation are adequately integrated (i.e. to avoid attention-switching demands). That is, in order to keep cognitive load at a minimum, effort must be made to present information to students in a way that maximises the potential of dual-modality forms of representation, such that students can successfully integrate the information in working memory and store it in long-term memory. Given that AMs support dual-coding of information in working memory via integration of text into a diagrammatic representation, cognitive resources previously devoted to translating prose-based arguments into a coherent, organised and integrated representation are 'freed up' and can be used to facilitate deeper encoding of arguments in AMs, which in turn facilitates later recall (e.g. Craik

& Watkins, 1973), as well as subsequent, higher-order thinking processes (Halpern, 2003a; Maybery, Bain and Halford, 1986). Thus, it is hypothesised that AM can decrease cognitive load and enhance overall learning (van Gelder, 2003).

The second related reason for why AM is hypothesised to enhance overall learning is because AM also makes use of Gestalt grouping principles. Research suggests that when to-be-learned items are grouped according to Gestalt cues, such as proximity and similarity, they are better stored in visual working memory (Woodman, Vecera, & Luck, 2003; Jiang, Olson & Chun, 2000). For example, Jiang, Olson and Chun (2000) examined how visual stimuli are organised in working memory by asking participants to identify target stimuli in a series of change-detection tasks, based on both spatial configuration (i.e. location) and colour. Results revealed that when the spatial organisation, or relational grouping cues denoting organisation (i.e. similar colour, close proximity) are absent, memory performance is worse, and that when multiple spatial organisation cues (such as colour and location) are used, recall is better.¹² These findings suggest that visually-based information in working memory is not represented independently, but in relation to other pieces of presented information; and that the relational properties of visual and spatial information are critical drivers of successful memory and subsequently, learning. Given that related propositions within an argument map are located close to one another, the spatial arrangement complies with the Gestalt grouping principle of proximity.

In addition, modern argument mapping software such as *Rationale*TM (van Gelder, 2007) adopts a consistent colour scheme within argument maps in order to highlight propositions that support (green box) or refute (red box) the central claim,

¹² Notably, Jiang, Olson and Chun (2000) also found that for colour to influence memory performance, enhancement is dependent upon how spatial configuration is manipulated. On the other hand, there is no dependency upon colour for spatial location to influence memory. This is important to consider in Study 1, where the colour of study materials provided to students is manipulated.

thus complying with the Gestalt grouping principle of similarity (i.e. *greens* are grouped based on similarity, as are *reds*). Research by Farrand, Hussain and Hennessy (2002) suggest that in comparison with traditional methods of study, such as rote rehearsal, study materials (such as mapping strategies) which utilise integrated colour and visual-spatial arrangement can significantly improve recall. This finding suggests that collating items according to grouping cues, such as similarity (i.e. green *because*s and red *buts*) and spatial proximity (Jiang, Olson & Chun, 2000; Luck & Vogel, 1997), may simplify the method of representing information and increase the capacity of visual working memory (i.e. the visuospatial sketchpad; see Chapter 1).

The third reason for why AM is hypothesised to enhance overall learning is because it presents information in a hierarchical manner. When arguing from a central claim, one may present any number of argument levels which need to be adequately represented for the argument to be properly conveyed (see Figure 3.8). For example, an argument that provides a (1) support for a (2) support for a (3) support for a (4) claim has four levels in its hierarchical structure. More complex or ‘deeper’ arguments (e.g. with three or more argument levels beneath a central claim) are difficult to represent in text due to its linear nature; and yet it is essential that these complex argument structures are understood by a student if their goal is to remember, comprehend, analyse, or evaluate the argument, and to infer their own conclusions. As discussed in Chapter 1, the linear nature of text sometimes makes it troublesome to assimilate information because the reader must often switch attention from one paragraph, or even one page, to another and back and forth, in order to create some structural representation of the argument. This is consistent with research that suggests that individuals often have difficulty in extracting key propositions and evidence from text (c.f. Brem, 2000; Phillips, Norris & Macnab, 2010). On the other

hand, the hierarchical nature of AM allows the reader to choose and follow a specific branch of the argument in which each individual proposition is integrated with other relevant propositions in terms of their inferential relationship.

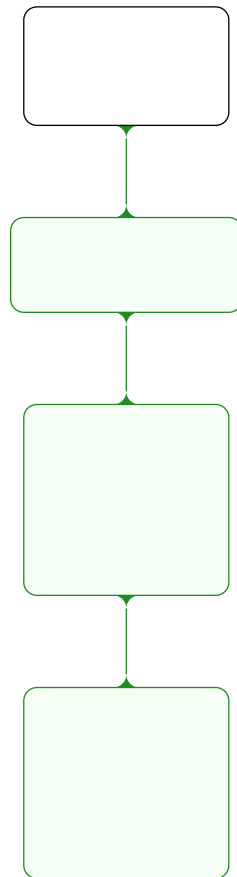


Figure 3.8: An example of a branch from an argument map

In the following chapters, empirical research will be presented which examines the broad claim that AM can be used to enhance a range of learning outcomes. In Chapter 4, the hypotheses for Study 1 will be presented, where focus is placed on the effects of manipulating study materials' format (i.e. text v. AM), colour (i.e. colour v. monochrome), size (i.e. 30-proposition v. 50 proposition), learning strategy (i.e. active v. passive learning) and study environment (i.e. isolated v. group setting) on outcomes of immediate recall, delayed recall and comprehension. The methodology, results and discussion of Study 1 will also be presented in Chapter 4.

Next, in Chapter 5, the details of Study 2 will be presented, where the focus is placed on the effects of AM, hierarchical outlining and CT disposition on the outcome variables of CT and reflective judgment ability. In Chapter 6, Study 3 will be presented, where the effects of AM training in an e-learning context are examined. A series of additional specific hypotheses around the potentially moderating effects of engagement, motivation and need for cognition on CT training gains in AM contexts will also be examined in Study 3. This will be followed by a general discussion of results in Chapter 7.

Chapter 4

Study 1: A Multi-Experiment Analysis of the Effects of Argument Map Studying & Construction on Students' Memory and Comprehension

4.1 Purpose

This chapter presents the empirical research carried out in Study 1, which involved a series of four experiments designed to examine the effects of argument map (AM) studying and construction on students' memory and comprehension¹³. In each experiment, AM studying and construction was compared with alternative learning techniques (e.g. text reading, text summarisation and hierarchical summarisation). As previous research has shown that the teaching of organizational strategies can help improve students' memory (e.g. Berkowitz, 1986; Farrand, Hussain and Hennessy, 2002; Taylor & Beach, 1984) and comprehension (e.g. Meyer, Brandt & Bluth, 1980; Robinson and Kiewra, 1995; Taylor, 1982), it was hypothesised that AM may be a particularly effective organisational strategy for learning (e.g. Butchart et al., 2009; van Gelder, 2000, 2001).

The four experiments reported below were conducted over a two year period, in the context of an annual module on *Psychology in Practice* delivered to first year psychology students. Experiment 1 was conducted in Year 1 and Experiments 2-4 were conducted in Year 2. The core manipulations and hypotheses associated with each of the experiments are summarized in Table 4.1.

¹³ Portions of Study 1 have been published and submitted for publication; and are cited in the References section, respectively (see Dwyer, Hogan & Stewart, 2010; 2011b).

Table 4.1: Core manipulations and hypotheses associated with each of the four memory experiments

Experiment	Design	Hypotheses
1. The effects of study format, size and colour on recall and comprehension.	3 x 2 MANCOVA – independent variables (IVs): study format (text, colour AM, and monochrome AM) and size (30, 50 propositions). Dependent variables (DVs): comprehension, recall memory.	Colour AM group will show better memory and comprehension than other groups. Monochrome map group will show better memory and comprehension than text group. Beneficial effects of AM condition will be greater for larger argument structures.
2. The effects of study format and setting on recall and comprehension.	2 (study format) x 2 (setting) MANCOVA – IVs: study format (text, AM) and setting (group testing, individual testing). DVs: comprehension, recall memory.	AM group will show better memory and comprehension than text group, regardless of environmental context. Isolated setting will yield better memory and comprehension performances.
3. The effects of study format and argument size on immediate and delayed recall.	2 x 2 x 2 Mixed ANCOVA. IVs: format (text, AM) and argument size (30, 50 propositions). DVs: immediate and delayed recall.	AM group will show better immediate and delayed recall than text group. Beneficial effects of AM condition will be greater for larger argument structures and greater for delayed assessment.
4. The effects of different active learning strategies on immediate recall.	Between Subjects ANCOVA. IV: learning strategy (3 levels: summarisation, outlining, AM). DV: immediate recall memory.	AM translation exercise will foster better immediate recall than outlining or summarisation. Outlining exercise will foster better recall than summarisation.

Experiments 1 and 2 examined the effects of AM reading, in comparison with text reading, on memory and comprehension performance, through the manipulation of: the colour of study materials (i.e. coloured vs. monochrome), argument size (i.e. 30-proposition vs. 50-proposition) and format of study materials (i.e. text vs. AM), as well as the environmental setting that studying and testing took place (i.e. a lecture

hall setting v. an isolated booth). Experiment 3 examined the effects of argument size (i.e. 30-proposition v. 50-proposition) and study material (i.e. AM versus text reading) on both immediate and delayed recall performance. Experiment 4 compared the effects of actively learning via AM construction, hierarchical summarisation (i.e. outlining) and text summarisation on immediate recall performance. All of the experimental manipulations were administered in a lecture hall setting, except in Experiment 2 where half the sample was tested in an isolated booth.

In each experiment, an argument was presented to students for purposes of study. In Experiments 1 and 2, the argument *Computers Can Think* was presented to students, which is an argument based on Robert Horn's (1999) AM, *Can Computers Think?* This argument was chosen because (1) the current research was inspired in large part by the work conducted on argument mapping by Robert Horn, who kindly provided this research programme with the full set of 7 wall maps on the artificial intelligence debate; and (2) because a smaller subset of arguments could be extracted from Horn's logically sound, pre-existing AM and could thus be used to test Horn's hypothesis (personal communication) that reading arguments from his maps is more conducive to learning than reading the same arguments from text. In Experiment 3, the argument *Aggression is Biologically Caused* was presented to students. This argument was chosen because it served as an introduction to the topic of *aggression* in the context of the *Psychology in Practice* module - a topic which would later be presented to students in their 1st Year Social Psychology module. In this sense, while the learning topic selected for Experiments 1 and 2 allowed for the development of an AM based on an existing AM that was widely recognised in the AM community, but was largely unfamiliar to first year psychology students, the topic used in Experiment 3 allowed for the development of an AM through the use of existing psychology

textbook arguments and afforded students the opportunity of learning these arguments either via text or AMs. In Experiment 4, the argument *There was nothing wrong with the United States Supreme Court's decision to uphold the individual's right to bear arms* was the chosen topic as it was assumed that most 1st Year Irish university students would have little knowledge of the topic, thus reducing the chances of students being assisted by any personal pre-existing schemas for such policies. The use of three different topics, two from outside and one within the domain of psychological science, also allowed for an examination of the effect of AM reading and construction on learning across a variety of different topics. More generally, the study topics used across the four experiments were selected because each could be readily presented in the form of a debate. That is, all three arguments possess a central claim that can be both refuted and supported. It was speculated during the construction of these arguments that the nature of debate within these arguments may be more interesting to students than a one-sided (i.e. justification only) argument.

After studying the presented argument, students' memory was tested via a cued recall test. Each memory test consisted of a set of statements (i.e. 7 – 10 statements) from the original study materials. Underneath each statement was either '*Because*', or '*But*', followed by a set of blank lines. Participants were asked to fill in the correct reason or objection to each statement based on the argument previously studied. Memory tests were scored by two independent raters who used a standardized scoring manual to code responses (see Appendix A).

Experiments 1 and 2 also included a comprehension test, which consisted of a set of 12 propositions from the original study material. Each question asked whether the statement was either a *support* or an *objection* to the central claim: *Computers can think*. The inclusion of both comprehension and memory tests proved unwieldy due to

time constraints (i.e. the procedure was administered during a lecture hour) and thus it was decided that Experiments 3 and 4 would assess memory only. All study materials and tests used in Study 1 (i.e. Experiments 1-4) can be found in Appendix B.

In each of the four experiments, the verbal and spatial sub-scales of the *Differential Aptitude Test* (Bennett, Seashore & Wesman, 1986) were administered. The verbal reasoning sub-test of the DAT tests analogical reasoning by providing sentences with missing words (e.g. “ ___ is to night as breakfast is to ___ ”) and five word-pairs to choose from that complete the sentence correctly (e.g. supper — corner; supper — morning; corner — morning; etc.). The spatial reasoning sub-test asks participants to mentally fold cut-outs to produce an object, the correct one being embedded in an array of four choices. Previous research on the DAT revealed that internal consistency reliability coefficients (as assessed by the Kuder-Richardson 20 method) for verbal reasoning are consistently in the .90s, while those for spatial reasoning are consistently in the .80s (Anastasi, 1988; Wang, 1993). Students were provided with 20 minutes to complete both reasoning tests, each of which consisted of 20 items.

Given that the reading and construction of AMs requires the ability to reason both verbally and spatially, both verbal and spatial reasoning ability were assessed as baseline measures and included as covariates in the analysis of group differences, in order to control for any pre-existing differences in ability among groups assigned to different experimental conditions. In addition, as described in Baddeley’s model of working memory (2000), Paivio’s (1986) dual-coding theory and the general framework for thinking (presented in Chapter 1), memory can be seen as consisting of both a visual-spatial coding system and a verbal coding system, which may work collaboratively or independent of one another during perceptual and encoding

processes. As a result, it is possible that students with different levels of verbal and spatial reasoning ability would approach the text reading and AM reading tasks in different ways. For example, when studying an AM, a student with higher verbal reasoning ability and lower spatial reasoning ability may focus on the verbal aspects of the map and not make full use of the spatial arrangement of propositions for the purpose of comprehending arguments or committing them to memory. On the other hand, a student with higher spatial reasoning ability and lower verbal reasoning ability may focus on the spatial arrangement of detail on the map and give less attention to the text. Those with high verbal and spatial ability may have the capacity to focus on both aspects, that is, the verbal detail of the propositions as well as their relational structure in the spatial array of the AM. In this context, given that high levels of both abilities may moderate the relationship between AM and text reading on memory and comprehension outcomes, it is necessary to control for both verbal and spatial reasoning in order to ensure that differences between the AM group and text group are not a result of higher levels of either reasoning ability. It was further hypothesised that higher verbal and spatial reasoning scores would correlate with higher memory and comprehension scores (Hitch & Baddeley, 1976; Jiang, Olson & Chun, 2000; Luck & Vogel, 1997; Wechsler, 1974).

4.2 Experiment 1

Experiment 1 examined the effects of different study material formats (i.e. colour AMs, monochrome AMs and text) and different argument sizes (i.e. 30-proposition vs. 50-proposition) on recall memory and comprehension. In line with the prediction that AM reading offers advantages over text reading (see section 3.6), it was hypothesised that given the same set of arguments to read (i.e. '*Computers can think*'), AM reading would facilitate superior memory and comprehension when

compared with conventional text reading. Given that AM allows for the organisation of information according to the Gestalt grouping principle of similarity (Farrand, Hussain & Hennessy, 2002; Jiang, Olson & Chun, 2000; Luck & Vogel, 1997), a secondary hypothesis was that AMs that contain colour to distinguish ‘reasons’ (i.e. green) from ‘objections’ (i.e. red) within the argument structure, would facilitate better comprehension and recall when compared with monochrome AMs.

It was further hypothesised that those who studied from smaller (30-proposition) argument structures (i.e. both text and AMs) would recall more information than those who studied from larger (50-proposition) argument structures, given that the latter increases the amount of element interactivity within the argument structure, thus increasing levels of intrinsic cognitive load (Pollock, Chandler & Sweller, 2002; Sweller, 2010). However, if facts and relations contained in large, complex argument structures are delivered in manageable chunks (as in AMs; see Chapter 3), it is more likely that the complex structures will be successfully encoded and recalled (Miller, 1956; Pollock, Chandler & Sweller, 2002; Sweller, 1999). On this basis, it was hypothesised that the benefits of AM reading over text reading would be greater for larger argument structures than for smaller argument structures, due to the increased cognitive load associated with the reading of larger text-based arguments and AM’s presumed ability to decrease cognitive load. To test this hypothesis, students read and then answered questions in relation to AMs and homologue texts that contained either 30 or 50 propositions.

4.2.1 Method

4.2.1.1 Design

A 3x2 between-subjects multivariate analysis of covariance (MANCOVA) was used to assess the effects of study materials on comprehension and recall of

arguments, while controlling for baseline verbal and spatial reasoning ability. The first between-subjects factor was study material (text, monochrome AM, AM with colour), while the second was size (30-proposition and 50-proposition argument structures).

4.2.1.2 Participants

Participants were first year psychology students ($N= 400$; 281 females, 119 males), aged between 17 and 25 years, from the National University of Ireland, Galway. In all experiments conducted in Study 1, students were awarded academic course credits in return for their participation. In addition, for purposes of ensuring confidentiality, students in each of the four experiments were identified by student ID number only. The participant consent forms and study information sheets provided to students in Studies 1 - 3 can be found in Appendix C.

4.2.1.3 Materials & Measures

Experimental study materials were constructed by extracting a sub-set of the arguments presented in Robert Horn's (1999) argument map: *Can computers think?* Six sets of study materials were developed, including: (1) a 30-proposition text; (2) a 30-proposition colour argument map; (3) a 30-proposition monochrome argument map; (4) a 50-proposition text; (5) a 50-proposition colour argument map; and (6) a 50-proposition monochrome argument map. All six experimental conditions included a common set of 30 propositions and the three 50-proposition experimental conditions contained an additional 20 propositions that were the same for all three. Regardless of the experimental condition, all participants took the same memory test.

Memory was assessed using a 'fill-in-the-blanks' cued recall test that assessed memory for reasons and objections linked to each of the major arguments supporting

or refuting the central claim: *Computers can think*. The test consisted of 14 items and was scored on a scale of 0-28. Using the intra-class correlation coefficient method, the inter-rater reliability (as measured by two raters) for the memory test was .90.

Comprehension was assessed by asking students to decide if a sub-set of 12 of the original propositions (present in all six study conditions) either supported (i.e. *reason*) or refuted (i.e. *objection*) the central claim.

4.2.1.4 Procedure

Experiment 1 took place over 3 weeks. Each class lasted for 50 minutes. In Week 1, participants were provided with an introductory lecture on critical thinking, in which the concept of AM was introduced and explained. In Week 2, participants were provided with 20 minutes to complete the verbal and spatial reasoning subtests of the DAT. When participants returned to class in Week 3, they were randomly assigned to one of the six study conditions as outlined above. Study materials were distributed. Participants were allotted 10 minutes to read the materials and were instructed to learn the material with a view to being tested. After 10 minutes had elapsed, study materials were collected. The cued-recall test was then administered, with 10 minutes provided for completion, after which tests were collected. Comprehension tests were then administered. Five minutes were provided for completion after which tests sheets were collected and participants were debriefed and thanked.

4.2.2 Results

Preliminary analysis evaluating the homogeneity-of-slopes assumption indicated that in the case of neither verbal nor spatial reasoning did relationships with the dependent variables differ as a function of experimental condition, $F(4, 372) =$

1.40, $p = .232$. Table 4.2 lists means and standard deviations for memory and comprehension test performance for each of the six study conditions.

Table 4.2: Descriptive statistics for Experiment 1

Condition	<u>Small Study Material</u>		<u>Large Study Material</u>	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<u>Comprehension</u>				
Text	7.28	2.12	6.45	2.11
Monochrome Map	6.86	2.28	6.86	2.12
Colour Map	6.47	2.16	6.74	1.94
<u>Memory</u>				
Text	6.36	4.39	2.76	2.85
Monochrome Map	7.60	4.42	4.28	3.14
Colour Map	7.87	4.48	5.13	2.53

With respect to memory, results from the MANCOVA revealed main effects of both study material and argument size, whereas in the case of comprehension, no significant effects were found for either study materials or argument size (see Table 4.3). There were no interaction effects. Post hoc analyses were conducted for all significant main and interaction effects that involved three or more comparisons. Only those effects that were significant after Bonferroni correction are presented below. Post hoc analyses on memory performance revealed a significant difference in favour of both colour argument maps ($t = -3.88$, $df = 225$, $p < .001$, $d = .34$) and monochrome argument maps ($t = -2.63$, $df = 247$, $p = .009$, $d = .28$) over standard text, with no significant difference between monochrome and colour argument maps. Overall, memory was better for those participants who read the smaller (30-proposition) argument structure than those who read the larger (50-proposition)

argument structures. Furthermore, results revealed that memory and comprehension performance were significantly correlated ($r = .31, p < .001$); and verbal reasoning ability was significantly correlated with both memory ($r = .27, p < .001$) and comprehension performance ($r = .35, p < .001$).

Table 4.3: Experiment 1 MANCOVA Summary

	df	df (error)	<i>F</i>	<i>p</i>	Partial η^2
<u>Memory</u>					
Study Material	2	358	7.73	.000	.04
Size	1	358	64.08	.000	.15
Study Material x Size	2	358	.02	.984	.00
<u>Comprehension</u>					
Study Material	2	358	1.07	.346	.01
Size	1	358	1.85	.176	.01
Study Material x Size	2	358	1.11	.333	.01

4.2.3 Discussion of Experiment 1

The findings from Experiment 1 suggest that the provision of a structural representation of the argument for the purpose of reading and learning (i.e. both colour and monochrome AMs) supported better recall performance in comparison with text reading. These results suggest that, when compared with traditional text-based information delivery methods, AM reading may significantly increase subsequent memory for arguments. Furthermore, though there was no significant difference between those who studied from coloured and monochrome AMs, on average, those who studied from colour AMs performed best on average. Notably, it is for this reason that all AMs provided to students from Experiment 2 onwards were presented in colour. Nevertheless, the null effect between the colour and monochrome AM groups does suggest that the benefits of reading AMs over reading text may

depend on the efficacy of AMs to organise information in a hierarchical manner, rather than the inclusion of colour to differentiate propositions. At the same time, when participants were asked to recall the arguments that both supported and refuted specific sub-claims, recall memory was poorer when participants were asked to remember from a larger stimulus set (i.e. a 50-proposition versus a 30-proposition argument structure).

The findings from this experiment also suggest a differential effect of stimulus materials and argument size on comprehension and recall performance. Although the comprehension test was sensitive to individual differences (i.e. no ceiling or basement effects were observed, $M = 6.78$, $SD = 2.14$) and correlated with verbal reasoning ability, the number of arguments a reader was asked to assimilate and the way in which those arguments were presented (AM versus text) had no effect on comprehension test performance.

In relation to comprehension, it may be that those with higher verbal reasoning scores were more likely to engage in some form of critical thinking during the reading of arguments. This would explain the correlation between verbal reasoning ability and both text and AM reading comprehension scores, and it would suggest that a study method that requires active critical thinking (as opposed to passive reading of material) may enhance comprehension (Berkowitz, 1986; Taylor, 1982). The comprehension test required students to decide if a sub-set of 12 of the original propositions either supported or refuted the central claim - *Computers can think*. It was a test that required more than simple memory for individual propositions: it required understanding of the relationships between propositions in the argument structure. Some students found this task surprisingly difficult and scored at chance levels (i.e. they made poor guesses and were correct only around 50% of the time).

Although it is surprising that AM reading did not facilitate performance in this context, it may be that over-and-above baseline reasoning ability and the spontaneous critical thinking efforts of participants during the reading of AMs does not motivate an additional tendency to critically engage with an argument, at least not for novice map readers. Some training in the analysis of arguments – using AMs as study materials – may be necessary for students to engage in the deeper relational analysis of arguments that is necessary for good performance on tests of comprehension.

Nevertheless, in the presence of a similar significant correlation between verbal reasoning and memory performance, and controlling for this co-variation, there remained a significant effect of stimulus materials on memory performance. In other words, although verbal reasoning ability also predicted memory performance, unlike for comprehension test performance, the reading of AMs provided some additional benefits in terms of memory for individual propositions in a cued-recall test. There are a number of possible reasons for this finding. It may be that a certain amount of information can be remembered in the absence of truly understanding the relational structure of an argument (Sweller, 1999), and that AMs facilitate memory for discrete propositions in this context by highlighting each proposition in a distinct box. Text does not demarcate discrete propositions in the same way, chunking them together into larger paragraph units instead. Furthermore, it may be that the benefits for memory of reading AMs differ significantly when novice and expert map readers are compared. For example, it may be that novice map readers will demonstrate some benefit over text readers when memory for discrete propositions is tested, whereas expert map readers will demonstrate some benefit over text readers for both discrete propositional memory and more complex forms of relational memory (as tapped into by comprehension tests, essay-writing tests and ad lib debating exercises). This

hypothesis is consistent with the work by Ericsson and Kintsch (1995) and their description of the function of long-term working memory in expert performance. Experts appear to have the ability to remember complex information structures not by virtue of working memory alone, but by linking information in working memory with knowledge structures in long-term memory. Expertise in argument map reading itself may facilitate the use of schema-driven strategies of chunking and relational mapping of discrete features of an argument map, which in turn may facilitate better subsequent comprehension of the overall argument structure.

Although both basement and ceiling effects were avoided for the memory test, it is clearly the case that participants responded less well when asked to read and recall a large 50-proposition argument. The data suggest that there is an upper limit to the number of arguments that can reasonably be assimilated in a short space of time. These findings suggest that being asked to recall 12 target memories from a set of 50 (after a study period of 10 minutes), is much more difficult than being asked, given a similar study period, to recall the same 12 target memories from a set of 30. To reiterate, this is consistent with the research on Cognitive Load Theory (Pollock, Chandler & Sweller, 2002; Sweller, 1988; 1999; 2010), which suggests that study or work environments that overburden working memory processes (e.g. being asked to memorise 50 propositions in the space of 10 minutes) will be associated with poor learning outcomes.

Though results from this experiment suggest that AM reading can facilitate enhanced recall of smaller, 30-proposition arguments in the context of a 10 minute study period, the results also suggest that the negative impact of intrinsic cognitive load may be greater than AM's beneficial effects. The findings of the current study are also consistent with the body of literature pointing to the superiority of distributed

learning over mass learning (Fiore & Salas, 2007); specifically, when seeking to remember a large quantity of information (e.g. 50-propositions), it is better to distribute the learning over a series of sessions, rather than attempt to learn all the information in one sitting. Future research could compare the effects of distributed learning from large (e.g. 50-proposition) AMs with the effects of distributed learning from large texts; and examine whether the memory benefits of AM reading over text reading transfers from smaller argument structures to large argument structures that are the focus of learning in distributed learning settings.

In summary, these results demonstrate that, when compared with traditional text-based information delivery methods, AM reading significantly increases subsequent memory for arguments, with no performance difference observed on a test of argument comprehension. In addition, participants who read smaller study materials scored significantly higher on memory performance when compared with those who read larger study materials, regardless of format (e.g. AM or text). Though the findings suggest that AM is potentially a useful teaching methodology - specifically in terms of improving memory performance - further research is needed to test the limits of its potential with respect to other important variables including study and testing setting, recall duration (i.e. immediate versus delayed recall) and learning strategy (e.g. silent reading of arguments versus active construction of arguments).

4.3 Experiment 2

Experiment 2 had two aims. The first aim was to replicate Experiment 1, by examining differences between AM reading and text reading using a 30-proposition version of the argument, *Computers can think*. Based on the results of Experiment 1, only colour AMs were used in the AM conditions; and only 30-proposition argument

structures were presented to students, given that the effects of AM were significant for 30-proposition structures and that 50-proposition structures may be too difficult to assimilate in the time provided for study. The second aim of Experiment 2 was to extend the research by examining whether or not memory and comprehension performance varied as a function of study environment, with students examined in both individual and group settings. Research suggests the environment in which an individual studies information may affect how the information is encoded and ultimately how it is retrieved (Godden & Baddeley, 1975; Tulving, 1983; Smith & Vela, 2001).

Given that one important goal of the current set of experiments was to examine the utility of AMs as a pedagogical aid, it was important to examine the difference between classroom use of AMs and the use of AMs as individual study aids. In the current experiment, it was hypothesised that the beneficial effects of AMs would be greater in an individual (i.e. isolated study) setting when compared with a group study setting as potentially less cognitive load would be imposed on students in the individual study condition (e.g. social distractions that may cause attention switching; Pollock, Chandler & Sweller, 2002; Tindall-Ford, Chandler & Sweller, 1997). Specifically, Baron (1986) suggests that the presence of others may act as distracting stimuli, which may cause the switching of attention from the intended focus of attention. Research by Mullen (1987) also suggests that group settings can distract an individual's focus away from the task at hand by focusing attention towards the self (e.g. *How do I look?* or *Is my behaviour acceptable?*). Furthermore, though past research suggests that completion of cognitive tasks in the presence of others often increases motivation, or drive, of those completing the task (Zajonc,

1965; 1980), a meta-analysis by Bond and Titus (1983) found that the presence of others can have a negative effect on performance for more complex tasks.

Based on the findings of Experiment 1, it was also hypothesised that students who use AMs to study would perform better than those who study from text on memory testing. It was further hypothesised that those who study from AMs would perform better on comprehension testing, particularly in the individual study setting. Specifically, because (1) comprehension is a more complex task than memorisation (Anderson & Krathwohl, 2001; Bloom, 1956) and because (2) group settings can negatively affect performance on more complex tasks (Bond & Titus, 1983); it was hypothesised that those in the isolated study group would perform better than those in the group study setting on comprehension.

4.3.1 Method

4.3.1.1 Design

A 2x2 between subjects MANCOVA was used to assess the effects of study materials and environment on recall and comprehension of arguments, while controlling for baseline verbal and spatial reasoning ability. The two between-subjects factors were study materials (text versus AM) and environmental setting (individual versus group setting).

4.3.1.2 Participants

Participants were first year psychology students ($N = 131$; 93 females, 38 males) aged between 17 and 25 years, from the National University of Ireland, Galway.

4.3.1.3 Materials and Measures

Experimental reading materials were the same as those used in Experiment 1 (i.e. arguments that addressed the claim: *Computers can think*). Two sets of study materials were provided to students: a 30-proposition text or a 30-proposition argument map. The arguments in the two conditions were identical other than the form in which they were presented. Recall and comprehension for the argument *Computers Can Think* were assessed by the same tests administered in Experiment 1.

4.3.1.4 Procedure

The study took place over three weeks. Each class lasted for 50 minutes. In Week 1, participants were provided with an introductory lecture on critical thinking, in which the concept of AM was introduced and explained. In Week 2, participants were provided with 20 minutes to complete the verbal and spatial reasoning subtests of the DAT. When participants returned to class in Week 3, they were assigned to one of the four study conditions. Students were allocated to study in either a lecture hall (in the presence of other students completing the same tasks) or alone in an isolated booth. Students assigned to the isolated booth condition were provided with an appointment time in which they were asked to visit the psychology laboratory. They were facilitated upon arrival and provided with the same instructions, study materials and test procedure as those in the group setting. The same experimenter facilitated both group and individual test sessions. After the study environment was allocated, students were then randomly assigned to study the topic *Computers can think* from one of two different study materials: a 30-proposition text or a 30-proposition AM. They were allotted 10 minutes to read their assigned study materials and were instructed to learn the material with a view to being tested. After 10 minutes had elapsed, study materials were collected. The cued-recall test was then administered,

with 10 minutes provided for completion, after which tests were collected. The comprehension test was administered next. Five minutes were provided for completion, after which tests sheets were collected and participants were debriefed and thanked.

4.3.2 Results

Preliminary analysis evaluating the homogeneity-of-slopes assumption indicated that in the case of neither verbal nor spatial reasoning did relationships with the dependent variables differ as a function of experimental condition, $F(8, 230) = .39, p = .926$. Table 4.4 lists means and standard deviations for memory and comprehension test performance for each of the four study conditions.

Table 4.4 Descriptive statistics for Experiment 2

Condition	<u>Memory</u>			<u>Comprehension</u>		
	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>
<u>Lecture Hall Setting</u>						
Text	5.80	3.87	39	7.08	2.25	39
Argument Map	7.97	4.79	37	6.58	2.21	37
<u>Isolated Setting</u>						
Text	4.29	2.64	27	7.44	2.64	27
Argument Map	6.18	3.81	28	6.54	2.63	28

Results from the MANCOVA revealed a significant main effect of study materials (see Table 4.5) on recall, with those who studied from AMs scoring significantly higher on the memory test when compared with those who studied from text, regardless of environmental setting. There was also a significant main effect of environmental setting, with those who studied in the lecture hall setting scoring significantly higher on the memory test than those who studied in the isolated setting,

regardless of study material used. However, there was no study material x environmental setting interaction effect on memory.

Table 4.5: Experiment 2 MANCOVA Summary

	df	df (error)	<i>F</i>	<i>p</i>	Partial η^2
<u>Memory</u>					
Study Material	1	125	11.68	.001	.09
Environmental Setting	1	125	7.04	.009	.05
Study Material x Environmental Setting	1	125	.01	.967	.00
<u>Comprehension</u>					
Study Material	1	125	2.37	.126	.02
Environmental Setting	1	125	.05	.829	.00
Study Material x Environmental Setting	1	125	.02	.890	.00

With respect to comprehension, there was no significant effect of study material or environment setting, and no study material x environmental setting interaction effect. Memory and comprehension performance were significantly correlated with each other ($r = .28, p = .001$); and each was also significantly correlated with verbal reasoning (Memory: $r = .33, p = .015$; Comprehension: $r = .45, p = .001$), while neither was significantly correlated with spatial ability (Memory: $r = .15, p = .260$; Comprehension: $r = .25, p = .068$).

4.3.3 Discussion of Experiment 2

Results from Experiment 2 revealed that those who studied from AMs recalled more information than those who studied from text, regardless of the environmental setting, which is consistent with findings from Experiment 1. The results also revealed

that, contrary to the proposed hypothesis, students who studied and were tested in a lecture hall setting performed better than those who studied and were tested in an individual study booth. In this experiment, participants in the lecture hall setting were monitored to ensure that they did not consult or copy from one another. In the absence of consultation or copying, it may be that the lecture hall setting provided a more formal and stimulating testing venue. It is also possible that participants may have felt more *drive* and responded more competitively in a group setting - in other words, the effect may have been due to some sort of social facilitation (Bond & Titus, 1983; Zajonc, 1965; 1980), which facilitated memory performance even in the context of what was assumed to be a relatively complex argument reading task.

While a group setting may affect participants in particular ways that an isolated setting does not (e.g. distraction, attention switching and social facilitation), such effects were equal across experimental manipulations of materials in the current study. Also, post-hoc analysis revealed that, in comparison with text reading, AM reading was associated with better subsequent memory performance in the group setting and the isolated test setting ($p < .05$ for both comparisons). As such, the beneficial effects of AM reading on memory performance were robust in both an isolated test setting and a group setting, notwithstanding the potential impact of social distraction, social facilitation and other factors linked to group study and testing. In any event, one purpose of this series of experiments was to elucidate the effect of different study techniques in the context of the classroom; and thus, the large group setting has ecological and face validity. As such, all subsequent experiments were conducted in a lecture hall setting only.

Results of the current experiment also revealed that there was an effect of neither study environment, nor AM reading on comprehension performance; a finding

which is consistent with the findings of Experiment 1. Mean scores on the comprehension test suggest that students were performing on average not very far above chance levels (they scored between 6.5 and 7.5 out of 12 on average in a two-choice proposition classification test, as in Experiment 1). In other words, it may be that students did not deeply comprehend the argument and when they were unsure as to whether or not a proposition either supported or refuted the central claim, they simply guessed; and as there were only two answers to choose from, they could guess the correct answer approximately half the time. At the same time, it is notable that students who scored higher on the comprehension test also performed better on the verbal reasoning sub-test of the DAT. Much like in Experiment 1, those high on verbal reasoning ability may have engaged in some form of higher-order metacognitive thinking (e.g. critical thinking) while working to learn from the study materials. Thus, while the findings suggest that the reading of AMs facilitates memory for arguments, it may be that some additional training in AM is needed to foster the kinds of metacognitive skills that may be necessary to support deeper comprehension.

Research suggests that to perform optimally in tests of comprehension and memory, one should read first to understand and then re-read in order to memorise (Pollock, Chandler & Sweller, 2002). However, in the current study, it is possible that because students were aware that they would be tested after they studied, they assimilated information strictly for purposes of memorisation and sacrificed the need to truly comprehend the argument structure. As suggested in relation to Experiment 1, it may also be that a certain amount of information can be remembered in the absence of truly understanding the relational structure of an argument, and that AMs facilitate

memory for discrete propositions in this context by highlighting each proposition in a distinct box.

Given the apparent memorial advantage of AM reading over text reading as shown in the first two experiments, one further question that arises is whether or not this benefit would be sustained over time; that is, from immediate to delayed recall. Long-term memory performance is of course important; not alone for the purpose of acquiring knowledge, but also for deepening comprehension. For example, it may be the case that the deeper comprehension of arguments is not only dependent on extensive reading and learning, but also on sufficient construction of schemas, which are built in order to retain information for relatively long periods of time (see Chapter 1). This schema-building process may allow for connections between arguments encoded at various times to be integrated in different configurations and at different levels of complexity. In order to investigate whether advantages of AM reading versus text reading would be sustained over time, Experiment 3 sought to examine the impact of study materials on immediate and delayed recall for both smaller and larger argument structures.

4.4 Experiment 3

Experiment 3 had three aims. The first aim was to replicate the findings from Experiment 1 by examining differences between reading small (30-proposition) and large (50-proposition) arguments on subsequent recall of select arguments. This was done in order to examine, once again, the hypothesis that the beneficial effects of AM reading would be greater for larger argument structures, due to the increased cognitive load associated with the reading of larger text-based arguments and AM's hypothesised ability to decrease extraneous cognitive load. At the same time, the findings regarding argument size in Experiment 1 (i.e. those who studied from smaller

arguments scored higher on memory testing than those who studied from larger arguments), may have been due to students not having had enough time to assimilate all the propositions within the study material. Therefore, in Experiment 3, all students were provided with five additional minutes to study (i.e. 15 minutes in total).

The second aim of Experiment 3 was to determine whether or not the effects of argument size observed in Experiment 1 were a function of the topic being studied. In Experiment 1, it was hypothesised that the potentially beneficial effects of AM reading over text-based reading would be greater for larger arguments as cognitive load increases (Chandler & Sweller, 1991; Sweller, 1999). However, first year psychology students (i.e. the sample examined) may have been largely unfamiliar with the arguments presented for the topic *Computers can think*; and thus, AM reading in Experiment 1 may have evoked a slower, more analytical reading of arguments that moderated additional beneficial effects of AM reading over text reading, as argument size increased. Thus, the second aim of this experiment was to investigate the effects of AM reading in the context of a different topic. In Experiments 1 and 2, students studied and were tested on the topic *Computers Can Think*. In this experiment, they studied and were tested on the topic *Aggression is Biologically Caused*. It was hypothesised that those who studied the latter topic using AMs would recall significantly more than those who studied the same topic from text (as in Experiments 1 and 2), in the context of both small and large argument structures; and that the benefits of AM would be greater for 50-proposition arguments than 30-proposition arguments (i.e. the benefits of AM would increase as load increases).

The final aim of this experiment was to build upon the previous experiments by examining whether or not the beneficial effects of AM reading extend across time.

It was hypothesised that beneficial effects of AM reading over text reading would be greater for delayed recall than for immediate recall. This was because it was thought that benefits of organisation and dual-coding (i.e. visual and verbal coding) on memory (Mayer, 1997, 2003; Paivio, 1986; 1971) are sustained for longer, and thus the forgetting rate from immediate to delayed recall assessment should be less for AM reading when compared with text reading. This is consistent with past research by Vicari et al. (1999), who found that though the likelihood of forgetting increases over time (i.e. from immediate to delayed recall), the use of organisational strategies as learning aids reduces forgetting. To test the hypothesis that the beneficial effects of AM reading over text reading would be greater for delayed recall than for immediate recall, the effect of AM reading on memory performance was compared with that of text reading both immediately after studying and again a week later (i.e. delayed recall).

4.4.1 Method

4.4.1.1 Design

A 2 x 2 x 2 mixed analysis of covariance (ANCOVA) was used to analyse the effects of both format and size of study material on both immediate and delayed recall, while controlling for baseline verbal and spatial reasoning ability. The two between-subjects factors were study material (text versus AM) and argument size (30-proposition versus 50-proposition); and the within-subjects factor was assessment time (immediately versus one week later).

4.4.1.2 Participants

Participants were first year psychology students aged between 17 and 25 years from the National University of Ireland, Galway. There was some change from

week to week in the students arriving to class and thus, there was a reduction in sample size from immediate to delayed recall assessment. Participants who took part in the delayed component of this experiment ($N = 199$; 142 females, 57 males), also took part in the immediate component, but the sample data available for the immediate component was obtained from a larger sample of students ($N = 286$; 204 females, 82 males).

4.4.1.3 Materials and Measures

In Experiment 3, the study materials presented the topic *Aggression is Biologically Caused*. The topic used in this experiment was different from the two previous experiments and was changed in order to examine whether or not the enhanced performance of students who studied AMs (i.e. in Experiments 1 and 2) was dependent upon the topic studied. Arguments were extracted from core first year psychology textbooks and from secondary sources. Four sets of materials were developed for this topic including: (1) a 30-proposition text, (2) a 50-proposition text (3) a 30-proposition AM and (4) a 50-proposition AM. The information within the four different sets of study materials was identical to the extent that they all contained the same 30 propositions, though the larger sets also contained an additional 20 propositions (i.e. the same ones for both groups). Regardless of the size of the materials presented, all participants took the same memory test.

Memory was assessed using a 'fill-in-the-blanks' cued recall test that assessed memory for reasons and objections linked to each of the major arguments supporting or refuting the central claim, *Aggression is Biologically Caused*. The test consisted of 12 items and was scored on a scale of 0-24. Using the intra-class correlation coefficient method, the inter-rater reliability (as measured by two raters) for the memory test was .95. The delayed recall test was the same as the immediate recall test

except for the order of presentation of questions, which was deliberately changed to minimise practice effects.

4.4.1.4 Procedure

Experiment 3 took place over four weeks. Each class lasted for 50 minutes. In Week 1, participants were provided with an introductory lecture on critical thinking, in which the concept of argument mapping was introduced and explained. In Week 2, participants were provided with 20 minutes to complete the verbal and spatial reasoning subtests of the DAT. In Week 3, students were randomly allocated to one of four groups in which they studied from a 30-proposition AM, a 50-proposition AM, a 30-proposition text, or a 50-proposition text. Study materials were distributed and students were provided with 15 minutes to learn the material with a view to being tested. Students were provided with 15 minutes to study their respective materials in this experiment (i.e. 5 minutes longer than in Experiments 1 and 2), in order to provide those learning from larger argument structures with more time to assimilate the information provided. After 15 minutes had elapsed, study materials were collected and the cued recall test was administered. Students were given 10 minutes to complete this test, after which all the tests were collected and the class ended. In Week 4, students returned and were re-tested on what they had studied the previous week. They were given ten minutes to complete the delayed recall test. After the delayed recall tests were collected, students were debriefed and thanked.

4.4.2 Results

Given the substantial variation in sample size for immediate and delayed recall assessments, three separate ANCOVAs were conducted. Focusing on participants who were present both during immediate and delayed memory assessments, a 2 (study

materials) x 2 (size) x 2 (testing times) Mixed ANCOVA was conducted to examine memory performance differences across groups at two testing times, both immediately after the study period and one week later. Next, two 2 (study materials: text vs. AM) x 2 (size: 30-proposition vs. 50-proposition) ANCOVAs were conducted to examine the effects of study materials and argument size on immediate recall and delayed recall, separately, controlling for verbal and spatial reasoning ability in each case. Means and standard deviations are presented in Table 4.6.

Table 4.6: Descriptive statistics for Experiment 3

Condition	<u>Small Map</u>			<u>Large Map</u>		
	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>
<u>Immediate Recall</u>						
Text	7.76	3.65	63	5.26	3.41	81
Argument Map	10.83	4.41	65	6.47	3.62	77
<u>Delayed Recall</u>						
Text	6.28	3.64	50	3.94	3.37	56
Argument Map	7.58	4.20	52	4.05	2.90	41

In relation to the 2 (study materials) x 2 (size) x 2 (testing times) Mixed ANCOVA, preliminary analysis evaluating homogeneity-of-slopes indicated that for both verbal reasoning and spatial reasoning, the relationship with recall did not differ significantly as a function of experimental condition, $F(4, 186) = .279, p = .891$. Results from the ANCOVA revealed a main effect of study materials (see Table 4.7), with the AM group ($M = 7.20, SE = .34$) scoring significantly higher than the text group ($M = 5.87, SE = .31$). There was a main effect of size, with the smaller argument group ($M = 8.17, SE = .32$) scoring higher than the larger argument group ($M = 4.90, SE = .33$). Though there was no main effect for time, there was a significant time x study material interaction effect, whereby the benefits of AM

reading over text reading were greater for immediate recall ($p < .001$) relative to delayed recall ($p > .05$; see Figure 4.1). No other interaction effects were observed.

Table 4.7: Experiment 3 ANCOVA Summary

	df	df (error)	F	p	Partial η^2
Study Material	1	193	8.29	.004	.04
Size	1	193	51.46	.000	.21
Time	1	193	2.18	.142	.01
Study Material x Time	1	193	10.46	.001	.05

In relation to the first 2 (study materials) x 2 (size) ANCOVA on immediate recall test performance, results revealed a significant main effect of size ($F [1, 280] = 65.67, MSE = 12.21, p < .001, \text{partial } \eta^2 = .19$), with the smaller argument structure group showing significantly better immediate recall performance than the larger argument structure group. Results from the ANCOVA also revealed a significant main effect of study material ($F [1, 280] = 24.17, MSE = 12.21, p < .001, \text{partial } \eta^2 = .08$), with the AM group showing better immediate recall performance than the text group. Results further revealed that performance on the immediate recall test was significantly correlated with both verbal reasoning ($r = .33, p < .001$) and spatial reasoning ability ($r = .24, p < .001$). There was no size x study material interaction effect.

Finally, in the second 2 (study materials) x 2 (size) ANCOVA, conducted on delayed recall test performance, results revealed a significant main effect of size ($F [1, 193] = 39.83, MSE = 11.05, p < .001, \text{partial } \eta^2 = .17$), with the smaller argument structure group performing better than the larger argument structure group on delayed recall. There was no main effect of study material and no size x study material

interaction effect. Delayed recall performance was significantly correlated with both verbal ($r = .34, p < .001$) and spatial ($r = .20, p = .004$) reasoning ability.

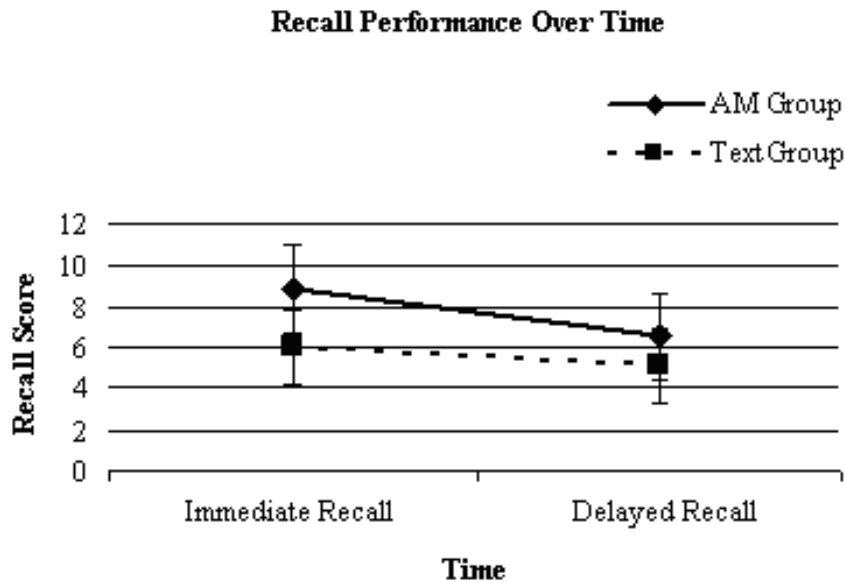


Figure 4.1: Immediate and Delayed Memory performance in text reading and AM reading groups.

4.4.3 Discussion of Experiment 3

The findings from Experiment 3 further confirm that studying from AMs can facilitate better immediate recall of arguments than text reading. Extending Experiments 1 and 2, which presented arguments in relation to the claim *Computers can Think*, the results of the current study, which presented arguments on the nature of aggression, suggest that the beneficial effects of AM reading over text reading is not dependent on the topic being studied. In addition, findings were contrary to the third hypothesis of Experiment 3, in that the superiority of AM with respect to immediate recall was not evident for delayed recall. Although the AM group scored higher on average than the text group on delayed recall, the former also showed a greater decrease in performance from immediate to delayed recall.

Although both basement and ceiling effects were avoided for the memory test, it is clearly the case that participants required to read and recall the larger argument did not perform as well as those required to read and recall the smaller argument. As in Experiment 1, the data suggest that there is a threshold in terms of the number of propositions that can be reasonably assimilated in a short space of time regardless of format or time of assessment. Notably, however, mean recall scores of those who studied larger arguments in Experiment 3 (Text Group: $M = 5.26$; AM Group: $M = 6.47$) were higher than the same groups in Experiment 1 (Text Group: $M = 2.76$; AM Group: $M = 5.13$), a finding which is considered to be the result of the additional five minutes provided to participants to study their arguments in Experiment 3. Findings once again indicate that, consistent with Cognitive Load Theory (Sweller, 1988; 1999; 2010), recalling 12 target memories from a set of 50 is much more difficult than recalling the same 12 target memories from a set of 30. In this respect, the same basic pattern was observed in Experiment 3 as in previous experiments, even though participants all received an additional five minutes of study time.

Findings from Experiment 3 also appeared to confirm the pattern of findings from the first two experiments in which there were significant correlations between verbal reasoning and memory performance (i.e. both immediate and delayed recall); providing further evidence that those who are proficient at verbal reasoning are able to recall more information after studying, even when it has been a week since they studied that information. However, unlike Experiments 1 and 2, there was a significant correlation in Experiment 3 between spatial reasoning and both immediate and delayed recall. The reasons for this difference in outcome are not immediately apparent. The change in topic studied from the first two experiments to the current one may have been a factor. For example, students may have been more familiar with

this topic, more interested in this topic, or possibly perceived it as less difficult to assimilate; thus latent spatial reasoning ability may have emerged as a correlate of recall as a result of the change in topic. However, this is purely speculative given that familiarity, interest and perceived difficulty were not measured in the current set of memory experiments, which is a limitation discussed below. Perhaps, though, the correlation between recall and spatial reasoning, in the context of this experiment, is ultimately a matter for additional empirical research to determine whether familiarity, interest, perceived difficulty or other variables might cause such a difference.

4.5 Experiment 4

The aim of Experiment 4 was to compare the effects of active argument mapping (AM), hierarchical summarisation (i.e. hierarchical outlining), and standard text summarisation on immediate recall performance in the context of the active learning and study of arguments. Previous research has demonstrated the beneficial effects of active study on learning outcomes (e.g. Burbach, Matkin & Fritz, 2004; Hake, 1998; Laws, Sokoloff & Thornton, 1999; Perry et al., 1996; Redish, Saul & Steinberg, 1997). In addition, all three strategies examined in Experiment 4 can be usefully employed to actively study and assimilate text-based arguments (Butchart et al., 2009; Kintsch & van Dijk, 1978; Taylor, 1982; Taylor & Beach, 1984; see Chapter 2). However, it can be argued that AMs offer particular advantages in this regard, given that the active construction of AMs may make the relationships among propositions clearer via its ‘box-and-arrow’ format, thus enhancing the ability to assimilate arguments (van Gelder, 2003). Therefore, it was hypothesised that students who actively construct and study AMs would perform better on immediate recall assessment than those who actively construct and study either outlines or text summaries. It was also hypothesised that students who actively construct and learn

from hierarchical outlines would perform better on immediate recall assessment than those who actively learn using text summarisation techniques. The latter prediction is based on the facts that hierarchical outlines, like AMs, present information in a hierarchically structured manner and that they have been shown to significantly enhance learning in previous research (as discussed in Chapter 2; Taylor, 1982; Taylor & Beach, 1984).

4.5.1 Method

4.5.1.1 Design

A between-subjects ANCOVA was used to analyse the effects of the three active learning techniques on immediate recall. The between-subjects factor was study technique (text summarisation, outlining and AM) while the covariates were baseline verbal and spatial reasoning ability.

4.5.1.2 Participants

Participants were first year psychology students ($N = 136$; 97 females, 39 males), 17 to 25 years old, from the National University of Ireland, Galway.

4.5.1.3 Materials and Measures

All groups were provided with a common study text which contained 30 propositions which argued for and against the claim: *'There was nothing wrong with the United States Supreme Court's decision to uphold the individual's right to bear arms'*. Participants were also provided with materials appropriate to their particular condition: Argument Mapping: a 30-proposition AM with 12 boxes left blank; Hierarchical Outlining: a 30-proposition outline with 12 proposition lines left blank; Text Summarisation: a blank page. The texts provided to those in all three conditions

were identical, but study tools (i.e. AM, Outline, Text summarisation) and instructions were different for each condition.

Memory was assessed using a 'fill-in-the-blanks' cued recall test that assessed memory for reasons and objections linked to each of the major arguments supporting or refuting the central claim: *'There was nothing wrong with the United States Supreme Court's decision to uphold the individual's right to bear arms.'* The test consisted of 12 items and was scored on a scale from 0-24. Using the intra-class correlation coefficient method, the inter-rater reliability (as measured by two raters) for the memory test was .92.

4.5.1.4 Procedure

Experiment 4 took place over three weeks. Each class lasted 50 minutes. In Week 1, participants were provided with an introductory lecture on critical thinking, in which the concept of AM was introduced and explained. In Week 2, participants were provided with 20 minutes to complete the verbal and spatial reasoning subtests of the DAT. In Week 3, participants were randomly allocated to one of three groups in which they studied via one of three active learning techniques: text summarisation, hierarchical outlining or AM. In all three groups, participants were given a 30-proposition text to read, based on the central claim *'There was nothing wrong with the United States Supreme Court's decision to uphold the individual's right to bear arms.'* and were asked to transfer a portion of the information (i.e. 12 propositions) into another document using a particular technique. Participants in the AM condition were given an incomplete AM and were asked to complete the map by transferring propositions as appropriate from the text into the 12 blank boxes on the AM. Participants in the outlining group were given an incomplete hierarchical outline and were asked to complete it by transferring propositions as appropriate from the text

into the 12 blank lines in the hierarchical outline document. In the text summarisation group (i.e. the control group), participants were given a blank sheet of paper and were instructed to take notes as they would naturally and write a summary of the text on the sheet of paper - with a specific focus on the same portion of text that was subject to transfer in the other conditions.

Participants were given five minutes to read the text (which contained 30 propositions) and a further 20 minutes to actively learn using their designated method. After the 25 minutes had elapsed, study materials were collected. The cued-recall test was then administered with 10 minutes allotted for completion. Finally, the completed recall tests were collected and participants were debriefed and thanked. Only those who took part in the active learning portion of the study session (i.e. specifically, those who filled in at least four boxes on the AM, four blank lines on the outline or wrote down four propositions on the blank page) were included in the data analysis. Taking into account the amount of time allowed to complete the task, it was decided that completion of a third or more of the allotted exercise indicated a genuine effort to both read and actively learn.

4.5.2 Results

Preliminary analysis evaluating homogeneity-of-slopes indicated that in the case of both verbal and spatial reasoning, immediate recall did not differ significantly as a function of experimental condition, $F(3, 126) = .729, p = .536$. Results from the ANCOVA revealed a significant main effect of study technique ($F[2,131] = 16.35, MSE = 11.09, p < .001, \text{partial } \eta^2 = .20$), with the AM group ($M = 9.42, SD = 3.77$) scoring significantly higher on memory testing than the text summarisation group ($M = 4.84, SD = 3.25; F[1,131] = 18.27, MSE = 11.09, p < .001, d = 1.30$); and with the outline group ($M = 7.76, SD = 3.68$) also showing significantly higher memory

performance than the text summarisation group ($F [1,131] = 30.68$, $MSE = 11.09$, $p < .001$, $d = .84$). No difference was found between the AM and outline groups on memory performance. Results further revealed that memory performance was significantly correlated with verbal reasoning ($r = .40$, $p < .001$), but not with spatial reasoning ($r = .14$, $p = .098$).

4.5.3 Discussion of Experiment 4

The findings from Experiment 4 showed that both AM and hierarchical outline construction facilitated better immediate recall performance than text summarisation. One reason for the apparent superiority of both AM and outline construction over text summarisation is that in the two former methods, information must be hierarchically organised, which research suggests can improve memory performance (Taylor, 1982; Taylor & Beach, 1984), while in the latter, though participants may have organised information hierarchically, they need not have. For example, participants in the text summarisation condition may have elected to use a summarisation method that did not employ hierarchical organisation; thus, putting them at a potential disadvantage compared with those in the other groups.

It was hypothesised that the spatial organisation of propositions within an AM might facilitate superior recall in comparison with hierarchical outlines. Though the AM group did show better recall on average than the outlining group, the difference between the two groups was non-significant. This pattern suggests that though the spatial organisation of propositions in an AM may benefit memory, the hierarchical organisation of propositions may be the most critical factor influencing subsequent recall.

Finally, consistent with the results from previous experiments, a correlation was found between verbal reasoning and immediate recall. This finding provides

further confirmation of the link between verbal reasoning ability and memory for arguments in classroom learning contexts - a relationship that appears to hold both for more passive reading and more active learning conditions. However, as in Experiments 1 and 2, there was no correlation between recall and spatial reasoning ability, thus suggesting that spatial reasoning ability is not a critical factor determining one's ability to recall specific propositions in an argument, regardless of whether or not AMs were used to study the argument.

4.6 Discussion of Study 1

4.6.1 Interpretation of Results

Results from this series of experiments suggest that learning through AM reading and construction facilitates subsequent recall of propositions in an argument; and that AM studying seems to be superior in this respect to a number of more traditional study formats, including passive text-reading and active text summarisation. In addition, the superiority of AM reading and construction appears to hold across study settings (i.e. isolated study settings and lecture hall settings), colour of study materials and topics studied.

Though these experiments highlight the benefits of AM reading and construction on immediate recall over text-based study methods, as seen in Experiment 3, the relative benefits of AM reading over traditional text-based reading on subsequent recall of arguments was not well maintained over time. In addition, there were no effects of AM on comprehension performance in either Experiment 1 or 2. The null effects of AM on delayed recall and comprehension are appropriate to consider together, since it is possible that a similar process may be at work in both cases. In explanation, (1) delayed recall is essentially a function of LTM, in that the ability to recall information over an extended period of time depends on how well it

was encoded into LTM (Tulving & Thompson, 1973; Tulving, 1984) and (2) research suggests that both LTM and comprehension function as a result of schema construction (Sweller, 2005). Therefore, it is not surprising that AM has similar effects on both LTM and comprehension. In this sense, it is possible that the reason for the null effects of AM reading on both LTM and comprehension was because AM reading did not facilitate enhanced levels of schema construction relative to traditional text-reading.

Results from Experiments 1 and 3 revealed that those who studied small arguments, regardless of format, scored significantly higher on both immediate and delayed recall performance than those who studied large arguments. Given the increased cognitive load associated with learning larger arguments, this finding was not surprising, as research suggests that study environments that overburden memory are associated with poor learning outcomes (Sweller, 1999). In addition, findings from Experiment 3 indicated that studying AMs seemed to provide less of an advantage over other formats in the context of the larger AMs (50-proposition) than in the context of the smaller AMs (30-proposition). This finding contradicted the initial hypothesis, which stated that the relative superiority of AM reading over text reading would be greater under conditions of higher relative to lower cognitive load, as AMs are designed partly to decrease cognitive load. Overall, the data suggest that there is a threshold in terms of the number of propositions that can be reasonably assimilated in a short space of time, regardless of study format. In relation to argument size, the findings are broadly consistent with previous research demonstrating the limited capacity of working memory (Cowan, 2000; Miller, 1956) and how an increase in information processing demands can contribute to cognitive load and poorer overall levels of learning (Pollock, Chandler & Sweller, 2002; Sweller, 1988; 1999).

The purpose of Experiment 4 was to compare argument mapping to the alternative techniques of outlining and text summarisation in a situation where students were required to construct their own study materials as opposed to simply reading and learning from study materials supplied by instructors. Results revealed that students in the AM and outlining groups performed significantly better than those in the text-summarisation group on recall performance, most likely as a result of the active involvement of students in the hierarchical organisation of information that was the focus of learning (e.g. Taylor, 1982; Taylor & Beach, 1984).

As previously discussed, a growing body of literature suggests that organisational strategies can be used to facilitate recall performance and learning in general (e.g. Berkowitz, 1986; Meyer, Brandt & Bluth, 1980; Myers, 1974; Oliver, 2009; Robinson & Kiewra, 1995; Taylor, 1982; Taylor & Beach, 1984). Importantly, the organizational structure of text may not always facilitate learning and memory, given that the linear nature of text does not allow one to readily link propositions that support or dispute specific claims (van Gelder, 2003). Overall, the superiority of AM studying over text-based studying in the current research may be attributed to the hierarchical structuring of information in AMs. Specifically, results revealed that the removal of colour from AMs in Experiment 1 had no significant effect on the AMs' ability to facilitate recall and any group that studied from hierarchically organised study materials (e.g. AM reading, AM construction and active hierarchical outlining) recalled significantly more than groups which did not have hierarchically organised materials. Again, this finding is consistent with previous research by Taylor (1982) and Taylor and Beach (1984) who found that the use of hierarchical summarisation increased recall performance of those who used the technique.

The correlation between memory performance and both verbal and spatial reasoning ability was measured in all four experiments and each time, a significant correlation was found between verbal reasoning ability and recall performance (i.e. both immediate and delayed). It seems possible that verbal reasoning ability and recall performance are correlated because those who are proficient at verbal reasoning are able to retain information long enough in order to assimilate the reasoning within a proposition or a set of propositions, which in turn facilitates their ability to answer subsequent test questions correctly (Colom et al., 2005; Hitch & Baddeley, 1976). This interpretation is consistent with the findings from Experiments 1 and 2, in which verbal reasoning was also found to be correlated with comprehension. Spatial reasoning was only found to be correlated with recall (i.e. both immediate and delayed) in Experiment 3. This suggests that spatial reasoning ability may influence memory for arguments in specific learning contexts (e.g. dependent on topic studied), but that verbal reasoning ability is a more robust and consistent predictor of memory in these situations.

Though perhaps unexpected, the absence of a significant correlation between memory performance and spatial reasoning for the majority of these experiments may be a result of the dominant verbal nature of the learning task used in each experiment and the fact that, according to Hegarty (2004), there may be a strong dissociation between spatial and verbal reasoning. Alternatively, as suggested in Experiment 3, the relationship between memory and spatial reasoning may simply come down to the topic studied, in that students' familiarity, interest or perceived difficulty in assimilating certain topics may have affected their ability to draw upon their spatial reasoning ability as a means by which to facilitate memory consolidation. In any case, the key reason that the spatial reasoning assessment was administered in this research

was in order to control for baseline ability (i.e. as it was hypothesised that those better at spatial reasoning, may be better suited to study from AMs than from text).

However, it seems reasonable to conclude, in light of findings, that students' memory for arguments can be enhanced via the reading of AMs and that these benefits of AM are observed even when statistically controlling for verbal and spatial reasoning ability.

4.6.2 Limitations & Future Research

Though the results of these experiments are informative in terms of the comparison between AM and alternative learning techniques, there are some limitations to the research that must be considered. One possible limitation is that there are student characteristics other than verbal and spatial reasoning ability that may have influenced learning and may have been usefully included in the analysis, such as motivation or interest in study topics and learning methods, for example. In retrospect, it is reasonable to assume those with higher levels of motivation to learn or to out-perform others, or those who found certain study methods novel and interesting (e.g. argument mapping or active learning), may have been more motivated to succeed and thus, may have outperformed others not similarly motivated or interested (Marzano, 1998; Pintrich et al., 1991). For example, the lack of an effect of AM reading on comprehension in Experiments 1 and 2 may have been caused by a lack of motivation to engage in higher-order thinking processes (e.g. critical thinking) that may, as hypothesised above, be critical drivers of comprehension in novel learning environments. In addition, participants who were interested in new learning techniques – even by the slightest manipulation to the technique, for example, colour in an AM, may have been further motivated by this novelty to perform better. In other words, in addition to a possible main effect of motivation on performance (Garcia,

Pintrich & Paul, 1992; Pintrich et al., 1991), it is possible that there was an unobserved interaction effect between motivation and novelty of learning materials across experiments. Having some knowledge of students' motivation may also have shed light on the attrition of students from testing time 1 to 2 in Experiment 3. Thus, the differential impact of instructional techniques on participants who were highly motivated should have been compared with the effects on those participants who were poorly motivated.

Another possible limitation that should be considered is that methods for gauging the difficulty of the topic studied, or students' familiarity with or interest in the topic, were not used. It may have been that memory performance was not a result of recall ability alone, but could also be a result of interest or familiarity with the topic studied (Taylor & Beach, 1984). Three different topics were used in this set of experiments and yielded similar patterns of findings. Nevertheless, participant ratings of familiarity, difficulty or intrinsic interest could yield potentially useful, additional information; perhaps especially with respect to unexpected patterns of findings such as the lack of an effect for comprehension in Experiments 1 and 2, or idiosyncratic patterns such as that seen for spatial reasoning in Experiment 3.

For example, the topic *Computers Can Think* (presented to students in Experiments 1 and 2) may have been perceived by students as difficult to assimilate, due to the language used within the argument. The study materials used in these two experiments were developed by adapting Robert Horn's (1999) AM, *Can Computers Think?*, into smaller AMs and corresponding texts. However, following these adaptations, it is possible that the actual prose within the resulting materials lacked some element of clarity and/or precision, thus making it difficult for students to assimilate. Again, this potential difficulty could explain the null effect of AM on

comprehension. However, it is also worth considering that the prose within Horn's (1999) original AM may have lacked some element of clarity prior to its' adaptation into the various study materials used in Experiments 1 and 2. At the same time, if in the event that the prose within the study materials did indeed lack clarity and/or precision, such deficits would have been equal across conditions, given that each study material consisted of the same prose and the same argument structure.

Notably, though student's motivation, interest and familiarity with the topics studied in Experiments 1-4 would all have been useful, one general challenge in these experiments was the issue of working within the allotted class time of 50 minutes. This time constraint made the inclusion of multiple assessments difficult for the efficient running of experiments with large numbers of students. Nevertheless, future research on AM's effect on key aspects of performance, such as memory and comprehension, needs to address some of the limitations discussed thus far, including motivation to learn, familiarity with the topic studied, perceived difficulty of the topic studied and interest in the topic studied.

The four experiments conducted in Study 1 aimed to compare AM and traditional presentation formats. However, one difficulty in making a fair comparison between these formats is that participants have been exposed to traditional formats such as text reading, all their lives, whereas they were new to AMs. Though the training provided in these experiments (i.e. a 50 minute lecture), may have been enough for participants to perform well in terms of studying from and constructing semi-completed AMs, it may not have been enough for the full potential of the advantageous effects of AM reading and construction on delayed recall and comprehension to be realised. Alternatively, it is possible that the 50-minute lecture used to introduce and explain AM might have primed students to favour the AM

strategy - possibly accounting for some latent effect of AM on immediate recall performance. Nevertheless, if the AM training provided was insufficient with respect to the reading of AMs, then this can be seen as a possible explanation for the null effects on comprehension performance (Experiments 1 & 2) and delayed recall performance (Experiment 3).

Furthermore, though the aim of the current study was to examine the effects of AM on both memory and comprehension, in comparison with traditional, text-based learning strategies; this is not to suggest that AM and text-based learning strategies are incompatible alternatives. For example, it is possible that the combined use of AM alongside text-based study can also facilitate learning. This recommendation is supported by findings from Experiment 4, in which the passive reading of text supported by active AM construction facilitated immediate recall beyond that of passive text-reading supported by text-summarisation. Thus, future research should aim to examine the effects of AM as an educational aid to text-based study, rather than strictly as an alternative. However, given the aims of this research, focus now returns to comparing the two learning strategies.

In order for a more equal comparison between AM and text-based materials to be made, participants would likely require more substantial training in the use of AM than that provided in these experiments. This view is consistent with van Gelder, Bissett and Cumming's (2004) view on the potential beneficial effects of AM on critical thinking; that is, in terms of achieving optimal growth in critical thinking ability, one must also consider what level of 'deliberate practice' is needed. Naturally, it is not simply the method or tools of instruction and learning that are important when working to cultivate critical thinking ability, but also the intensity and quality of practice. The question arises in this context as to what level of AM

familiarity and practice is needed to facilitate optimal memory and comprehension performance when using AM reading or construction as a classroom exercise.

Guidelines on what constitutes sufficient training in AM, in the context of critical thinking training studies, have been provided by van Gelder, Bissett and Cumming (2004) and suggest a semester-long course in critical thinking taught through argument mapping for purposes of improving critical thinking ability. Though it was not an aim of Study 1 to examine argument mapping's effect on critical thinking, it is hypothesized, based on van Gelder and colleagues' research, that any semester-long course that teaches a subject through AMs will adequately teach students how to use argument mapping as a study method to facilitate memory and comprehension ability. As a result, an aim for future research should be to examine the effects of explicit training in AM and the related method of hierarchical summarisation (i.e. outlining) on both lower-order learning outcomes (memory and comprehension) and higher-order learning outcomes (analysis, evaluation, inference, and reflective judgment).

Advancing upon the findings of Study 1 and also ascending Bloom's hierarchical taxonomy of thinking processes, Study 2 examined the effects of AM training and hierarchical outlining training on analysis, evaluation, inference and reflective judgment ability.

Chapter 5

Study 2: The Effects of Argument Mapping on Critical Thinking & Reflective Judgment Performance

5.1 Purpose

This chapter presents the empirical research carried out in Study 2, which compared the effects of a six-week critical thinking (CT) training course taught through argument mapping (AM) with both a traditional CT training course taught through hierarchical outlining (HO) and a no-CT training control condition on CT performance and CT sub-skill performance (i.e. analysis, evaluation, inference and reflective judgment)¹⁴. Both theory and previous research suggest that AM may be a useful learning aid, particularly when learning critical thinking skills (e.g. Butchart et al., 2009; van Gelder, 2001, van Gelder Bisset & Cumming, 2004). AM is thought to be superior to other strategies in this context, due to its useful organizational features that may help students to assimilate, analyse, and evaluate information (see Chapters 3 & 4). Thus, the first hypothesis of the current study was that those trained in CT through AM would outperform those in both the HO and control groups on overall CT, all CT sub-skills and reflective judgment (RJ). Though the results from Study 1 indicated that AM is no better than HO at enhancing immediate recall, the level of AM training received by students in Study 1 was limited and it was noted that some AM benefits may only emerge after more extensive training. Thus, effects of AM may be greater for CT outcome measures given the intervention of a six-week AM-infused CT training course.

¹⁴ Portions of Study 2 have been published and are cited in the References section (see Dwyer, Hogan & Stewart, 2011a).

Furthermore, in comparison with AM, HO is essentially a text-based learning strategy in which the structure of an argument is represented as a linear flow of text without the full complement of organisational advantages possessed by AM (e.g. 'box-and-arrow' structuring and colour cues to represent reasons, objections, and rebuttals). As such, the HO critical thinking training group was included as a comparison group in this study because it allowed for strict experimental control and comparison in the evaluation of the potential benefits of AM on CT outcome measures. Furthermore, given that third-level courses are generally taught using hierarchically structured, bullet-point styled slideshows, a HO condition also served as an ecologically valid comparison group.

Nevertheless, like AM, propositions within an HO are hierarchically organised and can be explicitly signalled in accordance with their inferential relationships with other propositions. In addition, the active construction of an HO involves the extraction and summarisation of key arguments and sub-arguments in a text. This active reconstruction and hierarchical ordering of arguments has been shown to benefit student learning (Taylor, 1982; Taylor & Beach, 1984). Thus, the second hypothesis of the current study was that those trained in CT through HO would outperform those in the control group on overall CT, all CT sub-skills and RJ. This hypothesis is also consistent with research which suggests that CT training in general, regardless of the format in which training is presented (e.g. AM, HO, prose-based text), can enhance CT ability (e.g. Hitchcock, 2004; Rimiene, 2002; Solon, 2007).

Notwithstanding the unique features of AM and the potential additional benefits of AM training over HO training, it is possible that AM and HO have many similar benefits when used as learning aids in the context of CT training. Given that reasoning toward a conclusion often requires the combined use of many different

thinking processes (e.g. memory, comprehension, analysis, evaluation, inference and reflective judgment; Facione, 1990b; Halpern, 2003a; Reeves, 1990), both AM and HO may be particularly useful in the context of CT training because their organisational features can help to decrease the demands associated with *extraneous* cognitive load while simultaneously enhancing various key cognitive processes (e.g. awareness of argument structure for analysis and coordinated evaluation of propositions marked as supports and rebuttals). In addition, CT training using AM and HO learning materials as supports may facilitate the growth of CT schemas, or strategies of analysis, evaluation and inference, that ultimately reduce the negative learning impact of intrinsic cognitive load when working with complex problems. As noted in Chapter 1, research suggests that efforts to promote schema-construction through some form of training can aid in the reduction of *intrinsic* load (Pollock, Chandler & Sweller, 2002; van Merriënboer, Kirschner & Kester, 2003), because those who possess relevant schemata, or some level of expertise in a knowledge or skill domain, are better equipped to assimilate information with high element interactivity than are those who do not possess the relevant knowledge or skills (Pollock, Chandler & Sweller, 2002; Sweller, 2010). Specifically, training in CT (regardless of whether it is trained through AM or HO) may provide students with the opportunity to develop the CT schemas necessary to decrease the intrinsic cognitive load associated with the application of CT in situations of high element interactivity and also facilitate enhanced CT performance. As such, CT instruction and exercises, supported by AM or HO, may decrease both extraneous and intrinsic cognitive load and facilitate the development of CT skills. Therefore, the third hypothesis tested in the current study was that those who attended either an AM- or HO-infused CT

training course would perform better on CT, all CT sub-skills and RJ as a combined group, than those in the control group.

A further aim of this research was to examine the effect of CT training on a variety of dispositional factors. Like CT, there are many conceptualisations of disposition towards thinking. According to Valenzuela, Nieto and Saiz (2011), some conceptualisations focus on the attitudinal and intellectual habits of disposition, whereas others focus on epistemological beliefs (see also Phan, 2008; Schommer-Aikins, 2002). Using both the California Critical Thinking Dispositions Inventory (Facione & Facione, 1992) and the Lectical Reflective Judgment Assessment (Dawson, 2008b), the current study measured the dispositions of truth seeking, open mindedness, analyticity, systematicity, confidence, inquisitiveness and maturity; as well as epistemological beliefs regarding the malleability, structure and stability of knowledge.

According to the *Delphi Report* (Facione, 1990b; see Chapter 1), good critical thinkers possess positive dispositions towards thinking. In addition, it has been argued that the *disposition* to think critically is as important to CT as is the ability to perform CT *skills* (Ennis, 1996; Halpern, 2003a; Perkins & Ritchhart, 2004). For example, though an individual may be aware of which CT skills to use in a given context and may have the capacity to perform well when using these skills, they may not be disposed to use them. Conversely, an individual may be prepared and willing to use CT skills, but may not know how to do so. In both contexts, it is unlikely that CT will be applied well (Valenzuela, Nieto and Saiz, 2011). Thus, along with the ability to engage in CT skills, “a critical thinker must also have a strong intention to recognise the importance of good thinking and have the initiative to seek better judgment” (Ku, 2009, p. 71). In other words, the combination of the ability to use CT skills and

possessing the disposition to apply these skills together determine a person's actual thinking performance (Ennis, 1998; Facione et al., 2002; Halpern, 2003a, 2006; Ku & Ho, 2010a).

Past research has demonstrated a significant relationship between CT dispositions and CT ability (Colucciello, 1997; Facione, Facione and Sanchez, 1994; Facione, 2000; Profeto-McGrath, 2003), as well as significant increases in both CT dispositions and CT ability as a result of CT training (Rimiene, 2002). For example, Colucciello (1997) examined the relationship between dispositions towards thinking and CT ability in a sample of nursing students of varying academic levels (i.e. sophomore, juniors and seniors) and found that CT performance was significantly correlated with truth seeking, open mindedness, analyticity, systematicity, confidence, inquisitiveness and maturity, as measured by the California Critical Thinking Dispositions Inventory (CCTDI; Facione & Facione, 1992). Profeto-McGrath (2003) similarly found a significant correlation between overall CT ability and self-reported positive dispositions towards thinking. This relationship was further corroborated in research studies by Facione (2000) and Facione, Facione and Sanchez (1994). The current study sought to confirm these findings by examining the relationship between the CCTDI measures of truth seeking, open mindedness, analyticity, systematicity, confidence, inquisitiveness and maturity and CT ability, as measured by the CCTST (Facione, 1990a); and differences among AM, HO and control groups on these same dispositions at post-testing.

In the current study, another dispositional factor, epistemological beliefs, was also examined. In the context of the current study, epistemological beliefs refer to beliefs regarding: the *malleability* of knowledge (i.e. 'the belief that knowledge is fixed at birth' or 'the belief that the ability to learn can be improved'); the *structure* of

knowledge (i.e. the ‘belief that knowledge is best characterized as complex interrelated networks of facts and relations’); and the *stability* of knowledge (i.e. ‘the belief that knowledge is evolving rather than unchanging’). Though longitudinal research suggests that beliefs in relation to the malleability, complex structure and evolving nature of knowledge increase over time, reflecting more mature, complex beliefs about the nature of knowledge (Schommer et al., 1997); to this researcher’s knowledge, no studies have explicitly examined the effect of CT training on epistemological beliefs. Research does suggest that beliefs in relation to the malleability, complex structure, and evolving nature of knowledge are good predictors of academic performance and reflective thinking ability (Cano, 2005; Kitchener & King, 1994; Phan, 2008; Schommer-Aikins, 2002), however, it is unclear if CT interventions have any effect on these beliefs.

Based on the previous research on disposition towards thinking and epistemological beliefs, it was hypothesised in the current study that CT ability would be significantly, positively correlated with both dispositions towards thinking (i.e. truth seeking, open mindedness, analyticity, systematicity, confidence, inquisitiveness and maturity), as well as epistemological beliefs regarding the malleability, complex structure and evolving nature of knowledge. It was also hypothesised that those who took part in CT training (i.e. both the AM and HO groups) would score significantly higher than controls on these same measures at post-testing.

5.2 Method

5.2.1 Design

A series of six one-way ANCOVAs were used to compare the argument mapping (AM), hierarchical outlining (HO) and control groups on overall CT, analysis, evaluation, inference, inductive reasoning and deductive reasoning ability at

post-test¹⁵, while controlling for the corresponding ability at baseline. Gain was not measured from pre-to post-testing; rather, in accordance with research and recommendations of Jacobs (1995) regarding the California Critical Thinking Skills Test (see *Materials and Measures*) Form B, performance was used as both a baseline measure and covariate, while Form 2000 was analysed as the dependent variable. Four 3 (condition: AM, HO and control groups) x 3 (time: pre-testing, post-testing and six-month follow-up testing) Mixed ANOVAs were conducted to examine the effects of experimental conditions on reflective judgment (RJ) and epistemological beliefs (i.e. malleability, structure, and stability of knowledge). A further series of eight ANOVAs was used to examine and compare the effects of the three conditions on overall disposition scores and the CT dispositions of truth seeking, open mindedness, analyticity, systematicity, confidence, inquisitiveness and maturity at post-test.

Formal power analysis was completed using Cohen's (1988) guidelines for small ($r = .1$), medium ($r = .3$), and large ($r = .5$) effects. Two-tailed alpha of .05 was assumed for all tests. Previous similar CT intervention research has shown medium to large effect sizes for core outcome variables analysis, evaluation, and inference (Alvarez-Ortiz, 2007; Hitchcock, 2004; Rimiene, 2002). Sample size calculations assumed medium effect sizes for ANOVA and sought to provide power greater than 0.8. Calculations, using a conservative medium effect size (.25), revealed that this would require approximately 40 participants completing the research in each group.

¹⁵ Though not identified in Bloom's taxonomy (1956), inductive and deductive reasoning are included on the California Critical Thinking Skills Test (Facione, 1990; Facione et al., 2002), a measure which was used in this study (see *Materials and Measures*).

5.2.2 Participants

Participants were first year psychology students ($N = 81$; 57 females, 24 males), aged between 18 and 25 years, from the National University of Ireland, Galway. The initial pre-test sample size ($N=134$) was decreased by an attrition rate of 40%. In the six-month follow-up portion of this study, the sample size was further reduced ($N= 40$; secondary attrition rate of 51%). Participants who withdrew from the course were administered a questionnaire which asked why they did not complete the study. Reasons for attrition included having conflicting schedules, being too busy with other courses and other personal reasons. In return for their participation, students were awarded academic course credits.

5.2.3 Materials and Measures

The CT intervention materials included DVD recordings of the CT courses (i.e. lectures and exercises), exercise handouts, a laptop computer and a projector. Exercise handouts and DVD recordings of the lectures and exercises can be found in Appendix D. The intervention materials used in Study 2 were developed: (1) based on a review of previous CT intervention studies and CT course designs (see Chapters 2 and 3); (2) under the direction of both primary supervisor and co-supervisor; (3) with a specific focus on sequentially presenting the skills of analysis, evaluation and inference (i.e. given that these are the core CT skills identified by the *Delphi Report* and assessed by the California Critical Thinking Skills Test); and (4) with a focus on moving from simple to more complex CT exercises.

The CT course used a mixed approach of *didactic instruction* and *active learning*. According to Mayer (2004), students must be cognitively active during learning and educators must provide students with guided practice. This type of active learning provides students with a form of *scaffolding* (Wood, Bruner & Ross, 1976),

whereby students are guided didactically by their educator and also actively *learn by doing*. Research suggests that people learn more through active learning (e.g. Hake, 1998; Laws, Sokoloff & Thornton, 1999; Redish, Saul & Steinberg, 1997), and more specifically, that intensive practice of CT skills increases critical thinking ability more so than didactic teaching of CT (Burbach, Matkin & Fritz, 2004).

As the quality of learning materials was thought to be a vital factor in facilitating enhanced CT ability, all lectures and exercises were piloted prior to using them in Study 2. Feedback on lectures and exercises was provided by students in two separate focus group sessions (i.e. one for the HO materials [$N = 5$] and one for the AM materials [$N = 6$]). On whole, students found the CT lectures and exercises to be “*simple, and interesting*”; “*straightforward*”, “*explained very well*”, “*engaging*” and “*easy to relate to*”. The students also reported that the CT courses appropriately revealed how to “*form*”, “*dissect*” and “*break down arguments*” and it was suggested that the course would “*aid in their undertaking of other academic courses*”. Students also noted that “*loads of time*” was provided for the completion of exercises. The students reported that they neither “*struggled to pay attention*”, nor found the course “*confusing*”. Regarding the video presentation of the course, students found the materials were “*very clearly presented*”, “*appropriate for the setting*” and allowed for “*good discussion*”. A number of the students reported that the audio/visual features of the pre-recorded lectures were “*great*”. Only one student found the course, in general, “*boring*”.

Students also suggested that exercises required a “*shorter amount of time to complete than what was provided*” to them; and that “*sometimes, too much time was given for each exercise... so time was sometimes wasted.*” To rectify this issue for the CT course in this study, a classroom facilitator (i.e. the researcher) walked around the

room and observed the progress of students during the completion of their exercises. When all students appeared to have finished a given exercise, the classroom facilitator would ask whether or not anyone needed more time. On no occasion did the time allowed for completion of exercises surpass the time permitted by the voiceover in the recording (i.e. the time provided to students during pilot testing). Students in the pilot testing focus group also suggested that the lecture sessions “*needed more, shorter, breaks*”, that is, more than just a 10 minute break provided after completion of the first hour of class. To rectify this issue for the CT course in this study, classroom facilitators provided students not only with a 10 minute break after completion of the first hour of the lecture, but also an additional two minute break after each class exercise. During the 10 minute break, students were permitted to leave the room and do as they wished. During the two minute breaks, students often remained seated and spoke amongst themselves.

The *California Critical Thinking Skills Test* (CCTST; Facione, 1990a; Facione et al., 2002) was used to measure critical thinking ability. The test consists of 34 multiple choice questions, which assess overall CT ability as well as five sub-skills: analysis, evaluation and inference (which are the core CT skills as identified by the *Delphi Report*), as well as inductive and deductive reasoning (see Figure 5.1 for examples of inductive and deductive reasoning problems). The CCTST Form B (Facione, 1990a) was administered at pre-testing. Kuder-Richardson 20 internal reliability coefficients for Form B range from 0.78 – 0.84 (Facione, 1991). The CCTST (Form 2000) was administered at post-testing. Kuder-Richardson 20 internal reliability coefficients for Form 2000 range from .74 - .84 (Facione et al., 2002).

The *Lectical Reflective Judgment Assessment* (LRJA; Dawson, 2008b), a cognitive development assessment of RJ, was administered as a baseline measure at

pre-test, as an outcome measure at post-test, and again six months later as a follow-up assessment. It consists of a series of seven open-ended questions about real-world problems. Responses to these questions are scored by Certified Lectical Analysts who (1) identify *scorable units of text* (i.e. an argument or related set of propositions) and (2) examine the structure of both the unit and the conceptual elements within by (3) mapping the responses for purposes of assessing the hierarchical complexity of the RJ performance (see Figure 5.2). These Certified Lectical Analysts maintain an inter-rater reliability that is at or above 85% agreement, while internal consistency is typically above .90 (Dawson-Tunik, Commons, Wilson, & Fischer, 2005).

Inductive Reasoning	Deductive Reasoning
<p>The teams in the city's youth recreational soccer program are meant to be evenly matched. Yet some teams are a little better than others. Suppose that last Saturday a team called the Sparklers defeated one called the Wildflowers. Suppose that the previous Saturday the Wildflowers had defeated a team called the Mustangs. What is likely to happen next Saturday when the Sparklers play against the Mustangs?</p> <p>A = The Sparklers will certainly win.</p> <p>B = The Sparklers will probably win, but might lose.</p> <p>C = The Sparklers will probably lose, but might win.</p> <p>D = The Sparklers will certainly lose.</p> <p>E = The soccer game will end in a tie.</p>	<p>Consider this argument: "Person L is shorter than person X. Person Y is shorter than person L, but person M is shorter than person Y. Therefore, person Y is shorter than person J." What information must be added to require that the conclusion is true, assuming all the premises are true?</p> <p>A = Person L is taller than J.</p> <p>B = Person X is taller than J.</p> <p>C = Person J is taller than L.</p> <p>D = Person J is taller than M.</p>

Figure 5.1: Inductive and Deductive Reasoning problems on the CCTST Form 2000 (adapted from Facione et al., 2002)

The score on the LRJA (e.g. 10:3) indicates the level of RJ ability of the test-taker (see Table 5.1). Specifically, in a score of 10:3, the first number (i.e. 10) represents the test-taker's developmental stage level (i.e. abstract mappings), while the second number (i.e. 3), which can range from 1-3, represents how far into that

level the individual has developed (i.e. how well the individual is able to integrate and coordinate the abstract mappings presented in their response to open-ended questions). Thus, at the abstract mappings level, a response may receive a score of 10:1 (i.e. the response presents a series of abstract mappings but does not integrate them), 10:2 (i.e. the response coordinates a series of abstract mappings, but does not elaborate on this integration) or 10:3 (i.e. the response coordinates a series of abstract mappings and elaborates upon this integration at a level of performance that approaches a systems level of thinking, or 11:1 thinking).

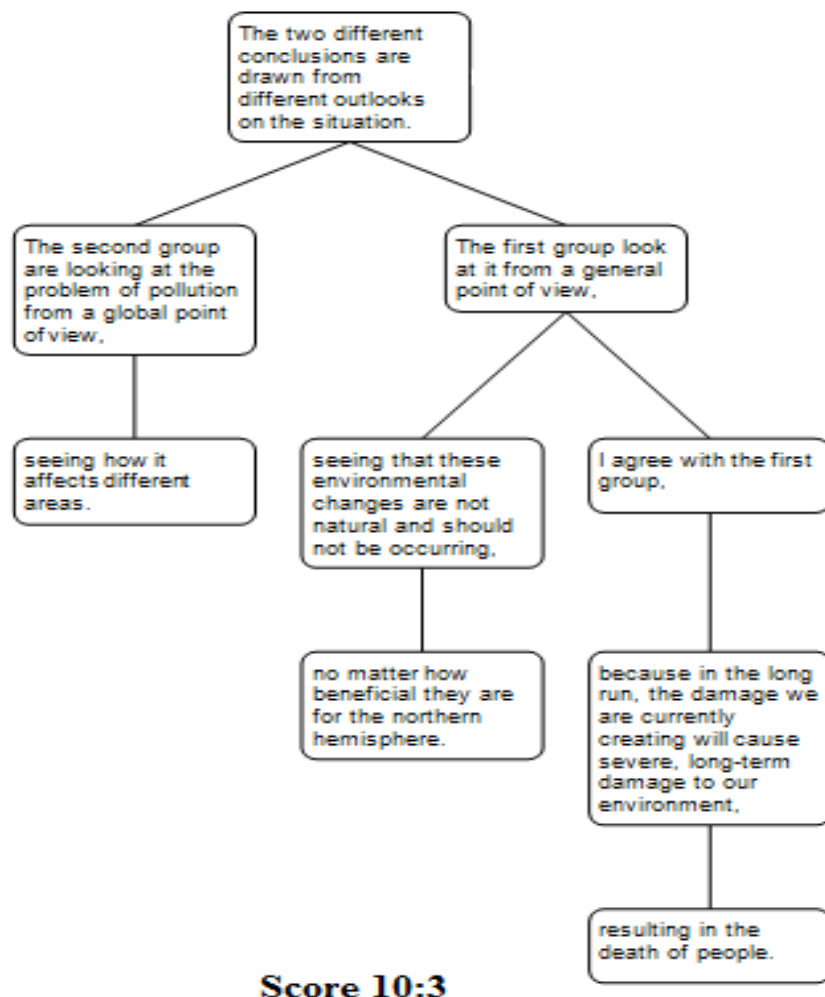


Figure 5.2: Example of how a response to an LRJA item might be mapped for purposes of scoring

Table 5.1: Example of a concept as it changes across the complexity orders (adapted from Hogan & Stein, 2010)

Level	Complexity Order	A Person is...	Abstraction	Structure
6	Single representations	My mom?	1 st order representation	
7	Representational mappings	A person cooks food a lot, and goes to his job. Like my Dad. My teacher is a person, and I am a person, but I'm a kid too.	2 nd order representation	Mapping: Person is a mapping of several 1 st order representations.
8	Representational systems	I'm a person you are a person. A person does things. Like talk to friends and eat to stay alive so they can have fun.	3 rd order representation	System: Person is a system of 2 nd order representations with two variables on the input.
9	Single abstractions	A person is a human being. Human beings live all over the world; they all know how to think except little babies. I heard humans used to be monkeys but then we got smarter. People are intelligent. They have feelings and ideas.	1 st order abstraction	Definitional: Person is defined as a 1 st order abstraction, which is a conceptual integration of representational systems.
10	Abstract mappings	A person uses reason; they are animals, but they are different from animals because they think about things better. A person has responsibilities, in life. You know, a person <i>is</i> somebody, with a personality of his own. It's hard to get to know somebody because the real person is on the inside. A person has a soul.	2 nd order abstraction	Mapping: Person is a mapping of several 1 st order abstractions.
11	Abstract systems	A person is like a whole world in themselves. So, you have to respect the unique emotional temperaments, life circumstances and perspectives of everyone. And you can't separate one of those aspects from the others because they all interact with each other. That's why people are so complex. That's why we need to be conscious of individuality.	3 rd order abstraction	System: Person is a system of 2 nd order abstractions. Here several variables which are at least 2 nd order abstractions are coordinated on the input.

Table 5.1: Example of a concept as it changes across the complexity orders (continued)

Level	Complexity Order	A Person is...	Abstraction	Structure
12	Single principle	The concept of <i>person</i> , as I define it, can be used to coordinate what often seems like a rift between an individual's unique system of meaning and a society's complex web of structures and institutions. The concept of 'person' is a way of thinking about the mutual-dependence between these two systems. It is in the 'person' that the citizen, (as defined by a system of rights, liberties, social roles etc.) and the individual, (as defined by system of unique choices, motivations and meanings), meet and embody their interdependence. You can not really have one system with out the other. Only in the social space provided to the citizen can an individual flourish, understand him or herself, and develop. And yet, only with the unique development of motivations and meanings can individuals fulfill their roles as citizens. So I understand 'persons' in this way, as they embody the fragile interdependence of the mind and society.	1 st order principles	Definitional: Person is defined as a 1 st order principle, which is a conceptual integration of two abstract systems.

In addition to the open-ended question portion of the LRJA, there is an additional 26 item questionnaire which measures epistemological beliefs (e.g. *The only thing that is certain is uncertainty itself*; and *The ability to learn is innate*), each of which is responded to using a seven-point scale ranging from 'strongly disagree' to 'strongly agree'. More specifically, epistemological beliefs refer to an individual's beliefs about "the nature of learning" (Schommer-Aikins, 2002, p. 104), as well as the beliefs about how they justify knowledge (King & Kitchener, 1994; Kitchener &

King, 1981). The questionnaire was derived from Schommer's (1990; 1993) Epistemological Questionnaire (SEQ) and measured: *malleability of knowledge* (e.g. 'the belief that knowledge is fixed at birth' or 'the belief that the ability to learn can be improved'); *structure of knowledge* (e.g. 'the belief that knowledge is best characterized as isolated bits and pieces' or the 'belief that knowledge is best characterized as complex interrelated networks'); and *stability of knowledge* (e.g. 'the belief that knowledge is unchanging' or the 'belief that knowledge is evolving'). The internal consistency of the three scales ranged from $\alpha = .61 - .73$.

The *California Critical Thinking Disposition Inventory* (CCTDI; Facione & Facione, 1992) was administered at post-testing. The CCTDI consists of 75 items (e.g. *It bothers me when people rely on weak arguments to defend good ideas*; and *People say I rush into decisions too quickly*), each of which is responded to using a six-point scale ranging from 'strongly disagree' to 'strongly agree'. Subscales of the CCTDI include: truth seeking, open mindedness, analyticity, systematicity, confidence, inquisitiveness and maturity. The estimates of test reliability for the CCTDI full scale range from .85 – .90; and range from 0.72–0.86 for the seven subscales (Facione & Facione, 1992; Suliman, 2006).

A customised questionnaire was administered upon completion of the intervention, which asked students to rate various aspects of the course, including their understanding of the materials and instruction (again, see Appendix D). The questionnaire also asked students to suggest ways in which the course might be improved.

5.2.4 Procedure

The study took place over eight months, during which the two experimental groups attended a six-week classroom based CT course (i.e. 12 hours – two hours per

week) designed to teach CT according to the framework provided by the *Delphi Report*. One group received an AM version of the course while the other received a HO version. The control group completed both the pre-test and post-test assessments, but did not participate in any CT course.

Prior to commencement of Study 2, participants were recruited via announcements in Arts Faculty lectures, posters placed around the university and a mass email sent to all Arts Faculty students. The duration of the recruitment phase was two weeks. Students who were interested in participating attended the testing session in Week 1. In Week 1, participants were provided with further information regarding the nature of the study, as well as the option of participation in it, based on the understanding that they could withdraw from the study at any time. The LRJA was also administered in Week 1. In Week 2, the CCTST (Form B) was administered and the course began. Lectures and classroom activities were delivered to four different groups: 2 AM and 2 HO groups. On average, there were 20 participants present in each group. Both the AM and HO lectures involved *Microsoft PowerPoint*[™] slideshows with *ECHO 360*[™] pre-recorded voiceover; and were identical with regard to substantive content for both AM and HO groups. More specifically, identical, scripted voice recordings were dubbed over both versions of the slideshows. Voiceover quality was rated highly by students and no differences between conditions as assessed by the post-intervention questionnaire were noted. The in-class handouts and slideshows used only differed across conditions in terms of AM or HO presentation (again, see Appendix D). The recordings and exercises were administered by two classroom facilitators, both of whom were trained and randomly assigned to facilitate both an AM group and a HO group; and both of whom were

rated similarly by students in the post-intervention questionnaire in terms of their ability to facilitate the course.

From Weeks 2-7, the CT course focused on three core CT skills: analysis, evaluation and inference. Students were taught how to use each of these skills by viewing lectures and worked examples, as well as by applying the skills in the context of class exercises (see Table 5.2). In the analysis-focused lectures, students worked to identify the sources of arguments (e.g. anecdotes, common sense, authority, statistics and experiments) and the presence of balance (of reasons and objections) and bias (dominance of either reasons or objections in relation to a claim) in argument structures. Students were also instructed to extract the structure of arguments for analysis using either AM or HO structural tools. During the evaluation training, students worked to evaluate the relevance, credibility, and logical strength of arguments, and the overall balance of relevant, credible and logical evidence in an argument structure. During inference training, students worked to gather the appropriate information from arguments and draw logical conclusions from either AM or HO structures.

As noted, both the AM- and HO-infused CT courses used both didactic instruction (i.e. pre-recorded lectures) and active learning. During the lectures, the recordings were often paused and restarted by the facilitator, in order to allow time for students to complete their active learning exercises after a section of the recording where a CT skill was introduced and a worked example was presented. This was a design feature of the standardized teaching strategy that had proved effective in the pilot research and which ensured adequate control over both didactic and active learning components of AM and HO training. Approximately 75% of the time allotted

Table 5.2: Critical Thinking Course Outline

Class No.	Title	What Was Taught
1	Pre-Testing	<ul style="list-style-type: none"> Students completed the LRJA pre-test
2	Session 1: “Introduction to Critical Thinking”	<ul style="list-style-type: none"> Students completed the CCTST (Form B) pre-test. 1. We think in order to decide what to do and what to believe. 2. We ultimately decide what to believe by adding supports or rebuttals to our own arguments (i.e. questioning our own beliefs). 3. Arguments are hierarchical structures. We can continue to add more levels to these hierarchical argument structures if we like. Exercises: <ol style="list-style-type: none"> Adding <i>but</i> to <i>because</i> in an argument structure. Add <i>but</i> to <i>because</i> with more <i>buts</i>. Add <i>rebuttals</i> to complete a 4-level argument structure.
3	Session 2: “Unpacking (analysing and evaluating) a persons’ belief”	<ol style="list-style-type: none"> In order to analyse an argument, we must extract the structure of the argument from dialogue or prose. Identifying types (sources) of arguments and considering the strength of each type is another form of analysis. Evaluation of the overall strengths and weaknesses of an argument can be completed after adequate analysis. <ul style="list-style-type: none"> Exercises: <ol style="list-style-type: none"> Extract the argument structure in the dialogue and place each proposition in its appropriate position in the template structure. Extract the argument structure in the text and place each proposition in its appropriate position in the template structure. Identify the source of each proposition and add a rebuttal to each.
4	Session 3: “Analysis & Evaluation”	<ol style="list-style-type: none"> Evaluation includes the recognition of imbalances, omissions and bias within an argument. Evaluative techniques can aid recall. Examining whether or not the arguments used are <i>relevant or logically connected</i> to the central claim is also an important factor in evaluation. <ul style="list-style-type: none"> Exercises: <ol style="list-style-type: none"> Read & Analyse the argument structure. Fill in the template with missing propositions from information in the text. Read the argument structure on <i>Critical Thinking</i> and try to memorise as much information as possible - Memory Test on <i>Critical Thinking</i>. Read & Analyse the text-based argument. Complete the argument structure from the information in the text.

Table 5.2: Critical Thinking Course Outline (continued)

Class No.	Title	What Was Taught
5	Session 4: "Evaluation"	<p>We must evaluate:</p> <ol style="list-style-type: none"> 1. Types (sources) of arguments based on credibility 2. The relevance of propositions to the central claim or intermediate conclusions within the argument 3. The logical strength of an argument structure 4. The balance of evidence in an argument structure <ul style="list-style-type: none"> • Exercises: <ol style="list-style-type: none"> 1. Identify and evaluate the credibility of each proposition based on source. 2. Evaluate the relevance and credibility of each proposition within the argument structure. Eliminate irrelevant propositions. 3. Evaluate the logical strength of the argument by identifying which 4 (out of 6) propositions act as justification.
6	Session 5: "Inference"	<ol style="list-style-type: none"> 1. Evaluation and inference are intimately related. 2. Inference differs from evaluation in that the process of inference involves <i>generating</i> a conclusion from previously evaluated propositions. 3. In larger informal argument structures, intermediate conclusions must be inferred prior to the inference of a central claim. <ul style="list-style-type: none"> • Exercises: <ol style="list-style-type: none"> 1. Syllogistic Reasoning: Identify whether or not each syllogism is valid (10 syllogisms). 2. Identify whether or not the 3 intermediate conclusions are valid. 3. Infer 3 intermediate conclusions based on the information within the argument structure. 4. Work from the bottom up to infer each intermediate conclusion and the central claim (3 inferences).
7	Session 6: "Making Another's Argument Your Own"	<ul style="list-style-type: none"> • Review of all the previous 5 sessions <ol style="list-style-type: none"> 1. Analysis 2. Evaluation 3. Inference • Exercises: <ol style="list-style-type: none"> 1. Identify 6 propositions from the passage on hypnotism, add 2 new propositions and use these to construct a logical argument and infer a conclusion. 2. Identify 6 propositions from the passage on downloading music, add 2 new propositions and use these to construct a logical argument and infer a conclusion. • Students completed the CCTST (Form 2000) post-test.
8	Post-Testing	Students completed the LRJA post-test and the CCTDI

to each class (i.e. approximately 1 hour 30 minutes) was dedicated to this active learning.

In Week 7, after the completion of the courses, the CT ability of all three groups was again measured, this time using the CCTST Form 2000. In Week 8, the LRJA and the CCTDI were administered to all three groups. Also in Week 8, participants completed a questionnaire which asked them to comment on various aspects of the course and make suggestions for improving it. The LRJA was administered a third time, six months after completion of the CT courses, in order to examine whether CT training aided the long-term development of RJ.

5.3 Results

Preliminary analysis evaluating the homogeneity-of-slopes assumption indicated that in the case of pre-test CT performance, relationships with the dependent variables did not differ as a function of experimental condition, $F(2, 75) = .391$, $p = .677$. Means and standard deviations for scores on the CCTST, LRJA and CCTDI are presented in Table 5.3. Table 5.4 presents a summary of the ANCOVAs conducted in the current study. Table 5.5 presents inter-correlations between all dependent measures included in this study.

5.3.1 Group differences in CCTST performance

Results from a series of six ANCOVAs revealed that there was no main effect of group on overall CT performance. However, there was a main effect of group on analysis performance (again, see Table 5.4), with the HO group scoring significantly higher than controls ($p < .05$) and the AM group borderline higher than controls ($p = .06$; again, see Table 5.3). There was no difference between the AM and HO groups. Post hoc analyses confirmed that, regardless of teaching strategy used, those who

Table 5.3: Means and standard deviations for the three groups on the CCTST, LRJA & CCTDI

	<u>Pre-Test</u>		<u>Post-Test</u>		<u>Six-Month Follow-Up</u>	
	M (%)	SD	M (%)	SD	M (%)	SD
<u>Overall CT</u>						
AM	.48	.15	.51	.14		
HO	.44	.16	.50	.13		
Control	.41	.15	.43	.13		
<u>Analysis</u>						
AM	.55	.15	.67	.17		
HO	.49	.18	.72	.16		
Control	.45	.19	.58	.22		
<u>Evaluation</u>						
AM	.46	.17	.45	.17		
HO	.41	.14	.41	.17		
Control	.36	.14	.34	.13		
<u>Inference</u>						
AM	.44	.16	.47	.16		
HO	.40	.15	.46	.15		
Control	.41	.16	.43	.16		
<u>Inductive Reasoning</u>						
AM	.49	.16	.59	.13		
HO	.42	.16	.58	.15		
Control	.40	.13	.50	.14		
<u>Deductive Reasoning</u>						
AM	.46	.15	.42	.16		
HO	.45	.16	.42	.16		
Control	.40	.13	.36	.16		
<u>RJ</u>						
AM	3.56	1.05	3.66	0.91	3.65	1.08
HO	3.42	1.05	3.60	1.17	4.12	1.54
Control	2.92	0.90	2.85	0.99	3.27	1.28
<u>EB Malleability</u>						
AM	25.24	4.36	24.20	2.84	24.86	3.32
Outline	24.71	3.54	24.39	3.24	23.91	3.75
Control	24.21	3.10	24.37	3.13	24.30	2.31

Table 5.3: Means and standard deviations for the three groups on the CCTST, LRJA & CCTDI (continued)

	<u>Pre-Test</u>		<u>Post-Test</u>		<u>Six-Month Follow-Up</u>	
	M (%)	SD	M (%)	SD	M (%)	SD
<u>EB Structure</u>						
AM	36.20	5.06	34.64	3.89	35.50	4.94
Outline	35.25	5.56	34.11	4.98	35.64	5.08
Control	35.11	3.83	36.16	3.34	35.30	2.11
<u>EB Stability</u>						
AM	45.72	5.60	46.04	5.50	45.93	6.55
Outline	47.07	3.58	45.50	4.19	45.09	4.50
Control	45.63	4.50	45.47	4.56	45.30	4.22
<u>Overall Disposition</u>						
AM			290.81	30.97		
HO			288.06	29.63		
Control			284.14	20.93		
<u>Truth-Seeking</u>						
AM			36.89	6.48		
HO			36.88	5.78		
Control			33.86	5.64		
<u>Open-Mindedness</u>						
AM			42.04	5.17		
HO			43.00	5.35		
Control			43.18	5.85		
<u>Analyticity</u>						
AM			43.44	5.14		
HO			42.21	5.75		
Control			41.93	4.80		
<u>Systematicity</u>						
AM			36.30	7.29		
HO			36.97	7.24		
Control			35.32	4.46		
<u>Confidence</u>						
AM			42.59	6.52		
HO			39.26	7.75		
Control			39.93	7.44		

Table 5.3: Means and standard deviations for the three groups on the CCTST, LRJA & CCTDI (continued)

	<u>Pre-Test</u>		<u>Post-Test</u>		<u>Six-Month Follow-Up</u>	
	M (%)	SD	M (%)	SD	M (%)	SD
<u>Inquisitiveness</u>						
AM			46.37	5.68		
HO			45.38	5.40		
Control			47.43	4.67		
<u>Maturity</u>						
AM			43.19	6.15		
HO			44.35	5.89		
Control			42.50	5.59		

obtained CT training (i.e. both AM and HO groups combined) scored significantly higher on analysis than controls, $F(1, 78) = 8.74, p = .004$, partial $\eta^2 = .10$.

There was no main effect of group on evaluation performance. However, post-hoc analysis did reveal that the combined AM and HO group did score significantly higher than controls on post-test evaluation performance, $F(1, 78) = 4.31, p = .041$, partial $\eta^2 = .05$. Results also revealed a borderline main effect of group on inductive reasoning. Post-hoc analyses revealed that both experimental groups scored significantly higher than controls at post-testing (AM: $F[1, 77] = 4.44, MSE = .02, p = .038$; HO: $F[1, 77] = 4.52, MSE = .02, p = .037$). There was no difference between the AM and HO groups. There were no effects of group on inference or deductive reasoning.

5.3.2 Group differences in LRJA performance

Results from a 3x3 Mixed ANOVA revealed that there was no main effect of group, no main effect of time, and no group x time interaction. Latent growth modelling analysis further confirmed a non-significant slope (i.e. change over time) in RJ scores. Three separate 3x3 Mixed ANOVAs revealed that there was no main effect

of group, no effect of time, and no group x time interaction on malleability of learning ability, structure of knowledge or stability of knowledge.

Table 5.4: ANCOVA Summary for Study 2

	df	df (error)	<i>F</i>	<i>p</i>	Partial η^2
<u>Group</u> <u>(AM v HO v Control)</u>					
Overall CT	2	77	1.65	.198	.04
Analysis	2	77	4.74	.011	.11
Evaluation	2	77	2.35	.103	.01
Inference	2	77	.49	.615	.01
Inductive Reasoning	2	77	3.08	.052	.01
Deductive Reasoning	2	77	.80	.452	.02
RJ	2	65	1.55	.220	.05
RJ (Six-month Follow-Up)	2	36	.77	.472	.04

5.3.3 Group differences on the CCTDI

A further series of eight ANOVAs was conducted to examine the effects of group on disposition and on CCTDI subscales. There were no main effects of group on overall disposition or on sub-scales of truth seeking, open mindedness, analyticity, systematicity, confidence, inquisitiveness or maturity.

5.3.4 Correlations

There was a significant, positive correlation between CT and RJ at both pre-testing ($r = .34, p < .05$) and post-testing ($r = .43, p < .01$). CT was not correlated with total disposition towards thinking at pre-testing (see Table 5.5), but the two were significantly, positively correlated at post-testing ($r = .45, p < .001$). RJ was

Table 5.5: Correlations (Pearson's) Among CT Skills and Disposition at Pre-testing (below diagonal) and Post-testing (above diagonal).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1 Critical Thinking		.15	.40 ³	.33 ²	.34 ²	.33 ²	.45 ³	.31 ¹	.26	.44 ³	.26	.34 ¹	.32 ¹	.20	.43 ²	-.03	-.28	.17
2 Analysis	.71 ³		.40 ³	.36 ³	.62 ³	.54 ³	.17	.34 ¹	.17	.17	.03	-.08	.05	.20	.22	-.15	-.01	-.15
3 Evaluation	.73 ³	.25 ¹		.48 ³	.80 ³	.57 ³	.17	.22	.20	.09	-.07	.16	.15	.13	.10	-.18	-.05	.11
4 Inference	.73 ³	.38 ³	.22 ¹		.60 ³	.86 ³	.15	.25	-.08	.03	.01	.18	.03	.21	.19	.02	-.06	.33 ¹
5 Inductive Reasoning	.75 ³	.43 ³	.79 ³	.35 ²		.47 ³	.12	.17	.13	.10	-.06	.11	.16	.03	.05	-.18	-.16	.13
6 Deductive Reasoning	.77 ³	.46 ³	.43 ³	.80 ³	.23 ¹		.22	.37 ²	.02	.08	.04	.13	.01	.34 ¹	.30	-.02	.06	.20
7 Overall Disposition	.16	.08	.10	.18	.17	.17		.66 ³	.54 ³	.81 ³	.72 ³	.69 ³	.78 ³	.54 ³	.33 ¹	.01	-.14	.02
8 Truth-Seeking	.28 ¹	.14	.22	.27 ¹	.18	.36 ²	.66 ³		.33 ¹	.34 ¹	.45 ³	.24	.31 ¹	.41 ²	.36 ¹	-.41 ¹	-.48 ²	-.05
9 Open-Mindedness	.23	.32 ¹	.01	.21	.21	.15	.54 ³	.33 ¹		.32 ¹	.20	.11	.38 ²	.35 ²	.50 ³	.02	-.18	-.01
10 Analyticity	.10	.02	.05	.14	.18	.01	.81 ³	.34 ¹	.32 ¹		.51 ³	.68 ³	.64 ³	.32 ¹	.24	.26	.11	.03
11 Systematicity	-.11	-.15	-.06	-.06	-.03	-.06	.72 ³	.45 ³	.20	.51 ³		.42 ²	.53 ³	.23	.09	.00	.11	-.07
12 Confidence	.09	-.01	.12	.08	.10	.10	.69 ³	.24	.11	.68 ³	.41 ²		.55 ³	.09	.15	.15	.10	.27
13 Inquisitiveness	.05	.08	.03	.01	.00	.13	.78 ³	.31 ¹	.38 ²	.64 ³	.53 ³	.55 ³		.31 ¹	.17	.20	-.06	.09
14 Maturity	.13	.07	.06	.15	.19	.07	.54 ³	.41 ²	.35 ²	.32 ¹	.23	.09	.31 ¹		.17	-.23	-.38 ¹	-.24
15 Reflective Judgment	.34 ¹	.28	.11	.38 ¹	.28	.23	.64 ³	.47 ²	.55 ³	.51 ³	.43 ²	.42 ²	.43 ²	.33 ¹		.05	-.18	.02
16 Malleability of Learning	-.23	-.11	-.31 ¹	-.03	-.26	-.18	-.07	-.27	-.01	.13	-.01	-.04	.17	-.30 ¹	-.05		.45 ³	.42 ³
17 Structure of Knowledge	-.28 ¹	-.29 ¹	-.14	-.19	-.15	-.29 ¹	-.17	-.44 ³	-.21	-.04	-.07	.00	.03	-.21	-.37 ¹	.47 ³		.25 ¹
18 Stability of Knowledge	-.17	-.19	-.04	-.17	-.15	-.14	-.05	-.29 ¹	-.14	.10	-.06	.12	.21	-.21	-.22	.36 ³	.35 ³	

Significance levels ¹ = p at the .05 level; ² = p at the .01 level; ³ = p at the .001 level

significantly correlated with disposition towards thinking at both pre-testing ($r = .64$, $p < .001$) and post-testing ($r = .33$, $p < .05$). Epistemological beliefs towards the structure of knowledge was significantly, negatively correlated with both CT ($r = -.29$, $p < .05$) and RJ ($r = -.37$, $p < .05$) at pre-testing ($r = -.29$, $p < .05$), but not at post-testing. Disposition towards thinking was also not correlated with epistemological beliefs at any testing time.

5.4 Discussion of Study 2

5.4.1 Interpretation of Results

Study 2 compared the effects of training CT through AM and HO with a no CT training control condition on overall CT, CT sub-skill and RJ performance; as well as epistemological beliefs, and disposition towards thinking. Using two different forms of the CCTST, CT was assessed both before and after the CT training period. Results revealed that the HO group outperformed controls on both analysis and inductive reasoning, while the AM group outperformed controls on inductive reasoning. Although results associated with observed trends and borderline significance must be evaluated with caution, generally, the findings from Study 2 point to a positive effect of CT training on the CT skills of analysis, evaluation and inductive reasoning, with no significant difference between the AM and HO training regimes.

Unlike in previous research by van Gelder and colleagues (2001, 2004), in Study 2, a significant effect of AM training on overall CT performance was not observed. However, a trend towards significance was observed in which those exposed to CT, whether trained through AM or HO, outperformed controls on overall CT performance, $F(1, 78) = 2.87$, $p = .094$. Furthermore, as noted above, there was a significant positive effect of CT training on the sub-skills of analysis, evaluation and

inductive reasoning observed (i.e. 3 of 5 sub-scales measured by the CCTST and two of the three CT skills identified by the *Delphi Report*).

One seemingly odd finding was that though inductive reasoning and inference are closely related skills involved in drawing conclusions; there was an effect of CT training for one (i.e. inductive reasoning) but not the other. One possible reason for this might be that inductive reasoning is measured by the CCTST in a manner that simultaneously measures other skills. More specifically, the CCTST consists of 34 questions, 17 of which examine inductive reasoning and 17 of which examine deductive reasoning. The same 34 questions also tap into the CT sub-skills of analysis, evaluation and inference. As a result, each question on the CCTST pertains to two CT sub-skills. It may be that students who took part in the CT course performed better on the inductive reasoning questions because roughly two-thirds of these questions also pertained to analysis and evaluation (which were shown to be improved by HO and AM).

The two experimental groups scored significantly higher on analysis and evaluation performance, but did not score higher on inference performance, when compared with the control group. One possible explanation for this is that relatively more time was devoted to training both analysis and evaluation skills. Specifically, six hours were devoted to training evaluation skills, four hours were devoted to training analysis skills, but only two hours were devoted to training inference skills. Thus, it may be that the time devoted to training inference skills was insufficient to bring about any positive change in inference test scores from pre-test to post-test.

Results showed no significant differences among groups or across testing times on RJ performance. Although the current study is unique in the sense that it sought to examine changes in RJ associated with the explicit training of related CT

skills, the results of the current study are broadly consistent with research by King and Kitchener (1994) and Dawson-Tunik et al. (2005), suggesting that it may take longer than 6 months for undergraduate college students to make significant advancements in RJ ability (e.g. to advance from abstract mappings to systems thinking). At the same time, Study 2 may have lacked statistical power due to the relatively small sample size, a possibility which once again suggests caution in the interpretation of these results. The effect of a relatively small sample size may have been further compounded by the sensitivity of the RJ measurement system to capture subtle changes in performance. Specifically, while the RJ scoring system is designed to distinguish significant shifts in the RJ level of test takers, some of the changes that occur in the nature of CT and RJ may be more subtle and difficult to capture by reference to these RJ level distinctions.

Though the lectical analysts maintain an inter-rater reliability that is at or above 85% agreement within $\frac{1}{4}$ of a level (with internal consistency typically above .90; Dawson-Tunik, Commons, Wilson, & Fischer, 2005), the difference between the numerical values of the LRJA's scoring output at two different testing times (e.g. the difference between 10:3 at pre-testing and 11:2 at post-testing) may not have reflected the significance of the gain in RJ made by the student. For example, the difference between one conceptual level (e.g. abstract mappings) and another (e.g. abstract systems) may be much more significant when evaluated *qualitatively* than when statistically compared in the context of *quantitative* analyses, and this may reflect the LRJA scoring focus on gross levels rather than subtle distinctions in marking qualitative changes in performance. As a result, it is important to approach the RJ findings in the current study with caution, due to the scoring method utilised by the

LRJA. Notably, an alternative method of assessing RJ is discussed below in section 5.4.2.

Alternatively, it is also possible that the CT course(s) did not facilitate RJ development because RJ was not explicitly trained. That is, though analysis, evaluation and inference were explicitly trained in the six training sessions, there was no dedicated focus on RJ in the training course. Although the literature on RJ suggests that RJ cannot be taught as a skill, but must somehow develop (Brabeck, 1981; Dawson, 2008a; Hofer, 2004), teaching students what RJ is and providing them with examples of RJ may have enhanced their ability to respond to LRJA questions with greater RJ and higher level responses that reflected more complex abstract mappings, abstract systems, single principles or principled mappings to structure their reasoning (e.g. again, see Table 5.1). At the same time, it is unclear how principle level thinking can be trained and whether or not standard CT intervention course methodologies are the best way to cultivate principle-level thinking.

As such, one last possibility to consider is that the intervention of a standard CT course like the one used in the current study, regardless of whether it is taught through AM or HO, will have little or no effect on RJ performance. Having said that, in the current study, the rate of attrition of participants over time was largest for the LRJA 6-months follow-up (i.e. a decrease from $N = 69$ to $N = 40$), which made the 6-month test of RJ development difficult to assess, given that some of those who did not complete the follow-up may have experienced continued development that was not captured in this analysis.

Results also revealed that RJ was significantly correlated with both CT performance and CT dispositions (i.e. overall disposition towards thinking as well as disposition towards truth-seeking, open-mindedness, analyticity, systematicity,

confidence, inquisitiveness and maturity) at pre-and-post-testing. These findings are consistent with past research, which suggests that RJ is a component of CT (Baril et al., 1998; Huffman et al., 1991; King, Wood & Mines, 1990) and that positive dispositions towards thinking are positively correlated with CT performance (Colucciello, 1997; Facione, Facione & Sanchez, 1994; Profeto-McGrath, 2003). Also consistent with the aforementioned past research, the results of the current study revealed a positive, significant correlation between overall CCTDI scores and CT performance at post-testing. The significant correlation between CT performance and total CT disposition scores was largely accounted for by significant, positive correlations between CT performance and CT dispositions of truth-seeking, analyticity, confidence and inquisitiveness. Furthermore, truth-seeking was significantly correlated with CT ability and inference at pre-testing and CT ability and analysis at post-testing. These findings suggest that perhaps the disposition towards truth-seeking is a critical and stable feature of good CT ability.

With regards to intra-group performance, disposition towards thinking was not correlated with the CT performances of controls or those who received CT training at pre-testing. However, at post-testing, the CT performance of those who received CT training was significantly correlated with truth-seeking ($r = .31, p = .023$), analyticity ($r = .44, p = .001$), confidence ($r = .34, p = .012$) and inquisitiveness ($r = .32, p = .017$). More generally, given that the post-test CT performance of controls was not correlated with a positive disposition toward thinking ($r = .25, p = .517$) and yet the post-test CT performance of those who received CT training was significantly correlated with CT dispositions ($r = .48, p = .001$), these findings suggest that a positive disposition toward thinking may emerge as a significant correlate of CT performance only after students have been exposed to some training in CT skills.

At the same time, the correlations between disposition scores and pre-intervention CT performance must be interpreted with caution, as disposition towards thinking was only measured at post-testing (i.e. the correlation between pre-test CT performance and post-test disposition are based on measurements at different times). Furthermore, researchers have identified problems with the measurement of dispositions, including the problematic nature of measuring CT dispositions using self-reports. It has been suggested that measuring CT ability and disposition separately is an inadequate method of “filling the gap” between what people do (i.e. applying CT on an actual test) and what they claim they do in a self-reported dispositional measure (Ku, 2009, p. 73).

Interestingly, results revealed that though there was no significant correlation between CT performance and epistemological beliefs regarding the malleability or stability of knowledge at pre-or-post-testing, CT performance was significantly, negatively correlated with epistemological beliefs regarding the structure of knowledge at pre-testing, but not at post-testing. It is unclear whether or not this is a chance finding or if a belief in the structure of knowledge as an interrelated network of isolated items may somehow predict poor CT performance prior to any CT training. With respect to the null correlations, it may be that the self-regulatory, dispositional factors that share a relationship with CT ability are ones that reflect an individual's *willingness* to conduct CT skills (Valenzuela, Nieto & Saiz, 2011), for example, disposition towards truth-seeking, analyticity, confidence and inquisitiveness, rather than ones that reflect an individual's *beliefs* about the nature of knowledge. Overall, given that there were no differences among the AM, HO and control groups on disposition towards thinking, any of the CCTDI sub-scales, or any of the epistemological belief sub-scales at pre-or-post-testing, these findings suggest

that post-test CT performance differences between groups were not a result of differences or changes in dispositional factors.

5.4.2 Limitations & Future Research

Study 2 produced a number of interesting findings, but was also characterised by a number of limitations. One limitation was the small sample size, which decreased the power of the statistical analysis. A possible reason for the small sample was that participation in the CT course was voluntary and thus, students may have been reluctant to register for an additional course module in an already busy academic schedule. From a pool of over 1,000 students, it was originally estimated that approximately 300 students would register for the course. Those who did sign up were to be randomly allocated to the AM group, the HO group or the control condition. However, only 101 students registered for the course. Thus, to ensure adequate statistical power in the comparison of the AM and HO conditions (in which case, each condition would require a sample size of 40), those 101 students were randomly allocated to one of two experimental conditions.

As a result, it was necessary to recruit an additional group of students for the control condition, which contributed to another limitation of the study – inadequate random assignment to the control condition. This is an important issue to consider, as inadequate random assignment is a violation of internal validity. Thus, all results involving a comparison of the control group with one or both of the experimental groups must be interpreted with caution.

The 33 students assigned to the control condition were students who had expressed an interest in attending the CT course but who could not attend due to conflicting schedules. Thus, although these students agreed to take pre-test and post-test measures, their motivation to perform well on CT and RJ tests may not have been

as high as those who attended the CT training. Having said that, results indicated that there were no differences among groups on disposition towards thinking at post-testing. Notably, there was also significant attrition from pre-to-post-testing. Of the 134 experimental and control group students who initially completed the CCTST pre-test, only 81 completed the post-test. Thus, the small sample size had a negative impact on statistical power in the current study, which made it more difficult to identify significant effects when comparing groups on CT and RJ performance at post-test.

Another possible limitation to this study was the limited instructor feedback provided to participants during training. A meta-analysis conducted by Marzano (1998), which examined the effects of different methods of instruction on learning, indicated that across a number of pedagogical studies, feedback on the type of strategy used to improve learning and on the efficacy of its use, produced a significant gain in student achievement (with an effect size of 1.31). Furthermore, van Gelder (2003) and van Gelder, Bissett and Cumming (2004) suggest that AM training requires provision of adequate feedback to students. Feedback can of course provide additional opportunities to evaluate and reflect upon one's own thinking; and can increase motivation, which might have curbed attrition levels in Study 2. However, in the current study, it was considered important to control as many extraneous variables as possible. Thus, it was judged that the provision of detailed feedback, that was both standardized and equal in every respect across two instructors and two experimental conditions, would be logistically difficult to achieve and might act as a potentially confounding factor in the comparative analysis of AM and HO conditions. Thus, only limited feedback was provided by classroom facilitators in an effort to ensure that all students understood instructions and were able to complete exercises. The reports of

students who completed the post-intervention questionnaire indicated that they found the recordings and exercises easy to follow. In addition, classroom feedback was provided by students as they occasionally worked cooperatively together to complete exercises in class. Also, the classroom facilitators reported that students generally completed classroom exercises with relative ease and enjoyment. In retrospect, while the provision of structured feedback, particularly focused on the quality of students' reasoning in their AM and HO workings, may have had an effect on motivation and performance, efforts made to control the nature of feedback was designed in the spirit of experimental rigour and an analysis of potential study confounds.

Future CT intervention research examining the effects of AM should aim to provide students in different experimental conditions with structured and rigorously controlled feedback such that experimental conditions can be reliably compared. As discussed in Chapter 3, research by Butchart et al. (2009) found that an AM-infused CT course that provided students with automated feedback resulted in mean gains in CT performance with an effect size of .45, whereas students who completed the AM exercises without automated feedback showed mean gains with an effect size of only .22. Though participants in the automated feedback group only received automated notice of a 'correct' or 'incorrect' response, the authors of that study note that this 'less than ideal' form of feedback still yielded an effect size double that of the effect size for the 'no-feedback' group. Thus, it is recommended that future research should include the provision of feedback on CT exercises as standard, and preferably provide feedback accompanied by explanations as to *why* a response is correct or incorrect, which seems particularly relevant in the context of CT training. The only caveat in this context is that if courses taught through AM are being explicitly compared with courses that use other tools of critical thinking, every effort should be made to

standardize feedback protocols and somehow ensure that the quality and intensity of feedback is similar across different experimental conditions and different instructors. This is important to consider for future studies, as previous research has found that such differences in treatment of groups may affect performance, for example, as a result of experimenter bias (e.g. Hancock, 2002; Rosenthal, 2006; Rosenthal & Jacobsen, 1966) and instructor teaching style (Deci et al., 1982; Felder & Silverman, 1988; Wentzel, 2002).

Another potential weakness of Study 2 was the choice of CT measure used. Though the CCTST measures CT and CT sub-skills according to the *Delphi Report* definition and framework, it is not necessarily ideal for evaluating gain in intervention studies. For example, according to Jacobs (1995), the various CCTST Forms are characterised by different levels of difficulty and should therefore not be used for purposes of measuring individual differences or gains from pre-to-post-testing. In addition, each version of the CCTST (i.e. Forms A, B and 2000) contains a different number of questions per subscale. Thus, Jacobs (1995) recommends that one form should be used as both a baseline measure and a covariate when examining intervention group differences. In accordance with Jacobs' recommendation, in Study 2, Form B was administered as a covariate and baseline measure of CT and Form 2000 as the outcome measure. This, however, is less than an ideal situation as such a design did not allow for the measurement of CT gain (i.e. from pre-test to post-test). Future research should seek to use CT assessments that allow for the measurement of gain, perhaps in the context of a reliable and valid scoring protocol that affords researchers the opportunity to measure the growth of CT ability throughout the training period.

The format of the CCTST itself is also less than ideal. More specifically, items are measured via multiple choice questions (MCQs). This CT testing format is problematic because it allows test-takers to simply guess when they do not know the correct answer, instead of demonstrating their ability to critically analyse and evaluate problems and infer solutions to those problems (Ku, 2009). Furthermore, as argued by Halpern (2003b), the MCQ format of the CCTST makes the assessment a test of verbal and quantitative knowledge rather than CT (i.e. because one selects from a list of possible answers rather than determining one's own criteria for developing an answer; Ku, 2009). The measurement of CT through MCQs is also problematic given the potential incompatibility between the conceptualisation of CT that shapes test construction and its assessment using MCQs. That is, MCQ tests assess cognitive capacities associated with identifying single right-or-wrong answers and as a result, this approach to testing is unable to provide a direct measure of test-takers' use of metacognitive processes such as CT, RJ and disposition (Halpern, 2003a; Ku, 2009).

Instead of using MCQ items, a better measure of CT might ask open-ended questions, which would allow test-takers to demonstrate whether or not they spontaneously use a specific CT skill. One commonly used CT assessment that employs an open-ended format is the Ennis-Weir Critical Thinking Essay Test (EWCTET; Ennis & Weir, 1985). However, this test has been criticised for its domain-specific nature (Taube, 1997), the subjectivity of its scoring protocol and its bias in favour of those proficient in writing (Adams, Whitlow, Stover & Johnson, 1996). Similarly, the LRJA, which was used in this research as a measure of reflective judgment, also employs an open-ended format of assessing thinking. Though, unlike the EWCTET, Dawson (2008b) argues that the LRJA is domain general. However, just as the EWCTET has been criticised for its potential bias in favour of those

proficient in writing, the same criticism can also be made in relation to the LRJA. Specifically, although it has been claimed that the LRJA is a reliable and valid measure of RJ ability, it is possible that this system of *lectical* assessment is biased in favour of those who are more skilled at writing or even those who possess a larger vocabulary (e.g. vocabulary reflecting knowledge of abstract concepts that are relevant to the problem domain).

Though the scoring of the LRJA is arguably domain general and very useful in characterising hierarchical complexity in the use of principles, abstractions and concrete representation, more definitive quantitative criteria for judging the quality of RJ responses is necessary. That is, perhaps a scoring protocol that focuses more on the key features of hierarchical complexity (e.g. the number and coherence of links in a chain of arguments) would allow for the development of an assessment that provides a measure not only of overall hierarchical complexity, but also of features of analysis, evaluation, and inference that reflect the underlying CT of participants; in addition to the range of representations, abstractions and principles they seek to coordinate in their argument. At the current time, both MCQ and open-ended tests for assessing CT have their respective limitations. Perhaps combining the two response formats into one test (Ku, 2009), as the recent Halpern Critical Thinking Assessment (HCTA; Halpern, 2010) has done, might facilitate more reliable and valid assessment of CT ability.

The HCTA asks open-ended questions based on believable, everyday situations and examples, followed by specific questions that probe for the reasoning behind the answer. The multi-part nature of the questions makes it possible to assess the ability to use specific CT skills when the prompt is provided (Ku, 2009). The HCTA's scoring protocol also provides comprehensible, unambiguous instructions for

how to evaluate responses by breaking them down into clear, measurable components (see Chapter 6). The scoring protocol used for the open-ended portion of the HCTA is a much richer method of evaluating responses than the LRJA's lectical scoring system, which essentially provides an index that is not easy to interpret.¹⁶

One final limitation of Study 2 concerns the possibility that the full potential of AM was not assessed. AM is a potentially powerful method of communicating an argument structure, but it is also relatively novel and thus, people may have little or no experience using it. As noted, van Gelder, Bissett and Cumming (2004) suggest that AM training is optimized in the context of a semester-long CT course. Study 2 did provide training; however, the intervention used included only six weeks of actual training (i.e. 2 hours per week), which might not have been sufficient to realise substantive improvements in CT ability.

On the other hand, though no differences between the AM and control groups was found for overall CT performance, it seems unlikely that this is due to a failure to provide students in the AM condition with sufficient AM-based in-class exercises. Butchart et al. (2009) provided students in their study with eight homework assignments and 10 sets of exercises (i.e. 18 exercises in total), and found gains in CT resulting from their intervention. In the present study, students were also provided with 18 exercises, which is comparable with the intervention provided by Butchart and colleagues. Thus, though the effects were modest in the context of the sample size tested, the CT training in the current study was similar to that of other studies in many other respects.

Furthermore, unlike Butchart et al. (2009), in the current study, the three groups were statistically compared with each other, which was an improvement in

¹⁶ Notably, scoring of the LRJA can only be legally conducted by (hired) certified lectical analysts, who may or may not be affiliated with the research.

certain respects upon Butchart et al's (2009) study, which did not explicitly compare groups in their statistical analysis. In addition, the majority of past research studies conducted on the effect of CT training courses have lacked a control condition or an alternative, comparable treatment group (e.g. Adams, Stover & Whitlow, 1999; Burbach, Matkin & Fritz, 2004; Daly, 2001; Gadzella, Ginther & Bryant, 1996; Hitchcock, 2004; Scott, Market & Dunn, 1998; Hagedorn et al., 1999; van Gelder, 2001; van Gelder, Bissett & Cumming, 2004). Thus, while the current study provided a harsh test of the efficacy of AM in CT training, it was a reasonably fair test, and one which can be built upon in future experimental designs.

One improvement on the current design for future studies incorporating AM might be to include more AM practice *outside* the classroom, perhaps in an online environment. Study 2 sought to control the level of practice for both experimental conditions by restricting work to class-time only; however, this may have had a negative impact on the overall efficacy of the course and may have particularly disadvantaged AM. For purposes of internal validity, it is recommended that future research provide extended training in AM and/or other systems of argument representation, both inside and outside the classroom, before the final assessment of CT ability takes place.

Providing students with a CT course in an online environment may be one way of extending AM training, while also facilitating participation and engagement, as it would avoid timetable clashes for student participants. In this context, AM software may be provided to students outside the class setting (as suggested by van Gelder, Bissett & Cumming, 2004). In Study 2, students were not trained in the use of AM software per se, but rather in the method of AM using a 'pen-and-paper' method of

AM examination and construction, with AM exercises completed on printed handouts provided in class.

Furthermore, in future AM research, students should be assessed more often than at just pre-and-post-testing, or alternatively (as discussed above), provided with feedback on their exercises throughout the training period, so that they may monitor their own progress (i.e. as an incentive for continued participation). Given the challenge of recruiting participants and maintaining high levels of engagement in voluntary CT courses of this nature, both the use of online delivery systems and the provision of feedback that helps students to assess how they are improving over time may act as incentives to initial and continued engagement with the courses.

Though there were limitations within Study 2 that require consideration, this study does present further evidence that both AM and the CT skill instruction can enhance certain aspects of CT performance, even in the context of a relatively short course. This should provide further support for the argument that a focus on teaching CT can reap important gains. One of the conclusions reached on the basis of the work conducted thus far is that extended training may be needed to fully realise the benefits of AM. Research carried out in the next chapter (i.e. Study 3) sought to examine the effects of AM training on CT performance in the context of student engagement in an online CT course; and investigated the effects of differing levels of engagement and motivation towards learning on CT performance.

Chapter 6

Study 3: The Effects of Argument Mapping on Critical Thinking Performance in e-Learning Environments

6.1 Purpose

This chapter presents the empirical research carried out in Study 3, which compared the effects of an online critical thinking (CT) course taught through argument mapping (AM), with a no-CT course control group on measures of CT ability¹⁷. The current study also examined the relationship between student engagement in the CT course and CT performance changes over time. This study was conducted in light of the discussion of results from Study 2, with the primary goal of improving upon the design, implementation, and evaluation of AM-infused CT training.

6.1.1 Improvements in Design & Implementation from Study 2

One limitation of Study 2 was the large attrition rate and small sample size. The proposed solution to this problem was to redesign the CT course and administer it in an online e-learning environment. The administration of the CT course in an online environment was also recommended by students in their feedback on Study 2. Additionally, it was believed that administering CT training in an e-learning environment would make the provision of feedback to students more feasible and provide students with more opportunities to engage with AM training. In this context, it is important to consider in broad terms what e-learning entails.

¹⁷ Portions of Study 3 have been submitted for publication and are cited in the References section (see Dwyer, Hogan & Stewart, 2011c).

6.1.1.1 e-Learning

e-Learning is a method of instruction delivered through the use of multimedia technologies on a computer (e.g. software programmes or the internet) designed to foster the transfer of information for purposes of achieving specific learning goals (Clark, 2005; Huffaker & Calvert, 2003). A report by the U.S. National Research Council, which surveyed 1,206 public schools, suggests that the use of e-learning resources, in conjunction with students' application of metacognitive processes (e.g. CT), can potentially improve learning (Huffaker & Calvert, 2003).

Past research has shown that online e-learning instruction can have a beneficial effect on learning in a variety of training contexts (e.g. Agarwal & Day, 1998; Brown, 2001; Chen, 2009; Hugenholtz et al., 2008; Johnson et al., 2004). e-Learning can prove efficacious under certain conditions, provided that certain guidelines are adhered to (Clark, 2005; Mayer, 2003). For example, Mayer (2003) proposed that in order to design a multimedia e-learning course appropriately, the designer of the course must adhere to three basic 'cognitive assumptions' about learning. These assumptions are that (1) people process information both visually and phonologically (Mayer, 1997; Paivio, 1986); (2) people possess a limit for the amount of information that can be processed by either the visuospatial or the phonological coding system (Baddeley, 2000; Cowan, 2000; Miller, 1956); and (3) people learn more through active learning (e.g. Hake, 1998; Laws, Sokoloff & Thornton, 1999; Redish, Saul & Steinberg, 1997).

With respect to the design of e-learning courses, Clark (2005) proposes that e-learning is an efficacious method of both encouraging active learning (i.e. by providing students with the opportunity to complete multimedia exercises) and minimising cognitive load. The use of e-learning interfaces can be used to reduce

cognitive load in a variety of different ways, for example, by explaining complex visual content with audio narration, and by using integrated visuospatial presentation strategies to reduce levels of element interactivity. By redesigning the AM-infused CT training course in light of these principles of e-learning course design, and in light of the hypothesised benefits of AM described throughout this thesis, the first hypothesis of Study 3 was that those who participated in the AM-infused e-learning CT training course would outperform those in the control group at post-training on measures of CT ability.

6.1.1.2 Sample Size

As noted, e-learning was chosen as the CT course delivery method in part as a means of maximising the participant sample size, thus, addressing a limitation of the previous study. Notably, the majority of controlled trials examining the benefits of e-learning focus on asynchronous e-learning, which is informed by a wide array of models for e-learning (Clark, 2005; Colette, 2001; Garrison & Anderson, 2003; Govindasamy, 2002; Greenagel, 2002); and is designed for self-paced, individual study (Huffaker & Calvert, 2003; Twigg, 2002). Asynchronous e-learning courses are not instructor-led, but instead, are pre-recorded and made available online. As a result, students who participate in an e-learning course are able to engage in learning at a time and place that suits them (Clark, 2005; Clark & Mayer, 2007) and can thus avoid clashes with other compulsory courses. The option to rewind and replay recordings and instructions also facilitates different learners who work at different paces to optimize the rate of delivery to suit their own needs. It is for these reasons that the administration of the CT course through e-learning was speculated to make participant disengagement less likely and thus lower rates of participant attrition.

Apart from the implementation of e-learning into the CT intervention, the study design was also altered to maximise the available sample of student participants. That is, Study 3 compared the effects of a CT e-learning course taught through AM with a no-CT control condition. Unlike Study 2, a hierarchical outlining group was not compared with an AM group in Study 3, as a significant advantage for one method of teaching CT over the other was not observed in Study 2. It was decided to teach the CT class in Study 3 exclusively through AM, in light of (1) the overall goal of the thesis, which is to rigorously evaluate the hypothesised benefits of AM; (2) previous research claims that AM can enhance CT (e.g. Butchart et al., 2009; van Gelder, 2001); and (3) the beneficial effects of AM observed in Studies 1 and 2. As such, only two groups were compared in this study (i.e. a group of students who were taught CT through an AM-infused CT e-learning course and a control condition who received no CT training). By using e-learning and by comparing only two groups in the analysis of the effects of AM on CT, it was expected that an attrition rate similar to the one observed in Study 2 would still leave an adequate sample size for comparison purposes, assuming the initial recruitment of participants was optimized.

6.1.1.3 Feedback

Another limitation from Study 2 that could be addressed through the administration of the CT course in an e-learning environment was the limited feedback provided to students. In addition to the broadly valuable nature of feedback provision as a means to accelerating learning (Marzano, 1998), it is often argued that the provision of performance feedback to students is of critical importance in the context of e-learning (Clark, 2005; Mayer, 2003). Though there is usually no instructor physically present to guide students and offer advice to them in relation to their work in an e-learning environment, an e-learning course can provide certain

advantages in comparison with classroom environments in relation to the timely and efficient provision of feedback to students. For example, as part of the design of the AM-infused CT e-learning course used in this study, it was possible to provide students with feedback on common mistakes made and different strategies used by students in completion of their exercises. In addition, it was possible to provide this feedback in an efficient and standardized manner via e-mail and web-posts; and students were given the opportunity to view feedback and respond to the feedback in their own time.

6.1.1.4 Engagement in AM Training

Providing students with the opportunity to engage with AM training outside of the classroom environment was another issue discussed in relation to Study 2. Given that the administration of e-learning is computer-based, it was feasible in the design and delivery of Study 3 to instruct students to complete their exercises using AM software (i.e. *Rationale*TM; van Gelder, 2007) in a computer laboratory setting, as opposed to creating AMs via pen-and-paper (as in Study 2). Students were also encouraged to use this software for other subjects outside of the CT course. Given that one of the main objectives in improving the design of the intervention within Study 3 was to provide students with more opportunities to engage in AM training, a key aim in the analysis of results was to examine the effects of the level of engagement in AM training on CT performance. Previous research by van Gelder, Bissett and Cumming (2004) found that CT performance and AM practice hours were significantly correlated ($r = .31$). More generally, research suggests that higher levels of engagement with relevant course material is positively correlated with academic performance (Klem & Connell, 2004), as well as CT ability (Carini, Kuh & Klein, 2006). Thus, the second hypothesis examined in Study 3 was that students who

engaged more with the e-learning course would perform significantly better on CT performance than those who did not engage as much.

6.1.1.5 Method of Assessment

One final limitation of Study 2 that was addressed in the current study was the assessment used to measure CT performance (i.e. the California Critical Thinking Skills Test [CCTST; Facione, 1990a; Facione et al., 2002]). As discussed in Chapter 5, though the CCTST measures CT and CT sub-skills according to the *Delphi Report* definition and framework, it has been criticised as having limited reliability; that is, the multiple forms of the CCTST available differ in terms of difficulty and may therefore, measure different levels of CT ability (Jacobs, 1995). The validity of the CCTST has also come under scrutiny because, as argued by Halpern (2003b), the MCQ format of the CCTST makes the assessment a test of verbal and quantitative knowledge rather than CT. According to Ku (2009), this is because MCQ test-takers are not afforded the opportunity to determine their own criteria for developing an answer to the question; instead they are provided a list of possible answers choose from. Similarly, the MCQ format allows test-takers to simply guess the answer, in the event that do not know the correct answer, instead of critically analysing and evaluating the problem. The newly developed Halpern Critical Thinking Assessment (HCTA; Halpern, 2010) was recommended in Chapter 5 as an alternative CT measure. The HCTA asks open-ended questions based on believable, everyday situations and examples, followed by specific questions that probe for the reasoning behind the answer. The HCTA is further discussed below in section 6.2.3.

In addition to the administration of the HCTA in the current study, the Motivated Strategies towards Learning Questionnaire (Pintrich et al., 1991) and the Need for Cognition Scale (Cacioppo, Petty & Kao, 1984) were also administered to

participants at both pre-and-post-testing. As discussed in the previous chapter, there are many conceptualisations of disposition towards thinking. According to Valenzuela, Nieto and Saiz (2011), while some conceptualisations focus on the attitudinal and intellectual habits of thinking (e.g. as measured in Study 2 by the California Critical Thinking Dispositions Inventory; Facione & Facione, 1992), many others emphasise the motivational features associated with a positive disposition towards CT. That is, these motivation-focused conceptualisations emphasise the importance of motivation as a process used to activate the cognitive and metacognitive resources necessary to conduct good CT (Ennis, 1996; Norris, 1994; Perkins, Jay & Tishman, 1993; Valenzuela, Nieto & Saiz, 2011). Though few empirical studies have examined the motivational aspects of CT dispositions, research by Valenzuela, Nieto and Saiz (2011) revealed that motivation, or drive, to think critically is a more significant correlate of CT ability ($r = .50$) than is a general, positive disposition toward critical thinking ($r = .20$). Similarly, research by Garcia, Pintrich and Paul (1992) found a significant, positive correlation between CT ability and motivation towards intrinsic goal orientation ($r = .57$), elaboration ($r = .64$) and metacognitive self-regulation ($r = .64$) - three sub-scales of the Motivated Strategies towards Learning Questionnaire (MSLQ; Pintrich et al., 1991). In addition, research has also shown that motivation to learn positively influences CT and learning in general (Hattie, Biggs & Purdie, 1996; Robbins et al., 2004). Thus, while Study 2 measured a range of attitudes in relation to valued CT dispositions (i.e. disposition towards truth-seeking, open-mindedness, analyticity, systematicity, confidence, inquisitiveness and maturity) and beliefs in relation to the malleability, structure and structure of knowledge, Study 3 sought to clarify the impact of students' motivation to learn and behavioural engagement with course materials on subsequent training-

related CT performance outcomes. Therefore, in addition to recording student responses to the Motivated Strategies towards Learning Questionnaire (Pintrich et al., 1991) at pre-and-post-testing, the current study measured behavioural engagement by measuring the number of homework exercises completed by students in the online e-learning environment.

The Need for Cognition Scale (NCS; Cacioppo, Petty & Kao, 1984) was also administered in Study 3, as research suggests that, in addition to motivation to learn, dispositional need for cognition is also significantly correlated with CT performance. For example, research by Halpern (2006) found that performance by a group of college students on the HCTA was positively, significantly correlated ($r = .35$) with scores on the NCS. In addition, research by Jensen (1998) found a significant, positive correlation between the need for cognition and reflective judgment of 1st year college students; and also that the higher an individual's need for cognition, the more likely they were to generate solutions to ill-structured problems that reflected an understanding of the uncertainty of knowledge - thus demonstrating better reflective judgment (King & Kitchener, 2002). Furthermore, research by Toplak and Stanovich (2002) found that those with high levels of need for cognition are more likely to use disjunctive reasoning (i.e. reasoning strategies similar to CT "that require the exhaustive consideration of all of the possible states of the world"; p. 197) during problem-solving and are also more likely to demonstrate better reasoning ability.

On the basis of the theory and research described above, the third hypothesis of the current study was that both need for cognition and motivation towards learning strategies would be significantly, positively correlated with CT performance. The fourth hypothesis was that the potential, significant correlations between CT and motivation towards learning strategies would be higher in the AM group than in the

control group at post-testing, as training in CT may reinforce some aspects of motivation (e.g. goal orientation, elaboration and metacognitive self-regulation) and increase the strength of the relationship between dispositional features of CT and CT performance (cf. Garcia, Pintrich and Paul, 1992; Hattie, Biggs & Purdie, 1996; Rimiene, 2002).

6.2 Method

6.2.1 Design

A series of sixteen 2 (condition: AM and control groups) x 2 (time: pre-and-post-testing) Mixed ANOVAs was used to compare the effects of a CT e-learning course taught through AM with a no-CT control group on need for cognition, overall motivation towards learning strategies, motivation towards learning sub-scales, overall CT and CT sub-scales performance. A further series of sixteen Mixed ANOVAs was conducted in the AM training group alone in order to examine the effects of level of engagement in the CT e-learning course on need for cognition, overall motivation towards learning strategies, motivation towards learning sub-scales, overall CT and CT sub-scales performance. Correlations among dispositional need for cognition, motivation towards learning strategies and CT ability were also analysed. A hierarchical multiple regression was conducted to examine if need for cognition or motivation towards learning predicted CT performance. The predictor variables were motivation and need for cognition scores at pre-test (Block 1) and post-test (Block 2); and the criterion variable was CT performance. After completion of the CT intervention, a focus group interview was conducted in order to investigate students' perceptions of the CT e-learning course, with specific focus on the presentation and quality of lectures, exercises and feedback provided to students, as

well as the student's experience in the use of AM to facilitate CT ability. The interview transcripts were examined using thematic analysis.

6.2.2 Participants

Participants were first year psychology students, aged between 18 and 25 years, from the National University of Ireland, Galway. Two-hundred and forty-seven students (173 females, 74 males) expressed an interest in participating and attempted the online pre-tests. However, only 156 (108 females, 48 males) completed pre-testing; and only 74 participants (47 females, 27 males) completed post-testing. There were no baseline differences (i.e. in CT, need for cognition and motivation) between completers and non-completers. Non-completers reported not having enough time and a heavy workload from other mandatory courses as the primary reasons for why they withdrew. Participants in the focus group (N = 9; 1 female, 8 males) were students who took part in the CT e-learning course and completed both pre-and-post-testing. In return for their participation, students were awarded academic course credits.

6.2.3 Materials and Measures

Students required a PC and an internet connection to access the course. The materials used for the CT course consisted of web-based lectures, exercises, and feedback on exercises (see Appendix E).

The Halpern Critical Thinking Assessment (HCTA; Halpern, 2010) was administered at pre-and-post-testing. The HCTA consists of 25 open-ended questions based on believable, everyday situations, followed by 25 specific questions that probe for the reasoning behind each answer. Questions on the HCTA represent five categories of CT skills: hypothesis testing (e.g. understanding the limits of correlational reasoning and how to know when causal claims cannot be made), verbal

reasoning (e.g. recognising the use of pervasive or misleading language), argument analysis (e.g. recognising the structure of arguments, how to examine the credibility of a source and how to judge one's own arguments), judging likelihood and uncertainty (e.g. applying relevant principles of probability, how to avoid overconfidence in certain situations) and problem-solving (e.g. identifying the problem goal, generating and selecting solutions among alternatives). For an example of a question on the HCTA and how it is scored, see Figure 6.1. These skills, though labelled differently, also reflect the skills identified by the *Delphi Report* (Facione, 1990b). Test reliability is robust, with internal consistency of $\alpha = .88$ (Halpern, 2010). The internal consistency of the assessment in the current study was $\alpha = .73$.

The Need for Cognition Scale (Cacioppo, Petty & Kao, 1984) (short form) consists of 18 items coded on a seven-point likert scale (1 = very strong disagreement, 7 = very strong agreement) that assess one's willingness to explore and engage in relatively complex cognitive activities (e.g. *I would prefer complex to simple problems*" and *"I prefer to think about small, daily projects to long-term ones"*). The estimates of test reliability range from .85 – .90 (Sherrard & Czaja, 1999), and the internal consistency of the scale in the current study was $\alpha = .88$.

The version of the *Motivated Strategies for Learning Questionnaire* (MSLQ; Pintrich et al., 1991) used in this research consisted of 43 items (e.g. *I work hard to do well in class even if I don't like what we are doing*), each of which is responded to using a seven-point likert scale (1 = strongly agree, 7 = strongly disagree). Eight subscales of the MSLQ were used in this study including motivation towards: *elaboration* (i.e. motivation to paraphrase, summarise and/or create analogies to build connections between different items of information); *critical thinking* (i.e. the degree to which students report applying previous knowledge to new situations in order to solve

Suppose that you are a first-year student in a dental school. You realize that your new friend, who is also a first-year student in dental school, is getting drunk on a regular basis several times a week. You do not see any signs of her drinking problem at school, but you are concerned because you will both begin seeing patients at the school's dental clinic within a month. She has not responded to your hints about her drinking problem. As far as you know, no one else knows about her excessive drinking.

Part A: State the problem in two ways.

Scoring: There are two points possible for part A. Please answer the following question(s) in order to score the respondent's answers. Sum the scores from both questions.

Does the respondent's problem statement indicate that the new friend has a drinking problem and will be dealing with patients? Yes = 1 point; No = 0 points

Does the respondent's problem statement indicate that there are no signs that the drinking problem impairs performance? Yes = 1 point; No = 0 points

Part B: For each statement of the problem, provide two differed possible solutions.

Scoring: There are 2 sets of questions for part B. Two points are possible for each set of questions. Please answer the following question(s) in order to score the respondent's answers.

Set 1: Does the respondent suggest informing an authority figure about the problem? Yes = 2 points; No = 0 points

Does the respondent suggest that the friend should not deal with patients? Yes = 2 points; No = 0 points

Set 2: Does the respondent suggest showing the friend how the drinking problem could potentially impair her performance? Yes = 2 points; No = 0 points

Does the respondent suggest convincing the friend that she puts others in danger regardless of whether she knows it or not? Yes = 2 points; No 0 points

Part C: Given these facts, rate each of the following problem statements on a scale of 1 to 7 in which:

- 1 = extremely poor statement of the problem.
- 2 = very poor statement of the problem.
- 3 = poor statement of the problem.
- 4 = statement of the problem that is medium in quality.
- 5 = good statement of the problem.
- 6 = very good statement of the problem.
- 7 = excellent statement of the problem.

1. The friend may cause harm to patients because she is drunk.
2. You are the only one who knows she has a drinking problem.
3. Your friend's parents do not know she has a drinking problem.
4. You need to find a way to give your friend better hints about her drinking.
5. The friend may flunk out of school if she continues to get drunk so often.
6. The friend may hurt herself if she continues to get drunk so often.
7. You feel responsible for your friend's drinking problem.

Scoring: There are seven points possible in part C; one point is possible per question. If the respondent selected any number within the correct range they earn one point. If the respondent selected a number outside the correct range they do not earn a point.

- Question 1: Correct range: 5-7; Question 2: Correct range: 2-5
Question 3: Correct range: 1-2; Question 4: Correct range: 1-4
Question 5: Correct range: 4-7; Question 6: Correct range: 4-7
Question 7: Correct range: 4-7

Figure 6.1: Question 21 on the HCTA (of the Problem-Solving Sub-Scale) with Scoring Protocol (Halpern, 2010)

problems, reach decisions or make critical evaluations); *effort regulation* (i.e. the motivation to control both effort and attention in the face of distractions and uninteresting tasks); *metacognitive self-regulation* (i.e. the motivation to plan, monitor and regulate cognitive processes); *organisation* (i.e. motivation to select information; construct connections among information; and to cluster, outline and select main ideas in reading passages); *control of learning beliefs* (i.e. students' beliefs that their efforts to learn will result in positive outcomes); *intrinsic goal orientation* (i.e. the degree to which the students perceive themselves to be participating in a task for reasons such as challenge, curiosity and mastery); and *extrinsic goal orientation* (i.e. the degree to which the students perceive themselves to be participating in a task for reasons such as grades, rewards, evaluation by others and competition). Internal consistency for sub-scales in the current study ranged from $\alpha = .65 - .88$.

A *Zoom H4n*TM solid-state audio recorder was used to record the focus group interview as a 128 Kbs MP3 audio file. The set of questions used to guide the focus group discussion is provided in Appendix F.

6.2.4 Procedure

The study took place over a period of eight weeks. Two groups took part in the study: those who participated in the e-learning CT course taught through AM ($N = 43$) and a control group who received no CT intervention ($N = 31$). The AM group completed a six-week online CT course in which they viewed classes twice per week; completed two exercise sessions per week; and received detailed feedback for both exercise sessions at the end of each week. Each class presented CT materials (e.g. lecture slides and exercises) to students through AMs. The exercises involved the manipulation of AMs and completion of relevant CT tasks using AMs. Students who participated in the course used the *Rationale*TM AM software or *Microsoft*

PowerPoint[™] to complete their AM exercises and were also encouraged to practice using the *Rationale*[™] programme outside of the course environment. The e-learning course was designed to teach CT according to the framework provided by the *Delphi Report* (Facione, 1990b). Classes focused on training analysis, evaluation, and inference skills, which are core CT skills identified in the *Delphi Report*. The control group did not view or complete any of the CT lectures or materials.

In Week 1, participants were provided with information regarding the nature of the study. Participants were informed that they could withdraw from the study at any time. Also in Week 1, the NCS, the MSLQ, and the HCTA were administered via the *Survey Monkey*[™] website as pre-tests, prior to the commencement of the e-learning course. The course proper began in Week 2. The e-learning classes were voice recorded and dubbed over a *PowerPoint*[™] slideshow using *CamTasia*[™] recording software. Classes lasted a maximum of 15 minutes each, as suggestions made by students in the questionnaire from Study 2 recommended that if full participation was to be expected, the CT classes and exercises together, should not take longer than an hour to complete. In addition, research has shown that didactically teaching students for longer than 15 minutes can substantially decrease attention to the source of instruction (Wankat, 2002). In each class, students were taught to use CT skills via worked examples (in the form of AMs). Students were able to pause, rewind, and restart the class at anytime they wished. In reality, some of the exercises in the latter sessions took longer than an hour to complete, as one of the core objectives of the teaching set was to work from simple to more complex applications of AM throughout the course – a feature of the course that was welcomed by students in retrospect, notwithstanding the amount of time and effort the students needed to put into completing these exercises (see discussion for further details).

From Weeks 2- 7, the CT e-learning course focused on three, core CT skills: analysis, evaluation and inference, which students were taught how to use through worked examples. The lectures presented to students in Study 3 consisted of the same general content as lectures in Study 2. For example, in the analysis-focused lectures, students worked to identify the sources of arguments and the presence of balance and bias in argument structures. Students were also instructed to extract the structure of arguments for analysis using AM structural tools. During the evaluation training, students worked to evaluate the relevance, credibility, and logical strength of arguments, and the overall balance of relevant, credible and logical evidence in an argument through AM. During inference training, students worked to gather the appropriate information from arguments and draw logical conclusions through AM.

The lectures developed for Study 3 differed from those used in Study 2 only in that the Study 3 lectures were designed to be shorter, more concise and clearer. Another change in Study 3 was that the final lecture of the Study 2 course (which acted as a course review), was replaced by an alternative review lecture that focused on reflective judgment (specifically, on the ability to use reflective judgment when applying the skills of analysis, evaluation and inference). Though reflective judgment (RJ) was not explicitly assessed in Study 3, a lecture on RJ was added because it was thought that students with a more metacognitive sensibility, in relation to the attributes of RJ, would demonstrate greater gains in CT ability. The course outline is presented in Table 6.1.

Immediately after each class, students were asked to complete a set of active learning AM exercises and email the completed exercises back to the primary investigator. Engagement in the course was measured according to the amount of

Table 6.1: e-Learning CT Course Outline

Class No.	Title	Content
1	Pre-Testing	<ul style="list-style-type: none"> Students completed the HCTA, MSLQ and Need for Cognition pre-tests.
2	“Introduction to Critical Thinking: Part 1”	<ul style="list-style-type: none"> We think in order to decide what to do and what to believe. We ultimately decide what to believe by adding supports or rebuttals to our own arguments (i.e. questioning our own beliefs). Exercises: <ol style="list-style-type: none"> Adding <i>but</i>s and <i>because</i>s to argument structures. Adding justification and objections to <i>but</i>s and <i>because</i>s regarding one of your own beliefs.
3	“Introduction to Critical Thinking: Part 2”	<ul style="list-style-type: none"> Arguments are hierarchical structures. We can continue to add more levels if we like in order to increase the hierarchical complexity. Exercises: <ol style="list-style-type: none"> Adding more <i>but</i>s to <i>because</i>s in an argument structure. Adding <i>rebuttals</i> to <i>but</i>s.
4	“Unpacking (analysing and evaluating) a person’s belief: Part 1”	<ul style="list-style-type: none"> In order to analyse an argument, we must extract the structure of the argument from dialogue or prose. During extraction, we must be able to identify the function of each proposition (i.e. support, objection or rebuttal). Exercises: <ol style="list-style-type: none"> Extract the argument structure in the dialogue and place each proposition in its appropriate position in the argument structure. Extract the argument structure in the text and place each proposition in its appropriate position in the argument structure.
5	“Unpacking (analysing and evaluating) a person’s belief: Part 2”	<ul style="list-style-type: none"> In order to unpack a person’s belief we must identify the types (sources) of arguments used and consider the strength of each type. The evaluation of the overall strengths and weaknesses of an argument can be completed after adequate analysis. Exercises: <ol style="list-style-type: none"> Identify the source of each proposition and add a rebuttal to each. Identify and evaluate the relevance of each proposition and exclude each that is irrelevant. Answer the series of questions about the argument you just evaluated.

Table 6.1: e-Learning CT Course Outline (continued)

Class No.	Title	Content
6	“Analysis & Evaluation: Part 1”	<ul style="list-style-type: none"> • Analysis of argument structures through the extraction of supports, objections and rebuttals from arguments. • Re-creation of argument structures after analysis and evaluation. • Recognition of imbalances, omissions and bias within an argument. • Examining whether or not the arguments used are relevant or logically connected to the central claim is also an important factor in evaluation. • Exercises: <ol style="list-style-type: none"> 1. Read & analyse the argument structure. Fill in the template with missing propositions from information in the text. 2. Make a list of your observations based on any imbalances, omissions, bias in the argument Answer the series of questions about the argument you analysed.
7	“Analysis & Evaluation: Part 2”	<ul style="list-style-type: none"> • It is sometimes difficult to establish the truth of a claim. • Evaluating sources of arguments, whether or not the arguments used are relevant; and evaluating the possibilities of imbalance, omission or bias. • Exercises: <ol style="list-style-type: none"> 1. Read & analyse the text-based argument. Complete the argument structure from the information in the text. 2. Examine as best you can whether or not the arguments used are <i>relevant and logically connected</i> to the central claim. Examine the argument by looking for <i>imbalances, omissions, bias</i> in the argument.
8	“Evaluation: Part 1”	<ul style="list-style-type: none"> • We must evaluate: <ol style="list-style-type: none"> 1. Types (sources) of arguments based on credibility 2. The relevance of propositions to the central claim or intermediate conclusions within the argument 3. The logical strength of an argument structure • Exercises: <ol style="list-style-type: none"> 1. Read the argument map and assess the relevance and credibility of each proposition. Eliminate propositions that are irrelevant and those that lack credibility; and suggest alternative arguments wherever possible. 2. Place each proposition in its correct place within its correct position in the argument map. There are two core argument structures (each containing 3 propositions) that support the central claim. However, in each set of three, two of the three propositions support the third. Please decide how best to arrange the propositions in the argument map. That is, which two propositions support the third in each case?

Table 6.1: e-Learning CT Course Outline (continued)

Class No.	Title	Content
9	“Evaluation: Part 2”	<ul style="list-style-type: none"> • Evaluating credibility after the identification of argument sources. • We must evaluate the balance of evidence within an argument structure. • Exercises: <ol style="list-style-type: none"> 1. Evaluate the argument structure on attraction. Take time to do each of the following: (1) assess the credibility of arguments; assess the relevance of arguments; assess the logical strength of the argument structures; assess the balance of evidence in the argument structure. 2. If a proposition lacks credibility, relevance, or logical strength, exclude it and explain why you did so. Also, explain the balance of the argument.
10	“Inference: Part 1”	<ul style="list-style-type: none"> • Evaluation and inference are intimately related. • Inference differs from evaluation in that the process of inference involves generating a conclusion from previously evaluated propositions. • Inferring conclusions with syllogisms. • Exercises: <ol style="list-style-type: none"> 1. Assess each syllogism. 2. Identify whether each syllogism is valid (10 syllogisms).
11	“Inference: Part 2”	<ul style="list-style-type: none"> • In larger informal argument structures, intermediate conclusions can be inferred. • Intermediate conclusions can be used to infer a central claim. • Exercises: <ol style="list-style-type: none"> 1. Infer 3 intermediate conclusions based on the information within the argument structure. 2. Work from the bottom up to infer each intermediate conclusion and the central claim (3 inferences).
12	“Reflective Judgment: Part 1”	<ul style="list-style-type: none"> • Reflective judgment is our ability to reflect upon what we know and the knowledge the world presents us; and our ability to think critically and reflectively in this context. • Understanding of the nature, limits and certainty of knowing and how this can affect our judgment. • Recognition that some problems cannot be solved with absolute certainty (i.e. ill-structured problems). • Exercises: <ol style="list-style-type: none"> 1. Read the passage and answer the two questions about the passage beneath. Each answer will be a solution to the problem presented in the passage. Please use the argument maps to structure your answers. 2. In the argument maps, you are asked to provide an (1) answer; (2) one reason below to support your answer; (3) two reasons for the support you provided above and (4) two alternative answers to the problem.

Table 6.1: e-Learning CT Course Outline (continued)

Class No.	Title	Content
13	“Reflective Judgment: Part 2”	<ul style="list-style-type: none"> • The importance of structure and complexity in reflective judgment. • We can analyse and evaluate information when there is uncertainty and then organise our findings in order to infer a conclusion and exhibit the reflective judgment in our thinking. • Exercises: <ol style="list-style-type: none"> 1. Read the short passage on “Violent Television” and reflect upon this issues and what you know about the topic. 2. Analyze the arguments derived from your own knowledge and think about whether or not what you think amounts to propositions of the following type: anecdotes, common sense, research findings, or authority opinions. 3. Research the topic to aid in the development of your argument. 4. Evaluate the relevance, credibility, and logical strength of any arguments you gather as part of your investigation, and reflect upon what the ‘world’ presents you in this context and how it is related to what you already know about the topic. 5. Consider whether or not you were biased in your search for reasons and objections (i.e. whether or not you analyzed and evaluated both confirming and disconfirming evidence in relation to central claim).
14	Post-Testing	Students completed the HCTA, MSLQ and Need for Cognition post-tests

exercises emailed to the primary investigator. The engagement scores ranged from 0 to 24. Feedback was provided to students at the end of each working week, that is, after they had completed and returned two sets of exercises. Feedback focused on the structure of arguments provided by students; inferential relationships among propositions in their arguments; and the relevance and credibility of the propositions they used (again, see Appendix E for sample feedback provided to students). In Week 8, after the completion of the e-learning CT course, the HCTA, the MSLQ and the NCS were again administered via the *Survey Monkey*TM website as post-tests. After the completion of the CT intervention, a focus group interview was conducted with a

sub-set of the full sample of participants, after which all participants were debriefed and thanked.

6.3 Results

Means and standard deviations for scores on overall CT, all CT sub-scales, need for cognition and motivation are presented in Table 6.2. Table 6.3 presents a summary of the ANOVAs conducted in the current study. Table 6.4 presents inter-correlations between all dependent measures included in this study.

6.3.1 Group differences in critical thinking, need for cognition, and motivation

A series of 2 (time: pre-and-post-testing) x 2 (condition: AM group and control group) Mixed ANOVAs was conducted to examine the effects of both time and CT training condition on need for cognition, motivation, CT and CT sub-skills. Post hoc analyses were conducted for all significant main and interaction effects. Only those effects that were significant after Bonferroni correction are presented below. Results revealed a main effect of both time and condition on overall CT ability (see Table 6.3). Both groups scored higher at post-testing when compared with pre-testing (see Table 6.2). Participants in the AM group also scored higher overall on CT ability when compared with the control group. Notably, this difference between the two groups was largely accounted for by post-test differences. However, the condition x time interaction effect was not significant.

An examination of CT sub-skills revealed the following pattern of results: there was a significant effect of time on hypothesis testing, verbal reasoning, likelihood/uncertainty and problem-solving. In each case, it was found that performance improved from pre-test to post-test. There was no effect of time on argument analysis. However, there was a condition x time interaction effect for argument analysis, with a significant improvement in the AM group from pre-test to

Table 6.2: Means and standard deviations for CT Performance by Condition

	<i>N</i>	<u>Pre-Test</u>		<u>Post-Test</u>	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<u>Overall CT</u>					
AM	43	96.42	15.03	108.35	14.57
Control	31	87.55	24.72	98.58	14.05
<u>Hypothesis Testing</u>					
AM	43	22.26	4.07	24.44	4.55
Control	31	21.10	4.53	21.90	5.96
<u>Verbal Reasoning</u>					
AM	43	6.40	1.99	7.56	2.68
Control	31	5.71	1.94	6.03	2.35
<u>Argument Analysis</u>					
AM	43	22.67	3.83	24.30	4.40
Control	31	22.00	5.01	21.00	4.58
<u>Likelihood/Uncertainty</u>					
AM	43	10.07	3.59	12.44	3.73
Control	31	9.07	3.44	10.29	4.02
<u>Problem-Solving</u>					
AM	43	36.42	4.10	39.36	3.70
Control	31	36.21	4.69	39.61	5.75
<u>Need for Cognition</u>					
AM	43	67.00	14.87	67.26	16.45
Control	26	65.85	19.10	65.08	15.65
<u>Motivation</u>					
AM	41	170.98	28.46	164.88	33.77
Control	26	181.73	24.15	174.81	25.49

post-test ($t = -2.14$, $df = 42$, $p = .038$, two tailed, $d = .40$), and a non-significant decline in the control group from pre-test to post-testing ($p = .248$; see Table 6.2).

A series of paired samples t-tests further revealed that the AM group scored significantly higher on post-testing compared with pre-testing on overall CT ability ($t = -6.65$, $df = 42$, $p < .001$, two tailed, $d = .81$) and all other CT sub-scales: hypothesis testing ($t = -3.89$, $df = 41$, $p < .001$, two tailed, $d = .55$), verbal reasoning ($t = -2.97$, $df = 41$, $p = .005$, two tailed, $d = .49$), likelihood/uncertainty ($t = -4.64$, $df = 41$, $p < .001$, two tailed, $d = .67$) and problem-solving ($t = -4.47$, $df = 41$, $p < .001$, two tailed, $d = .64$). Conversely, students in the control group scored significantly higher on the post-test compared with the pre-test on overall CT ability ($t = -3.01$, $df = 30$, $p = .005$, two tailed, $d = .55$) and for the problem-solving sub-scale ($t = -3.77$, $df = 27$, $p = .001$, two tailed, $d = .65$).

There was a main effect of condition on verbal reasoning and argument analysis, with the AM group scoring significantly higher than the control group. There was a similar trend towards better performance in the AM group relative to the control group for hypothesis testing ($p = .089$) and unlikelihood/uncertainty ($p = .063$). The significant differences between the experimental and control conditions were accounted for largely by differences in post-test performance. Specifically, there were no statistically significant differences between the experimental and control groups on any of the five sub-skills at pre-test; however, those in the AM group scored significantly higher than those in the control condition at post-test on the sub-skills of verbal reasoning ($t = -2.55$, $df = 72$, $p = .013$, two tailed, $d = .61$) and argument analysis ($t = -3.13$, $df = 72$, $p = .003$, two tailed, $d = .74$).

Results also revealed that there was no effect of condition or time on need for cognition, and no condition x time interaction effect. Similarly, there was no effect of

Table 6.3: ANOVA Summary for Study 3

	df	df (error)	<i>F</i>	<i>p</i>	Partial η^2
<u>Overall CT</u>					
Condition	1	72	6.72	.012	.09
Time	1	72	37.42	.000	.34
Condition x Time	1	72	.06	.812	.00
<u>Hypothesis Testing</u>					
Condition	1	68	2.98	.089	.04
Time	1	68	11.40	.001	.14
Condition x Time	1	68	1.49	.227	.02
<u>Verbal Reasoning</u>					
Condition	1	68	5.50	.022	.08
Time	1	68	6.73	.012	.09
Condition x Time	1	68	2.17	.145	.03
<u>Argument Analysis</u>					
Condition	1	72	5.25	.025	.07
Time	1	72	.30	.586	.01
Condition x Time	1	72	5.23	.025	.07
<u>Likelihood/Uncertainty</u>					
Condition	1	68	3.58	.063	.05
Time	1	68	18.88	.000	.22
Condition x Time	1	68	1.10	.299	.02
<u>Problem-Solving</u>					
Condition	1	68	.10	.748	.00
Time	1	68	37.04	.000	.35
Condition x Time	1	68	.04	.845	.00
<u>Overall Motivation</u>					
Condition	1	65	2.24	.139	.03
Time	1	65	8.63	.005	.12
Condition x Time	1	65	.04	.853	.00
<u>Need for Cognition</u>					
Condition	1	67	.34	.565	.01
Time	1	67	.23	.635	.00
Condition x Time	1	67	.41	.526	.01

condition on motivation, and no condition x time interaction effect. However, there was a main effect of time on motivation, with total motivation scores decreasing from pre-test ($M = 175.15$; $SD = 27.20$) to post-test ($M = 168.73$; $SD = 31.00$) in the sample

as a whole. Analysis of the motivation sub-scales revealed that there was a main effect of time on effort regulation, with a significant decrease in effort regulation from pre-to-post-testing, $F(1, 65) = 17.42, p < .001$, partial $\eta^2 = .21$, in both the AM group ($t = 2.96, df = 40, p = .005$, two tailed) and the control group ($t = 3.01, df = 25, p = .006$, two tailed). No other effects were observed.

6.3.2 High-engagement versus low engagement differences in the AM group

In order to examine the effects of engagement on performance in the AM training group, a series of eight 2 (time) x 2 (levels of engagement: high [12-24 exercises completed; $N = 19$] and low [0-11 exercises completed; $N = 24$]) Mixed ANOVAs were conducted using CT performance, CT sub-skills performance, need for cognition, and motivation towards learning strategies as separate dependent variables. There was no main effect of engagement level on need for cognition, motivation, or CT performance. However, there was a significant time x engagement level interaction effect for the CT sub-scale of problem-solving, $F(1, 40) = 9.26, p = .004$, partial $\eta^2 = .19$, with those in the high-engagement group scoring higher on post-test ($M = 41.21, SD = 5.77$) when compared to pre-test ($M = 36.00, SD = 3.71; t = -5.41, p < .001$); and with a smaller, albeit significant, gain in problem-solving ability in the low-engagement group (pre-test $M = 37.00, SD = 4.37$; post-test $M = 38.74, SD = 5.26; t = -2.20, p = .039$).

Further analysis was conducted to examine differences between high-and-low-engagement groups on motivation sub-scales at pre-test and post-test. There was a significant decrease in effort regulation in both the low-engagement group ($t = 2.29, df = 21, p = .033$, two-tailed; pre-test: $M = 21.83, SD = 8.40$; post-test: $M = 19.87, SD = 8.37$) and the high-engagement group ($t = 2.16, df = 19, p = .044$, two-tailed; pre-test: $M = 24.21, SD = 6.38$; post-test: $M = 22.47, SD = 5.08$). Further analysis revealed

that there was no difference between the high-and-low-engagement groups on any of the motivation sub-scales at either pre-testing or at post-testing.

6.3.3 Correlations

There was a significant correlation between need for cognition and motivation at pre-testing ($r = .52, p < .001$) and at post testing ($r = .60, p < .001$), but neither need for cognition nor motivation were correlated with overall CT performance at pre-testing. There was a significant correlation between CT performance and both need for cognition ($r = .47, p < .001$) and motivation ($r = .28, p = .017$) at post-testing. The full set of correlations among CT, need for cognition and motivation sub-scales at pre-testing and post-testing are presented in Table 6.4. Results from a hierarchical multiple regression analysis revealed that the predictors - need for cognition ($\beta = -.09, p = .510$) and motivation ($\beta = .08, p = .541$) in block 1 (pre-testing) and block 2 (post-testing) - did not account for any variance (adjusted $r^2 = .02$) in CT over and above the effect of experimental condition, $F(2, 62) = .71, p = .749$. With regards to post-intervention intra-group correlations between CT and motivation, the CT performance of both AM and control groups were significantly, positively correlated with motivation towards elaboration ($r = .35, p < .05; r = .60, p < .001$, respectively) and motivation towards CT ($r = .35, p < .05; r = .37, p < .05$, respectively). Results also revealed that CT performance of those in the AM group was significantly correlated with overall motivation ($r = .35, p = .024$), intrinsic goal orientation ($r = .47, p = .002$) and control of learning beliefs ($r = .38, p = .012$). Conversely, overall CT performance of those in the control group was significantly correlated with motivation towards organisation ($r = .38, p = .038$).

Table 6.4: Correlations Among CT performance, Need for Cognition and Motivation Sub-scales at Pre-testing (below diagonal) and Post-testing (above diagonal).

	1	2	3	4	5	6	7	8	9	10	11	12.	13	14	15	16
1 CT Performance	-	.81 ³	.62 ³	.72 ³	.59 ³	.68 ³	.47 ³	.28 ¹	.28 ¹	.24 ¹	-.07	.02	.24 ¹	.15	.36 ²	.40 ³
2 Hypothesis Testing	.69 ³	-	.44 ³	.45 ³	.43 ³	.39 ³	.44 ³	.14	.15	.14	-.15	.03	.12	.04	.21	.33 ²
3 Verbal Reasoning	.48 ³	.24 ¹	-	.28 ¹	.18	.46 ³	.36 ²	.30 ²	.26 ¹	.12	-.13	.11	.26 ¹	.31 ²	.40 ³	.38 ³
4 Argument Analysis	.27 ¹	.12	.21	-	.33 ²	.32 ²	.30 ²	.17	.19	.20	-.06	-.07	.20	.09	.24 ¹	.20
5 Likelihood & Uncertainty	.58 ³	.30 ¹	.17	.28 ¹	-	.13	.19	.13	.05	.12	.06	.07	.12	.05	.14	.14
6 Problem-Solving	.72 ³	.41 ³	.27 ¹	.31 ²	.13	-	.33 ²	.28 ¹	.33 ²	.23	.02	.08	.19	.14	.32 ²	.37 ³
7 Need for Cognition	.17	.16	.12	.25 ¹	.30 ¹	.19	-	.60 ³	.66 ³	.44 ³	.20	.29 ²	.46 ³	.47 ³	.49 ³	.57 ³
8 Overall Motivation	.03	-.11	-.01	.02	.11	.21	.52 ³	-	.74 ³	.59 ³	.37 ²	.78 ³	.89 ³	.80 ³	.81 ³	.85 ³
9 Intrinsic Goal Orientation	.01	.17	-.09	.06	-.06	.24	.50 ³	.68 ³	-	.52 ³	.20	.48 ³	.57 ³	.46 ³	.67 ³	.60 ³
10 Control of Learning Beliefs	.12	.14	.08	.16	.11	.40 ³	.40 ³	.52 ³	.56 ³	-	.34 ²	.29 ^{2*}	.40 ³	.40 ³	.38 ³	.48 ³
11 Extrinsic Goal Orientation	.03	-.04	-.06	-.18	.01	-.01	.21	.46 ³	.30 ²	.37 ²	-	.28 ¹	.16	.15	.13	.12
12 Organisation	-.14	-.23	.03	-.03	.02	.05	.30 ²	.73 ³	.38 ³	.18	.16	-	.67 ³	.56 ³	.53 ³	.68 ³
13 Metacognitive Self-Regulation	.03	-.11	.08	.03	.09	.20	.36 ²	.86 ³	.50 ³	.30 ²	.20	.65 ³	-	.73 ³	.72 ³	.71 ³
14 Effort Regulation	-.02	-.22	-.04	.05	.17	.13	.33 ²	.75 ³	.37 ²	.24 ¹	.35 ²	.50 ³	.63 ³	-	.59 ³	.64 ³
15 Motivation Towards CT	.04	-.08	-.01	.06	.11	.12	.54 ³	.74 ³	.52 ³	.29 ¹	.12	.54 ³	.62 ³	.33 ²	-	.67 ³
16 Elaboration	.14	-.06	-.02	-.02	.08	.19	.39 ³	.82 ³	.48 ³	.32 ²	.16	.67 ³	.71 ³	.51 ³	.73 ³	-

Significance levels ¹ = p at the .05 level; ² = p at the .01 level; ³ = p at the .001 level.

6.3.4 Focus Group

After completion of the CT intervention, a focus group interview was conducted in order to investigate students' perceptions of the CT e-learning course, with specific focus on the presentation and quality of lectures, exercises and feedback provided to students, as well as the student's experience in the use of AM to facilitate CT ability. The interview transcripts were examined using thematic analysis. Overall, students interviewed in the focus group were very pleased with (1) AM as a method of presenting the CT course materials, (2) the CT course in general and (3) the likelihood that they will be able to apply what they learned in the course to future experiences. However, the students interviewed did not respond favourably to the feedback provided to them on their exercises, nor the usability of *Microsoft PowerPoint™* to complete their AM exercises. Rather than present a full results section of the thematic analysis, it was decided to present the findings in the Study 3 discussion (i.e. section 6.4) in order to fulfil the purpose of the focus group interview, which was to shed light upon the quantitative findings and the implications of these findings in light of the student learning experience.

6.4 Discussion of Study 3

6.4.1 Interpretation of Results

The aim of Study 3 was to test four hypotheses: first, that AM training delivered using an e-learning CT course would significantly enhance CT performance in comparison with a control condition; second, that students who engaged more with the e-learning course would perform significantly better on CT performance than those who did not engage as much; third, that CT performance would be positively correlated with both dispositional need for cognition and motivation at both testing times; and fourth, that the potential, significant correlations between CT and

motivation towards learning strategies would be higher in the AM group than in the control group. The results of Study 3 revealed that students in the AM group scored significantly higher on post-testing than on pre-testing on measures of overall CT ability and on all CT sub-scales. Results also revealed that those in the control condition improved from pre-to-post-testing on overall CT ability and the problem-solving sub-scale of the HCTA. It is possible that the improvement in overall CT of those in the control group was largely a function of the increase in problem-solving ability, given that it was the only CT sub-scale where significant improvement was observed in the control group from pre-to-post-testing. The most likely cause of improvement in problem-solving and overall CT ability of the control group participants, all of whom were first year college students, is maturation and engagement in third level studies. This interpretation is consistent with research by Pascarella and Terenzini (1991), who found that the experience of college has beneficial effects on CT. Alternatively, it could be that the improvements in CT ability observed in the control group were a result of practice effects (i.e. repeat administration of the HCTA).

At the same time, however, positive effects of the AM training course were observed over-and-above any possible maturation or practice effects. Specifically, results revealed that those in the AM group scored significantly higher at post-testing than those in the control group on overall CT ability and the CT sub-scales of verbal reasoning and argument analysis. Given that there were no significant differences between the control and AM groups at the pre-testing stage on either overall CT or on any CT sub-scales, these findings suggest that the two groups were adequately matched on CT ability prior to the intervention and that AM training significantly enhanced CT performance.

Another interesting finding was the condition x time interaction effect for argument analysis. Specifically, results revealed that those in the CT group scored significantly higher on argument analysis at post-testing than at pre-testing, whereas the control group scored better (though not significantly) at pre-testing when compared with post-testing. These results suggest that while the AM group got significantly better at argument analysis over time, those in the control group, on average, got worse. Though it seems logical to consider that the decrease in argument analysis ability of those in the control group may have been caused in part by the significant reduction in motivation towards effort regulation, hierarchical regression results indicated that motivation towards effort regulation from pre-to-post-testing did not account for any variance in argument analysis ($\beta = .07$ $p = .797$). Thus, even though the overall CT ability of controls increased over the duration of the intervention, it is possible that the control group's decrease in argument analysis ability may simply have resulted from a lack of support or training in argument analysis during the intervention.

Strong evidence in favour of the second hypothesis was not found. Specifically, there was no difference between those who engaged more with exercises (i.e. those who completed 12-24 CT exercises) and those who engaged less (i.e. those who completed 0-11 CT exercises) on overall CT ability or on any of the CT sub-skills. However, there was a time x engagement interaction effect for the CT sub-skill of problem-solving, with those in the high-engagement group showing a greater gain in problem-solving ability than those in the low-engagement group. According to the HCTA manual (Halpern, 2010, p. 7), "problem-solving involves the use of multiple problem statements to define the problem and identify possible goals, the generation and selection of alternatives, and the use of explicit criteria to judge among

alternatives.” Problem-solving as defined by Halpern is akin to the CT sub-skill of inference, as defined in the *Delphi Report* (Facione, 1990b). Though results revealed that problem-solving can develop over time without engagement or explicit training in CT, as exemplified by those in the control group, the results of the current study further indicate that problem-solving skills may develop more in the context of high levels of engagement with active exercises in a CT training course.

One possible explanation for the overall positive effect of AM training condition but a weaker effect of exercise engagement may be due to the high quality of the lectures and/or the feedback posted on the course web-site (available to all students in the AM group, regardless of whether or not they completed more or less exercises), which may have provided students with enough engagement with AM to perform well. Also, given that the average level of engagement among all students who participated in the course was 11.67 (out of 24), with only 20.93% of the sample completing zero exercises, it is important to note that the majority of students engaged beyond simply viewing online lectures, including the majority of students in the low-engagement group.

Upon further analysis, the CT performance of those in the high-engagement group was positively correlated with motivation ($r = .38$; $p = .037$), as well as with need for cognition ($r = .49$; $p = .005$). In addition, though it was found that the CT performance of those in the low-engagement group was positively correlated with their need for cognition ($r = .38$; $p = .022$), CT performance in this group was not correlated with motivation. These findings indicate that motivation may have been a key feature of performance in the high engagement group, with motivation and performance being more closely coupled in the high-engagement group but not in the low-engagement group.

In the follow-up focus group interview, it was found that low-engagement on the part of some of the students, and possibly attrition, may have been a result of the manner in which the exercises were provided and the way in which students were instructed to complete them. Specifically, students were provided with instructions on how to complete each exercise at the end of every lecture recording and provided with a link to a *Microsoft PowerPoint*TM slideshow that contained the necessary materials to complete the exercises. Students were encouraged to use the *Rationale*TM AM software to complete these exercises, given that it is designed specifically for supporting AM construction. However, they were also provided with the opportunity to complete the exercises through *Microsoft PowerPoint*TM. The latter opportunity was provided for the simple reason that licenses for *Rationale*TM were provided for on-campus computers only; thus, if students found the time in their schedules to complete the AM exercises at home, they would not be able to access *Rationale*TM. As a result, the opportunity to complete their exercises through *PowerPoint*TM was provided in the event that some students wished to complete them at home. Unfortunately, as suggested in the focus group interview, providing students with the choice between the two software programmes for purposes of completing the exercises proved problematic; and most participants completed their exercises through *PowerPoint*TM. For example, participant CR5 chose to complete the exercises through *PowerPoint*TM for reasons of familiarity with the programme:

My own thought on it was like, pick one or the other, you know like, say OK, we're going to do this in PowerPointTM, everyone is using the same structure, everyone is using the same content, it's easy to use, everything flows, rather than saying you can do it this way or you can go to a PC suite and log in and use a whole new software that you're not familiar with. That was a fear that I had doing it - that, I don't know RationaleTM, but I'm fairly useful with PowerPointTM. Yet on PowerPointTM I struggled and now, like, do I really need to go off and learn a whole new software application to try and do this critical thinking course, so the way I dealt with it, I did them as best I could in PowerPointTM and submitted.

The ‘struggle’ mentioned by CR5 referred to the difficulty in constructing AMs in *PowerPoint*TM. For example:

I thought the PowerPoint slides were very difficult to manage. The stuff, like the actual content was fine, but it’s actually managing the editing of the PowerPointTM slides that was quite difficult.

Similarly, participant AO7 also found difficulty in using *PowerPoint*TM to complete the exercises:

The graphic of trying to move boxes around, you pick it, you pull it, didn’t necessarily work and that really took away from the purpose of the exercise and you wasted a lot of time just trying to get boxes in the right places.

These suggestions indicated, that though sufficient support in how to use *Rationale*TM was provided to students (i.e. links to instructional videos were provided on the course website), they felt, in general, that learning to use new software was an unnecessary use of their time, especially in the context of a voluntary course. However, at the same time, they felt that their ability to construct AMs was limited in doing so through *PowerPoint*TM. Participants further indicated that low-engagement may have been a result of their reported difficulties in completing exercises through *PowerPoint*TM.

Another possible reason for low-engagement and attrition, indicated by the focus group, may have been the length of time it took to complete the exercises. As mentioned above, some of the exercises in the latter sessions actually took longer than an hour to complete, as reported in the focus group. Though one of the core objectives of the course was to develop CT by scaffolding students to work from simple to more complex applications of AM throughout the course (i.e. a feature that was welcomed by participating students), not all students were entirely happy about the possibility of needing additional time to complete the more complex exercises. For example, CR5 stated that:

You should mention that [exercises] will take you more time than 20 minutes. I spent up to nearly two hours on one of them trying to figure it out. That was just the exercise. That wasn't including watching the lecture. The earlier ones were very simple - did them in a couple of minutes, didn't even really think about the first two; [but,] as they went along they got more complex.

Participant EF9 added:

I think the problem there was that you were using the same... the exercise will take you 15 minutes - the same strategy of preparing the student. But they did increase in complexity and perhaps in the first two lectures, yes, this will only take you 15 minutes; but then you need to 'up it' and that this will in fact take you 30 or 40 minutes. Is it necessary to put the time factor down? Why not just let them go through it in their own pace? I think that [the time factor] puts people under pressure.

Notably, even in the two accounts presented above, there is considerable variation in the time estimates provided by students in relation to how long it took them to complete more complex exercises (i.e. 30 minutes – 2 hours). However, rather than proposing to students how long exercise would take, based on the suggestions made by the focus group participants, students should have been simply encouraged to take their time and progress on the exercises at their own pace. As such, while it is possible that some students disengaged as exercises became more complex and took longer to complete, it is important to note that the bulk of loss of participants occurred between the times that students signed up for the course and completion of the pre-testing phase, and not necessarily in response to the exercises (i.e. 52% of attrition occurred during this time frame). As noted by some members of the focus group, many participants also struggled with completing pre-test assessments through *Survey Monkey*. For example, though *Survey Monkey*TM includes a *saving* function, many students seemed to have trouble saving their work and were then required to start over again in order to complete all pre-test measures. Focus group participants noted that this frustrated many of their classmates and this in turn may explain why the bulk of loss of participants occurred between the times specified above.

Study 3 also explored the relationship between need for cognition, motivation towards learning and CT ability in the sample as a whole. It was hypothesised that both need for cognition and motivation towards learning would be significantly, positively correlated with CT performance. It was observed that though there was a significant correlation between need for cognition and motivation at pre-testing and post-testing, neither was correlated with overall CT performance at pre-testing. Notably, however, need for cognition was correlated with the HCTA sub-scales of argument analysis and likelihood and uncertainty at pre-testing, as were motivation regarding control of learning beliefs and the HCTA sub-scale of problem-solving. Interestingly, the general pattern of results regarding the correlation between CT performance and dispositional factors are consistent with findings from Study 2, where it was found that overall CT ability was positively correlated with disposition towards thinking at post-testing, but not at pre-testing. The significant correlation between CT performance and both need for cognition and motivation towards learning at post-testing is also consistent with previous research reporting a positive relationship between CT performance and both need for cognition and motivation (e.g. Garcia, Pintrich & Paul, 1992; Halpern, 2006; Jensen, 1998; Valenzuela, Nieto & Saiz, 2011).

Notably, the positive correlation between overall motivation to learn and CT ability at post-testing was accounted for by five of the eight MSLQ subscales: intrinsic goal orientation, control of learning beliefs, metacognitive self-regulation, motivation towards critical thinking, and motivation towards elaboration. Motivation towards elaboration (i.e. motivation to paraphrase, summarise and/or create analogies to build connections between different items of information) was the motivation sub-scale with the highest correlation with CT performance at post-test, possibly due to

good critical thinkers choosing to build more connections, better connections, or possibly more intricate, complex connections among propositions within arguments. Motivation towards CT (i.e. motivation to apply knowledge to new situations in order to make evaluations, solve problems and/or reach decisions) was the motivation subscale with the second highest correlation with CT performance, which is perhaps unsurprising, given that the HCTA requires the application of knowledge to problem situations in order to make evaluations and reach a decision in relation to key probe questions.

However, results also indicated that though the CT performance of both groups was significantly, positively correlated with motivation towards elaboration and motivation towards CT, there was a higher correlation between both forms of motivation and the CT performance of the control group than the AM group. Furthermore, the CT performance of controls was also significantly correlated with motivation towards organisation, whereas the CT performance of the AM group was not. With respect to the control group, apart from the pre-screening experience itself, or a post-screening reflection period, the novel learning experience of the first year student at university was the only other significant educational factor that may have caused the increased coupling of particular motivations and CT ability over time, and this is perhaps unsurprising as most first year courses challenge students to think in new and different ways, and critically, about information that is being presented to them in lectures. At the same time, results also revealed that overall motivation, intrinsic goal orientation and control of learning beliefs were all significantly correlated with the CT performance of those in the AM group and not for those in the control group at post-testing. Results suggest that though CT training was not the critical factor binding CT ability and all motivations towards learning, CT training

may reinforce the coupling of CT ability with some aspects of motivation, namely motivation towards control of learning beliefs and intrinsic goal orientation.

The above findings are interesting to consider, given that there seems to be a consistent pattern of correlations between CT ability and dispositional factors across Studies 2 and 3, with disposition towards thinking, need for cognition and motivation towards learning all being significantly correlated with CT ability at post-testing, but not at pre-testing. At the same time, these findings are not entirely consistent with previous research, because contrary to the current findings, even in the absence of training, CT ability has been shown to be correlated with motivation and CT dispositions (e.g. Colucciello, 1997; Facione, Facione and Sanchez, 1994; Garcia, Pintrich & Paul, 1992; Halpern, 2006; Hattie, Biggs & Purdie, 1996; Profeto-McGrath, 2003).

Interestingly, in the current study, though students' need for cognition did not change over time in either the experimental or control groups, their motivation to learn significantly decreased over time. Upon closer analysis it was found that this global reduction in motivation was accounted for by a significant reduction in one of the eight motivation sub-scales: effort regulation. Notably, effort regulation was not correlated with CT ability, and it was seen to reduce in a similar way from pre-test to post-test in both the AM and control groups. Therefore, the decrease in effort regulation did not appear to be a function of variation in CT ability or a function of the CT training course; it may have been a result of some factor outside of the course itself. For example, as motivation towards effort regulation refers to the motivation to self-regulate effort and attention in the face of distractions and uninteresting tasks, it may be that those who participated in this course were distracted by requirements in other courses (i.e. impending examinations and class assignments), which made them

less interested in this voluntary CT course. However, at the same time, as this decline in effort regulation was seen in both the AM group *and* the control group, it may be that there was a general decrease in motivation towards effort regulation from early to late in the first semester of the first year at college. Specifically, it may be that in general, students' workload demands had a negative influence on effort regulation over time; or perhaps the novelty of being in college 'wore off' and students began to lose interest in maintaining their initial levels of effort.

Although the relationship between need for cognition, motivation towards learning, and CT ability changed from pre-test to post-test, the results of the current study revealed that there was no effect of AM training on need for cognition or motivation towards learning. This finding suggests that differences between the AM and control groups on CT performance were not caused by changes in students' dispositional need for cognition or motivation to learn. The results of regression analysis further clarified that need for cognition and motivation did not account for any variance in CT over and above the effect of experimental condition.

6.4.2 Limitations & Future Research

Though this study revealed that CT performance can be significantly enhanced by participation in an AM-infused CT e-learning course, there are some limitations that must be considered. Like Study 2, one limitation of the current study was the issue of attrition. With a prospective pool of approximately 1,200 first year undergraduate students, 720 of whom were psychology students, only 247 students volunteered to complete the preliminary testing sessions. Of the 156 students who *actually completed* the pre-testing session, only 74 completed post-testing after the intervention period. Again, it is worth noting that there were no differences between completers and non-completers on need for cognition or on motivation scores at the

pre-testing stage. These findings indicate that a lack of motivation or dispositional need for cognition was not accounting for attrition and that attrition may be more dependent on other personal factors (e.g. personal reasons, not having enough time to take part, or preferring to use any extra time between mandatory lectures for study; which is consistent with reports from non-completers in Study 2).

In order to overcome problems of attrition, future research might aim to implement and evaluate CT interventions in the context of a mandatory course, as opposed to a voluntary course (as employed in Studies 2 and 3). Although psychology students who participated in this study were promised credits towards their overall 1st Year Psychology mark and a certificate of completion, it seems that this was not enough to keep all participants involved in the study. By making such a CT intervention mandatory, attrition would have been significantly reduced and motivation (i.e. effort regulation) may have increased rather than decreased over time.

Though the AM group was compared with a control group, another limitation of this study was that the AM group was not compared with another CT training condition (as in Study 2). It was decided to include just two groups in Study 3 because it was predicted that a relatively high attrition rate would diminish the sample size and that the focus should therefore be on the comparison between an AM and a control group, which was deemed the most important comparison. While including a control group for comparison purposes is important for all AM-infused CT intervention studies, it is also important that future research include other active control conditions that involve training of CT skills using more traditional means, or alternate conditions where AM practice, type of feedback, or course delivery strategy is manipulated. For example, as discussed in previous chapters, although Alvarez-Ortiz's (2007) meta-analysis found that courses where there was "lots of argument mapping practice"

produced a significant gain in students' CT performance, with an effect size of .78 SD, CI [.67, .89], students who participated in CT courses that used at least some AM within the course achieved gains in CT ability with an effect size of .68 SD, CI [.51, .86]. Thus, while the amount of AM practice may be a significant variable worth manipulating in future intervention comparison studies, it appears that researchers will need to think carefully about how to maximize the benefits of AM practice and also test their various hypotheses in the context of realistic experimental designs that include reasonable sample sizes for the purpose of comparisons.

The comparison of multiple AM training strategies within a single study design, for example, where levels of AM practice (i.e. low versus high), AM feedback (i.e. feedback versus no feedback) and the delivery of AM (i.e. e-learning versus classroom) are manipulated, requires availability to larger samples of students. Though it is evident that a mandatory CT course would have yielded a larger sample in the current study context; unfortunately, at the time of this study, it was not feasible to make the CT training course mandatory, due to the existing curriculum structure across the 1st year Arts programme.

Future research on CT interventions could also move beyond measuring CT performance according to standardized tests that rely upon quantitative assessment, to include qualitative analyses of how students come to answer CT questions/problems. For example, research by Ku and Ho (2010b) examined students' CT ability using the HCTA, but did so by asking each student to 'talk aloud' when critically thinking about each question on the assessment. Results revealed that those who were proficient at CT engaged in more metacognitive activities and processes, including self-regulatory planning and evaluation skills. This type of qualitative examination can potentially shed greater light upon the nature of the skills used by students during

CT. Developing Ku and Ho's line of enquiry, this deeper qualitative analysis of the benefits of AM training may also shed light on the relationship between metacognitive processes other than CT, such as self-regulatory planning, and the increase (or decrease) in disposition toward CT and the coupling (or decoupling) of disposition and CT ability over time.

Furthermore, it should be an aim of future research to develop a method of assessing various dispositional, attitudinal, epistemological and motivational factors towards learning and thinking in a manner other than self-report. Perhaps a method of researcher observation can be developed to measure such dispositional traits. Ku (2009) suggests that the HCTA can be used to do this, in that its open-ended nature of questioning requires students to exhibit dispositional traits in order to answer sufficiently. However, Ku's recommendation is largely speculative, given that it is not accompanied by a clear method in which to quantify disposition.

Another important issue to consider in the current study is the nature of feedback provided to students. The discussion from the focus group suggested that though the general feedback provided to all students was helpful in certain respects, they did not always pay close attention to it, due to its impersonal, global nature. Participants suggested that feedback that used the students' names would provide more of a "*personal touch*" (CR5) and might encourage more students to engage the feedback (e.g. "*I'd like to give you an example of John's work, who argued as such and my response to John is as follows. Another example is David who...etc.*" [RA1]).

Interestingly, Butchart et al. (2009) encountered similar problems in providing AM feedback to students. Butchart and colleagues acknowledged that providing feedback on AM exercises is a time consuming process, especially if the goal is to provide personalised feedback. Though personalised feedback can be provided to

students, it becomes less feasible when working with large numbers of students. Notably, after commenting on the problematic nature of their ‘correct’ versus ‘incorrect’ proposition placement feedback provided to students, Butchart and colleagues cite van Gelder’s (2005) method of providing feedback via ‘model answers’ as possibly one of the best ways to provide feedback. Model answers (i.e. model AMs) allow students to compare their AMs with those of the instructor. In the current study, feedback was designed in light of Butchart et al. (2009) and van Gelder’s (2005) recommendations, in that ‘model answer’ AMs were provided to students and compared with a handful of AMs developed by students. In addition, common mistakes made by students were identified, represented and explained – an effort to show students how they might have built a stronger AM. Thus, the feedback provided in the current study improved upon the automated feedback provided by Butchart and colleagues, which simply declared whether an answer was correct or incorrect. The comparison of the adequacy of feedback in the two studies is reflected in the effect size of each intervention – Butchart et al.’s (2009) automated feedback AM condition yielded an effect size of .45, whereas the CT condition in the current study yielded an effect size of .68. Nevertheless, due to the difficulties associated with providing feedback, future research aimed at investigating the effects of *different types* of feedback (e.g. automated feedback, personalised feedback and general class-based feedback), the effects of manipulating the *presentational format* of feedback (e.g. through text, outlines and AMs) and the effects of manipulating the *delivery* of feedback (e.g. in e-learning environments, in didactic classroom settings and in active, collaborative learning settings) is necessary and may also prove informative to the field of instructional design.

6.4.3 Summary & Conclusion

In summary, results from Study 3 revealed that participation in a CT e-learning course taught through AM significantly enhanced overall CT performance and performance on each CT sub-scale (as measured by the HCTA). Results also revealed that those who participated in an AM-infused CT e-learning course performed significantly better than controls on tests of overall CT, verbal reasoning and argument analysis at post-testing. Results from the focus group seem to corroborate these findings, given the focus group members' general satisfaction with the course on the whole, as well as their more specific reports in relation to the benefits of using AM as a method of supporting CT.

RA1: I think it was very well structured from the start - from the first exercise to the last exercise. It brought you through step-by-step so you could really build on each ... It was very easy to understand as well, once you actually followed - once you actually paid attention to the lectures - did what you're supposed to do. It was good. It was very helpful.

CR5: The structure of the way the course was run - you know you started with very simple examples and they got a little bit more complicated and a little bit more complicated - you had to think a little bit harder to work out answers. I think the way it was structured was done very well. The way it started, you know the first two exercises were very simple, you know, and then just levels of complexity were raised all the way along and then the last one was particularly complicated, you know, but you were kind of, for want of a better phrase, spoon fed all the way along and you know you were used to the levels of complexity increasing, I think that was very well done.

AH6: It was very interesting, it kind of made you think more broadly about your own thoughts as well - your opinions, so at the end of each section I probably would slightly change my opinion on some things, so it was good.

In terms of the use of AM to help support CT during both the recorded lectures as well as the exercises, focus group participants reported that:

AH6: They're like an example - so you can build on that and put your argument into that example, whereas if you didn't have it you wouldn't know how to do it... I kind of thought it was all really about the visual aid, the visual system... It helped to identify the credible claims, because they're laid out there, visually, in front of you in the map.

CF4: *It's like you're using more parts of your brain when [the information] is mapped out like that. You're using a different part of your brain to compute just the text. If it was just text it's only using one part, but if it's in the map you're using two different parts. I also think that you're being stretched all the time - you can cope with whatever comes, every day I feel like it's something that I can cope with because I'm looking at different parts and angles. So the visual aid [of the maps] definitely helped for me. Like, I would have thought critically before, but not as critically as I would now, like you do with maps, and you're going through the process more and you're doing more argument maps and its going more into it, you know what I mean?*

In conclusion, consistent with reports which highlight the value of using e-learning to facilitate the development of metacognitive processes (Huffaker & Calvert, 2003), the results of Study 3 suggest that CT skills can be enhanced by participating in an AM-infused CT training course delivered in an e-learning environment. However, future research is necessary to further examine the effects of AM and e-learning on CT, as well as the conditions that most positively affect CT performance and CT dispositions in the context of CT training.

Chapter 7

General Discussion

This chapter presents a general discussion of the research conducted in this thesis, with a specific focus on findings related to the effects of argument mapping on the learning outcomes of memory, comprehension and critical thinking, as well as the dispositional factors that relate to these learning outcomes. This will be followed by a discussion of potential, future research that may be conducted in order to examine the effects of AM in alternative learning contexts, particularly collaborative e-learning environments. Next, broader implications for the use of argument mapping in education will be discussed, followed by a general summary and conclusion.

7.1 Interpretation of Results

The overall aim of the current programme of research was to evaluate the use of argument mapping as a learning tool. Specifically, over the course of three studies, this research examined the effects of argument mapping (AM) on immediate recall, delayed recall, comprehension and critical thinking (CT). Findings from these studies suggest that AM is an efficacious learning method, as it was shown to facilitate both recall and CT ability.

7.1.1 Memory & Comprehension

Study 1 involved four experiments. The collective findings from these experiments suggest that AM reading and construction can facilitate better immediate recall of propositions from arguments when compared with more traditional learning strategies, such as text-reading and text-summarisation. Study 1 experiments revealed that when compared with traditional text-based study materials, AM reading

significantly enhanced the immediate recall of arguments, regardless of (1) the presence or absence of colour to demarcate reasons and objections in AMs (Experiment 1), (2) the environmental setting in which AMs were studied (Experiment 2), and (3) the study topic used in experiments (i.e. positive effects of AM reading on immediate recall were observed for two different topics presented in Experiments 1 - 3). Results also revealed that those who actively learned through AM and hierarchical outline (HO) construction performed significantly better on immediate recall testing than those who actively learned through text-summarisation (i.e. Experiment 4).

Results from Experiment 1 also revealed that though those who studied from colour and monochrome AMs scored significantly higher on immediate recall when compared with a text-reading group; there was no difference between the colour and monochrome AM study groups. This finding suggests that the use of colour within AMs is not as important a contributor to memory performance benefits as other AM features (e.g. the hierarchical organisation of information). This suggestion was further supported by the results from Experiment 4, in which it was found that though both active AM and HO groups scored significantly higher on immediate recall than the text summarisation group, there was no difference between the AM and HO groups. Though the AMs used in Experiment 4 were both coloured and hierarchically organised, the HOs were hierarchically organised only. Therefore, though the processes through which AM and HO construction facilitated learning gains may have been different, the advantage of AM construction over text summarisation was most likely a result of the hierarchical organisation of information - a common feature of both AM and HO learning strategies (e.g. Robinson & Kiewra, 1995; Taylor, 1982; Taylor & Beach, 1984; van Gelder, 2003). Overall, it appears that the hierarchical

organisation of propositions within an AM is what facilitates enhanced memory, rather than the additional gestalt organizational properties of colour that were hypothesised as beneficial in Experiment 1.

Even though this programme of research did not identify colour as an influential variable, future experiments could conduct a more refined analysis of the effects of colour in the use of AMs as learning aids. For example, research suggests that the use of colour encoding to enhance visual short-term memory may be dependent upon the spatial configuration of the coloured target items that are the focus of encoding activity (Jiang, Olson & Chun, 2000). Jiang, Olson and Chun (2000) found that whereas the spatial location of target items is important for determining whether colour will enhance recall, the colour of the item is not necessarily important for determining whether location will enhance recall. Thus, future research might examine the interdependence between colour cues and spatial arrangement features of propositions in AMs, in order to further evaluate the potential benefits of colour use in AMs. Such research is important because, contrary to findings in Study 1, it may be that the use of colour to demarcate reasons from objections in the overall AM structure does support and enhance certain aspects of spatial memory for specific arguments - an effect that could potentially be revealed if a different, spatially-oriented memory testing method were used. For example, future research could examine the effects of AM on recall by presenting three groups of students an argument in the form of an AM, a hierarchical outline or a text and have them study these materials for a fixed amount of time. Following this study period, the three groups could be tested for their recall of the propositions within the argument by providing them with a matching template on a computer screen of the study material originally allocated to them. The testing template would not contain all

the text that was included in the study materials, but rather a large number of blank AM boxes or blank lines. Correct responses would then be identified by matching propositions from the original argument structure with the corresponding location in the AM, HO or text by clicking the box or line where the target proposition should be located based on its location in the original study materials (Ciranni & Shimamura, 1999).

Despite the apparent benefits of AM revealed in Study 1, passive reading of AMs did not appear to facilitate either comprehension or delayed recall. It is possible that the reason for these null effects was because AM reading did not facilitate enhanced levels of schema construction necessary for comprehension or long-term storage over traditional text-reading. For AM reading to enhance both comprehension and delayed recall over and above text reading, it may be that students need more time to assimilate the AMs presented to them. Specifically, it may be that the 10-15 minutes provided to students in Study 1 for AM reading was not enough time for students to conduct schema construction at a level of critical complexity and depth to facilitate full understanding of the material and/or long-term storage. Thus, future research could examine the effects of AM reading, in comparison with HO and text reading, on comprehension and delayed recall in study settings that last longer than those in the current research – possibly 20-30 minutes.

Furthermore, it may be that multiple readings of an AM, as opposed to the one encounter with an AM in Study 1, are necessary for the enhancement of schema construction, and subsequently, delayed recall and comprehension of target information. That is, perhaps students who read an AM two or three times might perform better on tests of delayed recall and comprehension than those who read a text the same amount of times. This suggestion is consistent with past research on

both distributed versus mass learning (Fiore & Salas, 2007; Willingham, 2002) and spacing effects (Challis, 1993; Greene, 1989; 2008), which suggests that multiple encounters with the same study materials enhance delayed recall performance more so than one encounter. In future research on the effects of AM on delayed recall and comprehension, students might, for example, be asked to read an argument via an AM, text or a hierarchical outline for 10 minutes on Monday, again Wednesday and once more on Friday; and then be assessed on their delayed recall and comprehension for that argument the following Monday. In this context, it is possible that the multiple encounters with the study materials might yield significant differences among those who studied from a text, outline and AM.

Due to the time constraints associated with the length of classes in the lecture hall setting of study and testing, it was not possible to administer multiple assessments in all four experiments. As a result, comprehension and delayed recall were not tested in active learning environments as was immediate recall. Consistent with findings from Experiment 4, in which active AM and HO construction facilitated enhanced immediate recall, as well as past research suggesting the beneficial effects of active learning on various learning outcomes (e.g. Hake, 1998; Laws, Sokoloff & Thornton, 1999; Redish, Saul & Steinberg, 1997), it may be that elements of active argument construction is necessary for AM use to develop enhanced levels of delayed recall and comprehension. As a result, future research may investigate the effects of AM construction on delayed recall and comprehension in comparison with the effects of HO construction and text summarisation. For example, students might be provided with a text to read for 10 minutes, followed by a 30-40 minute period of AM construction, HO construction or text summarisation of the text provided. After the active study period, students could be administered a test of comprehension for the

material they actively learned. Students might then be provided a delayed recall test for the same information the following week. It is speculated, given the inclusion of both active learning and the extra time to conduct this active learning, that such future research might reveal beneficial effects of AM on delayed recall and comprehension.

7.1.2 Critical Thinking

Whereas Study 1 examined the effects of AM on memory and comprehension, Studies 2 and 3 investigated the effects of AM on higher-order metacognitive processes, particularly CT. A key feature of both Studies 2 and 3 was that, over the course of six weeks, students who participated in the AM groups were deeply immersed in AM-infused CT training. It was speculated that the training regimes in Studies 2 and 3 might maximise the facilitative effects of AM on CT.

In Study 2, the effects of a six-week AM-infused CT training course were compared with those of a traditional (HO) CT training course and a no-CT training control condition. Study 2 findings revealed that though the CT course taught through AM did not facilitate overall CT or RJ performance over and above the traditional HO CT course or the control condition (i.e. no-CT course), participants in the AM training group performed significantly better on inductive reasoning than those in the control group. However, given that the HO group also scored significantly higher than the control group on inductive reasoning, findings suggested that teaching CT through AM was no better than teaching CT through HO. Interestingly, results from Study 2 are consistent with those from Study 1, where no difference was observed between the AM and HO groups on immediate recall performance. Furthermore, when analysed together, the CT course attendees (i.e. both AM and HO groups combined) outperformed the control condition on analysis, evaluation and inductive reasoning, suggesting that exposure to CT training infused with some hierarchical organisation

strategy may have beneficial effects on CT performance. Overall, these findings again indicate that though the use of colour and spatial organisation of propositions within AMs may provide beneficial visual cues to both the reader and constructor of AMs, the hierarchical organisation of propositions within AMs and HOs may be the critical feature associated with the beneficial effects observed for memory and CT learning outcomes.

Also in Study 2, the null effect of AM on overall CT, analysis, evaluation, inference and RJ was presumed to be a result of one or more of the following limitations (as discussed in Chapter 6): the small sample size of those who participated in the CT intervention; the limited instructor feedback provided to participants during training; the method of assessment used to measure CT and RJ; and possibly the level of students' engagement in the CT courses. Thus, Study 3 was conducted with the primary goal of improving upon the design, implementation, and evaluation of AM-infused CT training.

Study 3 examined the effects of an AM-infused CT e-learning course on CT performance by comparing the effects of an AM-infused CT e-learning course with a no-CT course control group on measures of CT ability. Study 3 also examined the relationship between student engagement in the AM-infused CT course and CT performance changes over time. Results from Study 3 revealed that those who participated in the AM-infused CT group outperformed those in the control group on overall CT, verbal reasoning and argument analysis. Results also revealed that performance on overall CT and all CT sub-scales (i.e. hypothesis testing, argument analysis, verbal reasoning, assessing likelihood and uncertainty, and problem-solving) of those in the AM-infused CT group were significantly enhanced from pre-to-post-

testing. These findings indicated that CT training through the use of AM can facilitate enhanced CT ability.

Interestingly, one sub-scale assessed by the HCTA which was not enhanced more for the AM group than for the control group was problem-solving. This finding is interesting because problem-solving, which “involves the use of multiple problem statements to define the problem and identify possible goals, the generation and selection of alternatives, and the use of explicit criteria to judge among alternatives” (Halpern, 2010, p. 7), is akin to the CT sub-skill of inference, which in Study 2, was also the only core skill of CT not improved by a CT course intervention. These findings suggest that perhaps inferential skills, such as gathering information to draw conclusions or solve problems, like RJ, require longer than six weeks (i.e. the duration of CT courses in both Studies 2 and 3) to significantly develop more than the inferential skills of those who do not participate in CT training. This is not surprising given the complexity of the skill of inference, which is heightened due to its dependency on the ability to also apply skills of analysis and evaluation. Having said that, the one important difference between Study 2 and Study 3 in this context is that in Study 3, the AM group scored significantly higher on the CT post-test than on the CT pre-test in terms of overall CT ability and all CT sub-scales, including problem-solving, suggesting a significant gain in all aspects of CT ability.

Nevertheless, future research might test the above speculation regarding the need for a longer period of AM training for inference skill development. For example, future research might examine the effects of AM- and HO-infused CT training, in comparison with those who do not participate in CT training on skills of inference and problem-solving, as well as RJ (i.e. which was also speculated to require longer periods of development) in a longitudinal study. In this context, AM- and HO-infused

CT training groups could participate in multiple CT courses over the course of two years (e.g. one per semester), in which both CT training groups as well as controls are tested on inference, problem-solving and RJ at testing periods of pre-testing, 6 months, 12 months, 18 months and 24 months. As suggested above, it may be that the development of inference, problem-solving and RJ ability over and above the performance level of controls requires a longer amount of time than that provided in the current research. Thus, such longitudinal research may reveal the speculated, beneficial effects of CT training and more specifically, AM-infused CT training on inferential skills and RJ.

Results from Study 3 also revealed that though there was no strong effect of level of engagement with AM-infused CT exercises on overall CT ability, there was a time x engagement interaction effect for the CT sub-skill of problem-solving, with those in the high-engagement group showing a greater gain in problem-solving ability than those in the low-engagement group. One possible explanation for the overall positive effect of AM-infused CT training, but weaker effect of exercise engagement (i.e. there was neither an effect of condition or time on CT or any of the CT sub-scales), may be the high quality of the lectures posted on the course web-site - available to all students in the AM group, regardless of how many exercises they completed; which may have provided students with sufficient engagement with AM to perform well. However, the positive effect of engagement levels on problem-solving performance in the AM group is promising, as it is broadly consistent with the suggestion above that more intensive levels of engagement may be crucial for the development of higher-level inference skills.

7.1.3 Dispositional Factors towards Thinking

The three studies conducted in this research produced some very interesting findings with regards to AM's effect on memory and CT. However, apart from the current research programme's focus on the effect of AM on lower-and-higher order thinking skills, the research also produced some very interesting findings with regards to dispositional factors associated with higher-order thinking performance outcomes. At the pre-testing stage of Study 2, it was found that overall disposition towards thinking (including disposition towards truth-seeking, open-mindedness, analyticity, systematicity, confidence, inquisitiveness and maturity) was significantly correlated with RJ performance and that disposition towards truth-seeking was significantly, positively correlated with CT performance. However, epistemological beliefs regarding the structure of knowledge was significantly, negatively correlated with both CT and RJ at pre-testing. At the post-testing stage, disposition towards thinking (including truth-seeking and open-mindedness) was again correlated with RJ. Disposition towards thinking (including disposition towards truth-seeking, analyticity, confidence and inquisitiveness) was also significantly, positively correlated with CT at the post-testing stage. In Study 3, it was found that though both need for cognition and motivation were not significantly correlated with overall CT performance at pre-testing, both were significantly correlated with CT performance at post-testing - specifically, need for cognition and the motivation towards critical thinking, metacognitive self-regulation, elaboration, intrinsic goal orientation and control of learning beliefs.

With regards to the overall relationships between CT ability and dispositional factors towards thinking, knowledge and learning, results suggest that disposition towards thinking, need for cognition and motivation towards learning are good

predictors of CT ability. In the case of RJ, this is particularly true, given that it was significantly correlated with overall disposition towards thinking at both pre-and-post-testing. However, these findings require further explanation, given that the correlations between overall CT and the dispositional factors did not exist at pre-testing (with the exception of disposition towards truth-seeking). As indicated in previous chapters, the significant correlation between overall CT and overall disposition, motivation and need for cognition at post-testing suggests that perhaps, CT training is a binding factor between dispositional factors and CT performance. However, it was also speculated that this was not the case for all dispositional factors (e.g. need for cognition and motivation towards both elaboration and organisation), given that there were post-test correlations between CT performance and both need for cognition and motivation in the AM group as well as the control group. Thus, in situations where there was a significant, positive relationship between the CT performance and dispositional factors in controls, it may be that maturation and the experience of college (i.e. which have been shown to enhance CT performance; Pascarella & Terenzini, 1991) might also bind disposition and CT ability over time, over-and-above any effect of CT training. At the same time, it may be that that CT training serves to strengthen the relationship between certain dispositions and CT performance.

In any case, future research is necessary to further examine the relationship between CT performance and the dispositional factors investigated in this programme of research, in order to confirm whether or not the observed pattern of correlations between these two variables holds true in replication studies. In the event that such research is conducted, it seems important to examine the changing nature of the relationship between CT performance and disposition towards truth-seeking,

analyticity, confidence and inquisitiveness, as well as motivation towards control of learning beliefs and intrinsic goal orientation – dispositional factors which were all exclusively correlated with post-test CT performance of those who received CT training. At the same time, consistent with past research (e.g. Colucciello, 1997; Facione, Facione & Sanchez, 1994; Garcia, Pintrich & Paul, 1992; Halpern, 2006; Profeto-McGrath, 2003; Valenzuela, Nieto & Saiz, 2011), in general, the findings from the current research suggest that disposition towards thinking, motivation towards thinking and need for cognition are good predictors of CT ability. These are important findings that are worthy of deeper, qualitative and quantitative analysis in future research.

Overall, results revealed that learning through AM can enhance immediate recall (in both passive and active learning settings), inductive reasoning (as measured by California Critical Thinking Skills Test; Facione, 1990a, Facione et al., 2002), overall CT, hypothesis testing, verbal reasoning, argument analysis, judging likelihood and uncertainty and problem-solving (as measured by the Halpern Critical Thinking Assessment). In addition, results revealed that, regardless of the method used for training CT (e.g. AM and HO), CT training can facilitate analysis, evaluation and inductive reasoning (as measured by California Critical Thinking Skills Test; Facione, 1990a, Facione et al., 2002). Furthermore, results revealed that certain dispositional factors towards thinking, knowledge and learning (e.g. disposition towards thinking, motivation and need for cognition) are good predictors of CT performance.

7.2 Future Research

In the previous three empirical chapters, a number of recommendations were made for future research, in light of both findings and limitations of the current

research programme. These recommendations suggest that further research needs to be conducted in order to further test the potential benefits of AM in relation to a number of learning outcomes (e.g. delayed recall, comprehension and inference). Furthermore, it was recommended that such research on AM be carried out with the goals of (1) controlling for motivation towards learning; (2) controlling for familiarity, interest and perceived difficulty in the topics studied; (3) providing sufficient training in AM to facilitate enhanced learning outcomes; and (4) providing sufficient feedback to students based on their work. Keeping these recommendations in mind, in light of further research and theory, focus now turns to discussion of broader suggestions for future research.

Given that research in this thesis has provided evidence to suggest that AM is a beneficial learning aid, the broad focus of future research on AM should be largely aimed at investigating different educational contexts (1) in which such benefits can be enhanced and (2) where AM can be used to enhance learning outcomes such as delayed recall, comprehension, inference and RJ (i.e. which were not shown to be enhanced by AM in the current research programme). Two educational contexts in which learning through AM may possibly benefit comprehension, delayed recall, CT and RJ are e-learning and collaborative learning. e-Learning may be a particularly fruitful educational context for future research given the beneficial effects of the AM-infused CT e-learning course on CT observed in Study 3, as well as past research on the beneficial effects of e-learning in general (Agarwal & Day, 1998; Brown, 2001; Huffaker & Calvert, 2003; Johnson et al., 2004; Jukes & McCain, 2002; see Chapter 6). Furthermore, given that AM software is used to visually represent the structure of arguments and allows for their manipulation (van Gelder, 2000; 2003); and that argumentation is a social activity (van Eemeren et al, 1996), it seems reasonable to

further speculate that the ability of computer-supported AM to enhance learning may be optimised in collaborative learning settings. Thus, it is recommended that future research on AM focus on its use in computer-supported collaborative learning (CSCL) environments.

Considering that a major focus of the current research was to examine AM's effects on CT, it may be useful for future research to consider the potential benefits of computer-supported AM in conjunction with collaborative learning in CT training environments. It has been argued by Paul (1987; 1993) that dialogue, a fundamental component of collaborative learning, is necessary for good CT. In the context of CT and computer-based AM, dialogue is advantageous because it provides thinkers with an opportunity to explain and question their own beliefs and arguments in light of the thinking and opinions of others involved in the dialogue. In this way, the thinkers involved in the dialogue are actively engaged in collaborative learning. Another benefit of the proposed future research is that over the course of a computer-supported AM-infused CT training intervention, the whole gamut of learning outcomes investigated in the current research programme (i.e. immediate recall, delayed recall, comprehension, analysis, evaluation, inference and reflective judgment) can be examined in the context of a collaborative learning environment. For example, five groups could be compared: (1) a collaborative learning AM-infused CT group, (2) an AM-infused CT group (i.e. without collaborative learning), (3) a traditional HO-infused CT group with collaborative learning (i.e. without AM), (4) a traditional HO-infused CT group (without AM and collaborative learning) and a (5) control group (i.e. without AM, collaborative learning and CT training); with regards to their performances on all the learning outcomes listed above. Specifically, immediate recall and comprehension could be examined during a break in class or at the end of class;

similar to the way it was examined in the current research in which students were provided a study material and asked to study it for an allotted amount of time. Following this study period, students' comprehension and immediate recall could be tested via a specially designed test, relevant to the information studied (i.e. similar to those used in the current research). The examination of immediate recall and comprehension could potentially work well in some of the introductory CT classes, in the context of emphasising the importance of the ability to remember and understand information one is required to think critically about. Furthermore, in the weeks following, delayed recall could be tested by providing the students with the same test completed the week prior. In addition, the higher-order thinking processes of analysis, evaluation, inference and reflective judgment could also be assessed similar to the way they were examined in the current research study, that is, via pre-post intervention-testing.

The rationale for emphasising future research on the effects of AM-infused CSCL environments is that past research indicates that the use of CSCL strategies (e.g. mapping strategies), can facilitate: (1) higher grades on academic course assessments; (2) reasoned discussion among students; and (3) aid in focusing students to transfer these dialogic skills to curriculum-based learning (Engelmann et al., 2010; Engelmann & Hesse, 2010; Hwang, Shi & Chu, 2011; Johnson, Johnson & Stanne, 2000; Ma, 2009; Wegerif and Dawes, 2004). According to Johnson, Johnson and Stanne (2000), there are over 900 research studies that have indicated positive effects of collaborative learning over and above other learning strategies, including individual and competitive learning. Furthermore, in Johnson, Johnson and Stanne's (2000) meta-analysis of 164 high quality studies, a number of cooperative learning strategies were shown to enhance scholastic performance; in particular, *learning*

together, which was shown to enhance academic achievement significantly more than both competitive learning ($d = .85$) and individual learning ($d = 1.04$). Slavin (1995) describes a number of reasons why collaborative learning enhances academic performance, including both motivational and cognitive factors. According to Slavin, when students share a common goal and work together, they positively reinforce each other during their collaborative work, which motivates them to achieve higher standards of performance. Similarly, students are motivated to help one another in collaborative learning environments because it is in their own interest to do so. That is, as students involved in group projects are graded based on what the group achieves, as opposed to their own individual efforts, it is in the interest of each student to motivate and help their peers to work hard and perform well. This is consistent with research by Ma (2009), where it was found that collaborative learning environments in which students are aided by their peers can enhance higher-order thinking skills.

Ma's (2009) research examined how interactivity in a CSCL environment contributed to the development of higher-order thinking skills. Six groups with four-to-five participants in each group engaged in a seven-week project examining developments in textile fibres and fabrics. The degree of intra-group interactivity was measured using the number of direct responses to messages posted in a discussion forum. It was found that the quality of interaction using the discussion forum significantly influenced the quality of learning constructed by the group. For example, Group A had the second highest number of forum messages, with intensive and rich interaction. Specifically, more than 32% of threaded messages contained between four and ten referents. Furthermore, Group A members' understanding was promoted through online group support and help, as those members who lacked understanding of a topic received help from fellow group members to address the lack of knowledge.

Essentially, increased discussion and collaboration led to improvements in the quality of thought and a deeper understanding of the topic.

Future research, similar to that by Ma (2009), could examine how interactivity in CSCL-infused CT training, taught through AM, contributes to the development of higher-order thinking skills, such as CT and RJ. Such research could compare Ma's method of having groups converse via prose/text-based responses and comments with groups who converse via posting AMs on a discussion board, in which responses and comments are made by amending other groups' AMs (e.g. by adding objections, rebuttals and/or supports). In this context, it is hypothesised that, given the support of collaboration and dialogue with peers, the AM groups would perform significantly better than the prose/text-based groups on outcomes of analysis, evaluation, inference and RJ.

In addition to motivational benefits, assisting peers in learning can also benefit the individual who is doing the helping in a cognitive manner (Slavin, 1995). That is, for individuals to build knowledge (i.e. through schema construction), it is beneficial for them to engage in some form of *elaboration*. Elaboration refers to the cognitive restructuring of information that is the target of learning (Wittrock, 1986). Research has shown that one of the most effective methods of elaboration is explaining the information to others (Dansereau, 1988; Newbern, 1994). AM software may be a useful tool in this context because it can be used as a strategy that promotes elaboration of information among students through the presentation and manipulation of information in both a verbal and visual manner.

The recommendation for future research to examine AM-infused CSCL environments is also made because research has shown that reasoning and argumentation skills increase if CSCL is used, given that CSCL aids students in

making their thoughts and solution strategies clear (Kuhn et al., 2008; Wegerif, 2002; Wegerif & Dawes, 2004). For example, Kuhn, et al. (2008) examined the effects of computer-supported instant-messaging on the development of argumentation skills. Twenty-eight sixth grade students participated in an argumentative discourse activity, where they collaborated in pairs and engaged in arguments against successive pairs of peers on the opposite side of an issue. Pairs were involved in discourse on the topics of capital punishment (i.e. for the initial and final topic) and home-schooling. Students were initially required to create their own, individual position on the given topics. Following this, using instant messaging software, they were asked to work in pairs to develop dialogic arguments. Students then engaged in a reflective session, where they were asked to compare the arguments of opposing pairs with their own, prior to actively debating these topics with their competitors. Across the three successive topics, students showed significant gains in their understanding of the nature of good argumentation skills. These advances remained once they were asked to construct individual arguments once more, bereft of the support of their collaborating peers. Though working with peers through computer-supported instant messaging is a method of collaborative e-learning, instant-messaging itself remains a text-based format of presenting/exchanging information. Thus, similar to future research recommended above, additional research could further examine the effects of AM on argumentation skills, by comparing the performance of a number of AM groups (i.e. each consisting of two or three students) with the performance of hierarchical outlining groups and groups using other text-based strategies, on their ability to argue and debate various topics in CSCL environments that utilise web-based communication (e.g. instant-messaging and e-mail).

These recommendations for future research regarding AM are further supported by recent research which suggests that collaborative learning through mapping strategies, similar to AM, can enhance learning performance. For example, research by Engelmann and Hesse (2010) examined the effects of concept mapping through CSCL on domain-specific problem-solving ability. Given that intra-group problem-solving can be hindered by a lack of shared domain-specific knowledge among groups, the authors suggested that the development of digital concept maps may be a possible solution to this problem. Such maps would contain the domain-specific knowledge of all group members regarding a given problem and allow for this information to be shared amongst the group. In Engelmann and Hesse's (2010) research, 20 groups (consisting of three participants each) who constructed such concept maps were compared to 20 other groups who did not construct concept maps. It was found that the concept map groups showed increased awareness of the knowledge of their peers. Moreover, these groups were more focused on problem-relevant information and solved the problems both faster and with more accuracy than the groups without access to the maps.

In a follow-up study, Engelmann et al. (2010) investigated whether access to concept maps alone was adequate for increased group problem-solving or whether communication among the group was also a necessity. Eighty-one participants were involved in the research. In all groups, members were separated from one another. Half of the groups were provided the facility of communicating with intra-group members and viewing their concept maps; whereas the other groups could view the concept maps of their fellow group members, but were not allowed to communicate with them. Results revealed that intra-group communication improved group problem-solving, as it resulted in fewer mistakes in the problem-solving process,

greater performance in complex problem-solving tasks, and less stress amongst group members.

In addition, Hwang, Shi and Chu (2011) examined the benefits of digital concept maps to improve collaborative learning within a context-aware ubiquitous learning environment. *CmapTools* (Novak & Cañas, 2006) concept mapping software enabled students to construct and then share knowledge representations in the form of concept maps. Seventy elementary school students, placed into one of three groups, were involved in the experiment and collaboratively learned about butterfly ecology through reading and observation with their own group only. The experimental group, consisting of 25 participants, were taught how to use concept maps and were then asked to create their own maps via *CmapTools* on a palmtop computer. Control group A, containing 21 participants, were taught how to use concept maps and were then asked to create their own individual 'pen-and-paper' concept maps. Control group B, containing 24 participants, received neither concept map instruction, nor were allowed to use concept maps. It was found that attitudes to science learning and group learning were all significantly higher in the experimental group than the two control groups. Furthermore, when compared to control group A, the collaborative possibilities of the shared digital concept maps led to increased academic performance.

Based on this review of past research on collaborative learning in computer-supported learning environments, it is recommended that future research should examine the effects of using AM in CSCL environments in comparison with (1) other computer-supported mapping strategies (e.g. concept mapping; Engelmann et al., 2010; Engelmann, & Hesse, 2010; Hwang, Shi & Chu, 2011; Roth & Roychoudhury, 1994); (2) computer-supported text-based strategies, such as participation in instant-messaging (e.g. Kuhn et al., 2008), online discussion forums (Ma, 2009) or social

networking websites; as well as (3) computer-supported hierarchical outlining strategies. Overall, it is recommended that future research should examine the effects of computer-based AM as instructional support in collaborative learning settings, because the use of AM, alongside collaborative e-learning strategies may potentially enhance a number of learning outcomes, such as those examined in the current research.

7.3 Broader Implications of Argument Mapping

A more global perspective on the findings from this research suggests that AM can potentially supplement traditional methods of presenting arguments that are the focus of learning in educational settings. For example, based on the findings of the current series of studies, it appears that AM can be successfully used: (1) to support didactic instruction or to potentially replace text-based learning strategies in certain situations; (2) as a study guide provided by the teacher to be used by the student, (3) as a partially completed study guide provided by the teacher to be completed by students when reading text, (4) as a means of supporting text-based study through passive reading or active construction; and/or (5) as a means of providing students with a method of constructing arguments from scratch using specific, class-based material as the basis for AM construction work.

Specifically, in didactic, instructional settings, instead of presenting students with slideshows filled with bullet points of information that they will need to recall in the future, it may prove more advantageous to place AMs within the slides as a means of presenting both the target information and the structure of the reasoning behind it. In this context, AMs may provide students with the opportunity to gain deeper insight and greater understanding of the subject being taught, through assimilating the propositions; drawing the necessary connections among those propositions; and

assessing the relevance, credibility and the logical strength of those propositions as well as their interconnectivity within an AM. Thus, as a result of potentially greater understanding and deeper insight, students may be better able to analyse and evaluate the class materials. However, based on suggestions made in Study 1 (i.e. in light of the null effect of AM on comprehension), AM's ability to potentially facilitate this deeper insight and understanding may be dependent upon the amount of training received in AM strategies. That is, provided that the process of AM is first trained (see Studies 2 and 3), AMs may provide students with a visual scaffold of the information expected to be learned; and may also aid in their ability to critically analyse and evaluate the target information for purposes of creating greater understanding.

Presenting information in this hierarchically organised manner may also allow students to more readily question the importance of propositions and their relationships within class materials, given that the structure of the information is made explicit; and may possibly motivate students to seek further justification from sources apart from class-based materials. That is, if an argument is explicitly laid out for students in class via an AM, it may facilitate their ability to see the logical flow of the argument more easily, given that they are spared the need to simultaneously assimilate the argument and take notes. This may increase the likelihood of a student raising their hand in class and questioning the logic or validity of the argument. That is, AMs in the classroom may promote student engagement in the classroom.

In addition to the benefits of AM in didactic settings, the ability to actively map arguments could potentially aid students to organise their notes outside of the classroom and more easily assimilate important information from additional readings. This in turn would allow them to actively learn information through their own

investigation of the given subject area. Furthermore, findings from Study 3 indicate that the provision of AM software can present students with the opportunity to actively learn, in that students are provided a means of structuring propositions into arguments, gathered from both classroom-based and extracurricular investigations, for purposes of analysing and evaluating the materials and inferring their own conclusions; thus providing them with the opportunity to actively gain a deeper understanding of the subject area.

AM may be utilised in a number of different contexts in educational settings. Whatever domain is the target of argumentation, AM provides a means of structuring one's thoughts so that the reasoning behind argumentation is made clear, thus aiding in problem-solving and the making of critical decisions. Notably, though AM can aid such processes as a result of its visual representation of arguments (which can simplify the process of evaluating reasons and objections and inferring conclusions; van Gelder, 2003), it does not *guarantee* either good reasoning or good decision-making. This is because an AM is only as good as the reasoning behind the argument. In other words, no matter how visually appealing and helpful the representation of the argument, if the argument is poor, the decision-maker will gain very little value from it, other than perhaps recognising how poor their reasoning is. Therefore, while AM can be an efficient tool for presenting arguments, its benefits are entirely dependent on both the critical thinking and reflective judgment ability of its creator. In general, AM may be used as a means of both presenting and constructing arguments. Given the results of the current research, it can be expected that the use of AM will increase users' ability to recall arguments and also facilitate their ability to think critically.

7.4 Summary & Conclusion

In order to acquire and adapt knowledge - the basis for all thinking processes necessary in educational settings, as outlined by Bloom (1956), Romiszowski (1981), Anderson and Krathwohl (2001) and Marzano (2001), students must often navigate many different sources of information, including class notes, text-books and classroom handouts. For students to successfully navigate this information and create new knowledge, they must assimilate, analyse and evaluate information and infer conclusions. More often than not, however, the sources of the information navigated by students present information in a text-based format. Text-based learning materials are problematic because they present information in a linear fashion, in which the inferential relationships amongst items of information are not always made explicitly clear. Without such clarity, a student's ability to assimilate information, regardless of whether lower-or-higher-order thinking skills are used, is compromised by the placement of additional cognitive load on the student reading the text.

As discussed, given that (1) all thinking processes are dependent on one's ability to encode and remember information via the operation of working memory; and that (2) cognitive load places additional demands on working memory processes, it is important to develop new ways of presenting information to students that allows them acquire knowledge with as little cognitive load as possible. In the current research, argument mapping, a method of organising information hierarchically in a visual representation, was speculated to keep constraining factors like cognitive load to a minimum (van Gelder, 2000; 2001; 2003).

Building upon Bloom's taxonomy, and drawing upon the work of Romiszowski (1981), Anderson and Krathwohl (2001), Marzano (2001), Baddeley (2000; 2002), Chi, Glaser and Rees (1982), Kotovsky, Hayes and Simon (1985),

Sweller (1999; 2005; 2010), Facione (1990), Halpern (2003a; 2006) and King and Kitchener (1994), the current research programme examined the effects of argument mapping on thinking processes identified as necessary in educational settings, including memory, comprehension, analysis, evaluation, inference and reflective judgment. Results from Study 1 revealed that AM reading and construction facilitated better immediate recall performance when compared with more traditional methods of text-based learning, such as text-reading and text summarisation. Study 2 found that participation in a CT training course significantly enhanced CT skills of analysis, evaluation, and inductive reasoning. Participants in the AM-infused course scored significantly higher than controls on tests of inductive reasoning, while participants in the traditional (hierarchical outlining) CT course scored significantly higher than controls on tests of analysis and inductive reasoning. Finally, results from Study 3 revealed that there was a significant gain from pre-to-post-testing on all aspects of CT performance for those who attended the AM-infused CT e-learning course; and that those who took part in the AM-infused course scored significantly higher on overall CT and several CT sub-scales, including verbal reasoning and argument analysis, than those in the control group.

Overall, the current research suggests that argument mapping can enhance many of the thinking processes identified in this thesis, above and beyond some more traditional, text-based methods. However, future research is necessary to further evaluate argument mapping's potential benefits, for example, by examining argument mapping's effects on learning outcomes in computer-supported collaborative learning environments. At the same time, the research conducted in this thesis recommends argument mapping as an efficacious educational tool that is worthy of future application and investigation.

References

- Abrami, P.C., Bernard, R.M., Borokhovski, E., Wade, A., Surkes, M.A., Tamim, R., & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research, 78*, 4, 1102-1134.
- Adams, M.H., Whitlow, J.F., Stover, L.M., & Johnson, K.W. (1996). Critical thinking as an educational outcome: An evaluation of current tools of measurement. *Nurse Educator, 21*, 23–32.
- Agarwal, R., & Day, A. E. (1998). The impact of the Internet on economic education. *Journal of Economic Education, 29*, 99-110.
- Allegretti, C.L., & Frederick, J.N. (1995). A model for thinking critically about ethical issues. *Teaching of Psychology, 22*, 1, 46-48.
- Allen, R., Feezel, J., & Kauffeld, F. (1967). *A taxonomy of concepts and critical abilities related to the evaluation of verbal arguments (Occasional Paper No. 9)*. Madison, WI: Wisconsin Research and Development, Center for Cognitive Learning.
- Alvarez-Ortiz, C. (2007). *Does Philosophy Improve Critical Thinking Skills?* Unpublished master's thesis, University of Melbourne, Australia.
- Anastasi, A. (1988). Explorations in human intelligence: Some uncharted routes. *Applied Measurement in Education, 1*, 3, 207-213.
- Anderson, J. R. 1990. *The adaptive character of thought*. Hillsdale, N.J.: Erlbaum.
- Anderson, L. W., & Krathwohl, D.R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Addison-Wesley.

- Anderson, R.D., (1983). A consolidation and appraisal of science meta-analyses. *Journal of Research in Science Teaching*, 20, 497-509.
- Armbruster, B. B., Anderson, T. H., & Ostertag, J. (1987). Does text structure/ summarization instruction facilitate learning from expository text? *Reading Research Quarterly*, 22, 3, 331-346.
- Association of American Colleges & Universities, (2005). *Liberal education outcomes: A preliminary report on student achievement in college*. Washington, DC: AAC&U.
- Athanassiou, N., McNett, J.M., & Harvey, C. (2003). Critical thinking in the management classroom: Bloom's taxonomy as a learning tool. *Journal of Management Education*, 27, 5, 533-555.
- Atkinson, R.C., & Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In K.W. Spence & J.T. Spence (Eds.), *The psychology of learning and motivation*, 2, 89-195. New York: Academic Press.
- Australian Council for Educational Research, (2002). *Graduate skills assessment: Stage one validity study*. Australia: Dept. of Education, Science and Training.
- Ayres, P., & Sweller, J. (2005). The split-attention principle in multimedia learning. In R.E. Mayer (Ed.) *The Cambridge handbook of multimedia learning*, 135-146. New York: Cambridge University Press.
- Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.
- Baddeley, A.D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4, 417-423.
- Baddeley, A.D. (2002). Is working memory still working? *European Psychologist*, 7, 85-97.

- Baddeley, A.D., Eldridge, M., & Lewis, V. (1981). The role of subvocalisation in reading. *The Quarterly Journal of Experimental Psychology Section A*, 33, 4, 439-454.
- Baddeley, A.D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory*, 8, 47-90. New York: Academic Press.
- Baddeley, A.D., & Wilson, B.A. (2002). Prose recall and amnesia: Implications for the structure of working memory. *Neuropsychologia*, 43, 4, 583-587.
- Baril, P. B., Cunningham, B. M., Fordham, D. R., Gardner, R. L., & Wolcott, S. K. (1998). Critical thinking in the public accounting profession: Aptitudes and attitudes. *Journal of Accounting Education*, 16, 4, 381-406.
- Baron, R. S. (1986). Distraction-conflict theory: Progress and problems. *Advances in Experimental Social Psychology*, 19, 1-40.
- Bartlett, F. (1932). *Remembering: A study in experimental and social psychology*. New York: Cambridge University Press.
- Beckie, T.M., Lowry, L.W., & Barnett, S. (2001). Assessing critical thinking in baccalaureate nursing students: A longitudinal study. *Holistic Nursing Practice*, 15, 3, 18-26.
- Bennett, G.K., Seashore, H.G., & Wesman, A.G. (1986). *Differential aptitude test*. Windsor, UK: Psychological Corporation & NFER-Nelson.
- Bensley, D.A. (1998). *Critical thinking in psychology: A unified skills approach*. Pacific Grove, CA: Brooks & Cole.
- Berkowitz, S.J. (1986). Effects of instruction in text organization on sixth-grade students' memory for expository reading. *Reading Research Quarterly*, 21, 2, 161-178.

- Bloom, B.S., (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain*. New York: McKay.
- Boekaerts, M., & Simons, P.R.J. (1993). *Learning and instruction: Psychology of the pupil and the learning process*. Assen: Dekker & van de Vegt.
- Bond Jr., C.F., & Titus, L.J. (1983). Social facilitation: A meta-analysis of 241 studies. *Psychological Bulletin*, 94, 2, 265-292.
- Bower G.H., Clark, M.C., Lesgold, A.M., & D. Winzenz. (1969). Hierarchical retrieval schemes in recall of categorised word lists. *Journal of Verbal Learning and Verbal Behavior*, 8, 23-343.
- Brabeck, M.M. (1981). The relationship between critical thinking skills and development of reflective judgment among adolescent and adult women. Paper presented at the 89th Annual Convention of the American Psychological Association, Los Angeles, August 24-26.
- Broadbent, D.E. (1958). *Perception and communication*. New York: Oxford University Press.
- Brown, A. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. Reiner & R. Kluwe (Eds.), *Metacognition, motivation, and understanding*, 65-116. Hillsdale, NJ: Erlbaum.
- Brown, K.G. (2001). Using computers to deliver training: Why employees learn and why? *Personnel Psychology*, 54, 271-296.
- Buckingham-Shum, S.J. (2003). The roots of computer supported argument visualization. In P. A. Kirschner, S. Buckingham-Shum, & C. Carr (Eds.), *Visualizing argumentation: Software tools for collaborative and educational sense-making*, 3-24. London: Springer-Verlag.

- Budd, J. (2004). Mind maps as classroom exercises. *Journal of Economic Education*, 35, 1, 35-49.
- Burbach, M., Matkin, G., & Fritz S. (2004). Teaching critical thinking in an introductory leadership course utilizing active learning strategies: A confirmatory study. *College Student Journal*, 38, 3, 482-493.
- Butchart, S., Bigelow, J., Oppy, G., Korb, K., & Gold, I. (2009). Improving critical thinking using web-based argument mapping exercises with automated feedback. *Australasian Journal of Educational Technology*, 25, 2, 268-291.
- Buzan, T., & Buzan, B. (1997). *The mind map book*. London: BBC Books.
- Cacioppo, J.T., Petty, R.E., & Kao, C.F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, 48, 306-307.
- Cano, F. (2005). Epistemological beliefs and approaches to learning: Their change through secondary school and their influence on academic performance. *British Journal of Educational Psychology*, 75, 203-221.
- Carini, R.M., Kuh, G.D., & Klein, S.P. (2006). Student engagement and student learning: Testing and linkages. *Research in Higher Education*, 47, 1, 1-32.
- Carr, C.S. (2003). Using computer supported argument visualization to teach legal argumentation. In P. A. Kirschner, S. Buckingham-Shum, & C. Carr (Eds.), *Visualizing Argumentation: Software tools for collaborative and educational sense-Making*, 75–96. London: Springer-Verlag.
- Carroll, J.B. (1993). *Human cognitive abilities. A survey of factor-analytic studies*. Cambridge: Cambridge University Press.
- Chaffee, J. (1988). *Thinking critically (2nd ed.)*. Boston: Houghton Mifflin.

- Challis, B.H. (1993). Spacing effects on cued-memory tests depend on level of processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 2, 389-396.
- Chan, J.C.K. (2009). When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language*, 61, 153-170.
- Chandler, P., & J. Sweller, J. (1991). Evidence for cognitive load theory. *Cognition and Instruction*, 8, 4, 351-362.
- Chase, W G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4, 5-81.
- Chen, C.M. (2009). Personalized e-learning system with self-regulated learning assisted mechanisms for promoting learning performance. *Expert Systems with Applications*, 36, 5, 8816-8829.
- Chi, M.T.H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R.S. Sternberg (Ed.), *Advances in the psychology of human intelligence*, 7-77. Hillsdale, NJ: Erlbaum.
- Ciranni, M.A., & Shimamura, A.P. (1999). Retrieval-induced forgetting in episodic memory. *Journal of Experimental Psychology*, 25, 6, 1403-1414.
- Clark, R.C. (2005). Multimedia learning in e-courses. In R.E. Mayer (ed.) *The Cambridge handbook of multimedia learning*, 589-615. New York: Cambridge University Press.
- Cohen, J. (1988). *Statistical power for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

- Colette, E. (2001). Electronic education system model. *Computers & Education*, 36, 171-182.
- Colom, R., Abad, F.J., Rebollo, I., & Shih, P.C. (2005). Memory span and general intelligence: A latent-variable approach. *Intelligence*, 33, 623-642.
- Colucciello, M.L. (1997). Critical thinking skills and dispositions of baccalaureate nursing students - A conceptual model for evaluation. *Journal of Professional Nursing*, 13, 236-245.
- Cottrell, S. (2005). *Critical thinking skills*. New York: Palgrave Macmillan Ltd.
- Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral & Brain Sciences*, 24, 87-185.
- Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research*, 169, 323-338.
- Craik, F. I. M. (1983). On the transfer of information from temporary to permanent memory. *Philosophical Transactions of the Royal Society of London*, 302, 341-359.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour*, 11, 671-684.
- Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104, 3, 268-294.
- Craik, F. I. M., & Watkins, M.J. (1973). The role of rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behaviour*, 12, 6, 599-607.

- Daly, W.M. (2001). The development of an alternative method in the assessment of critical thinking as an outcome of nursing education. *Journal of advanced nursing*, 36, 1, 120-130.
- Dansereau, D.F. (1988). Cooperative learning strategies. In C.E. Weinstein, E.T. Goetz, & P.A. Alexander (Eds.), *Learning and study strategies: Issues in assessment, instruction and evaluation*, 103-120). Orlando, FL: Academic Press.
- Darling-Hammond, L. (2008). How can we teach for meaningful learning? In L. Darling-Hammond (Ed.), *Powerful Learning*, 1-10.
- Davies, M. (2010). Concept mapping, mind mapping, and argument mapping: What are the differences and do they matter? *Higher Education*, 62, 3, 279-301.
- Davis, S.G., & Gray, E.S. (2007). Going beyond test-taking strategies: Building self-regulated students and teachers. *Journal of Curriculum and Instruction*, 1, 1, 31-47.
- Dawson, T.L. (2008a). *Metacognition and learning in adulthood*. Northhampton, MA: Developmental Testing Service, LLC.
- Dawson, T.L. (2008b). *The Lectical Assessment System: Validity and reliability*. Northhampton, MA: Developmental Testing Service, LLC.
- Dawson-Tunik, T.L., Commons, M.L., Wilson, M., and Fischer, K.W. (2005). The shape of development. *International Journal of Cognitive Development*, 2, 163-196.
- De Bono, E. (1985). *Six thinking hats*. New York: Key Porter Books.
- Deci, E. L., Spiegel, N. H., Ryan, R. M., Koestner, R., & Kaufman, M. (1982). The effects of performance standards on teaching styles: The behavior of controlling teachers. *Journal of Educational Psychology*, 74, 852-859.

- Dee-Lucas, D., & Larkin, J.H. (1995). Learning from electronic texts: Effects of interactive overviews for information access. *Cognition and Instruction, 13*, 3, 431-468.
- Demetriou, A. (1993). On the quest of the functional architecture of developing mind. *Educational Psychology Review, 5*, 1-18.
- Demetriou, A. (2000). Organisation and development of self-understanding and self-regulation: Toward a general theory. In M. Boekaerts, P.R. Pintrich, and M. Zeidner (Eds.), *Handbook of Self-Regulation*, 209-251. London: Academic Press.
- Dewey, J., (1910). *How to think*. Boston: Heath & Co.
- Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process*. Lexington, MA: Heath & Co.
- Donohue, A., van Gelder, T., & Cumming, G. (2002). *Reason! Project Studies, 1999-2002*. Melbourne: The University of Melbourne.
- Dressel, P.L., & Mayhew, L.B. (1954). *General education: Exploration in education*. Washington, DC: American Council on Education.
- Duchastel, P.C. (1990). Examining cognitive processing in hypermedia usage. *Hypermedia, 2*, 221-233.
- Dwyer, C.P., Hogan, M.J., & Stewart, I. (2010). The evaluation of argument mapping as a learning tool: Comparing the effects of map reading versus text reading on comprehension and recall of arguments. *Thinking Skills and Creativity, 5*, 1, 16-22.
- Dwyer, C.P., Hogan, M.J., & Stewart, I. (2011a). The promotion of critical thinking skills through argument mapping. In C.P. Horvart & J.M. Forte (Eds.), *Critical Thinking*. New York: Nova Science Publishers.

- Dwyer, C.P., Hogan, M.J., & Stewart, I. (2011b). An examination of the effects of argument mapping on students' memory and comprehension performance. *Thinking Skills & Creativity*. Submitted.
- Dwyer, C.P., Hogan, M.J., & Stewart, I. (2011c). An evaluation of argument mapping as a method of enhancing critical thinking performance in e-learning environments. *Metacognition and Learning*. Submitted.
- Engelmann, T., Baumeister, A., Dingel, A., & Hesse, F.W. (2010). The added value of communication in a CSCL-scenario compared to just having access to the partners' knowledge and information. In J. Sánchez, A. Cañas, & J.D. Novak (Eds.), *Concept maps making learning meaningful: Proceedings of the 4th international conference on concept mapping, 1*, 377–384. Viña del Mar, Chile: University of Chile.
- Engelmann, T. & Hesse, F.W. (2010). How digital concept maps about the collaborators' knowledge and information influence computer-supported collaborative problem solving. *Computer-Supported Collaborative Learning*, 5, 299–319.
- Ennis, R.H. (1987). A taxonomy of critical thinking dispositions and abilities. In J.B. Baron, and R.J. Sternberg (Eds.), *Teaching Thinking Skills: Theory and Practice*, 9-26. New York: W.H. Freeman.
- Ennis, R.H. (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher*, 18, 4-10.
- Ennis, R. H. (1996). *Critical thinking*. Upper Saddle River, New Jersey: Prentice-Hall.
- Ennis, R.H. (1998). Is critical thinking culturally biased? *Teaching Philosophy*, 21, 1, 15-33.

- Ennis, R.H., Millman, J., & Tomko, T.N. (1985). *Cornell critical thinking tests*.
CA: Critical Thinking Co.
- Ennis, R.H., & Weir, E. (1985). *The Ennis-Weir critical thinking essay test*. Pacific
Grove, CA: Midwest Publications.
- Ericsson, K.A., & Kintsch, W. (1995). Long-term working memory. *Psychological
Review*, 102, 211-245.
- Facione, P. A. (1990a). *The California critical thinking skills test (CCTST): Forms A
and B; The CCTST test manual*. Millbrae, CA: California Academic Press.
- Facione, P.A. (1990b). The Delphi report: Committee on pre-college philosophy.
Millbrae, CA: California Academic Press.
- Facione, P.A. (1991). Testing college level critical thinking skills. Paper presented at
the *11th Annual Conference on Critical Thinking and Educational Reform*,
Sonoma, CA, August.
- Facione, P. A. (2000). The disposition toward critical thinking: Its character,
measurement, and relationship to critical thinking, *Informal Logic*, 20, 1, 61 –
84.
- Facione, P.A., & Facione, N.C. (1992). *The California critical thinking dispositions
inventory (CCTDI) and CCTDI Test Manual*. California Academic Press,
California.
- Facione, P.A., Facione, N. C., Blohm, S. W., & Giancarlo, C. A. (2002). *The
California critical thinking skills test: CCTST. Form A, form B, and form
2000. Test manual, 2002 updated edition*. Millbrae, CA: Insight Assessment.
- Facione, P.A., Facione, N.C., & Giancarlo, C.A. (1997). *Setting expectations for
student learning: New directions for higher education*. Millbrae: California
Academic Press.

- Facione N.C., Facione, P.A., & Sanchez, C.A. (1994). Critical thinking disposition as a measure of competent clinical judgment: the development of the California Critical Thinking Disposition Inventory. *Journal of Nursing Education, 33*, 8, 345-350.
- Farrand, P., Hussain, F., & Hennessy, E. (2002). The efficacy of the 'mind map' study technique. *Medical Education, 36*, 426-431.
- Felder, R.M., & Silverman, L.K. (1988). Learning and teaching styles in engineering education. *Engineering Education, 78*, 7, 674-681.
- Fiore, S. M., & Salas, E. (2007). *Toward a science of distributed learning*. Washington, DC: American Psychological Association.
- Fischer, K.W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review, 87*, 477-431.
- Fischer, K. W., & Bidell, T. R. (2006). Dynamic development of action, thought, and emotion. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Theoretical models of human development (6th ed.)*, 1, 313-399. New York: Wiley.
- Fisher, A. (1988). *The logic of real arguments*. Cambridge: Cambridge University Press.
- Flavell, J. (1976). Metacognitive aspects of problem solving. In L. Resnick (Ed.), *The nature of intelligence*, 231-236. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of psychological inquiry. *American Psychologist, 34*, 906-911.

- Folsom-Kovarik, J.T., Schatz, S., Sukthankar, G., & Nicholson, D. (2010). What information does this question convey?: Leveraging help-seeking behavior for improved modeling in a simulation-based intelligent tutor. *Proceedings of the 2010 Spring Simulation Multiconference*, Orlando, FL, April 11-15.
- Fox, T., Grunst, G., & Quast, K.J. (1993). *HyPlan: A context-sensitive hypermedia help system*. Report No. 743, GMD, St. Augustin, Germany.
- Fuster, J.M. (2000). Prefrontal neurons in networks of executive memory. *Brain Research Bulletin*, 52, 331–336.
- Gadzella, B.M., Ginther, D.W., & Bryant, G.W. (1996). Teaching and learning critical thinking skills. Paper presented at the 26th *International Congress of Psychology*, Montreal, August 19.
- Gagne, R.M. (1985). *Conditions of learning (2nd ed.)*. New York: Holt, Rinehart and Winston.
- Garcia, T., Pintrich, P.R., & Paul, R. (1992). Critical thinking and its relationship to motivation, learning strategies and classroom experience. Paper presented at the 100th *Annual Meeting of the American Psychological Association*, Washington, DC, August 14-18.
- Gardner, H. (1985). The development of symbolic literacy. In Wrolstad, M.E. & Fisher, D.F. (Eds.), *Toward a New Understanding of Literacy*, 39-56. New York: Praeger.
- Garrison, D. R., & Anderson, T. (2003). *E-learning in the 21st century: A framework for research and practice*. London: Routledge/Falmer.
- Gerlach, V., & Sullivan, A. (1967). *Constructing statements of outcomes*. Inglewood, CA: Southwest Regional Laboratory for Educational Research and Development.

- Glaser, E. M. (1941). *An experiment in the development of critical thinking*. New York: Teachers College of Columbia University, Bureau of Publications.
- Gobet, F., & Clarkson, G. (2004). Chunks in expert memory: Evidence for the magical number four... or is it two? *Memory*, *12*, 732–747.
- Godden, D.R., & Baddeley, A.D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, *66*, 325-331.
- Good, T.L., & Brophy, J.E. (1986). *Educational Psychology (3rd ed.)*. New York: Longman.
- Govindasamy, T. (2002). Successful implementation of e-learning: Pedagogical considerations. *Internet and Higher Education*, *4*, 287-299.
- Granello, D.H. (2001). Promoting cognitive complexity in graduate written work: Using Bloom's taxonomy as a pedagogical tool to improve literature reviews. *Counselor Education and Supervision*, *40*, 4, 292-307.
- Greenagel, F.L. (2002). The illusion of e-learning: Why we are missing out on the promise of technology. Retrieved 10/12/2011 from <http://www.league.org/publication/whitepapers/0802.html>
- Greene, R.L. (1989). Spacing effects in memory: Evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 3, 371-377.
- Greene R. L. (2008). Repetition and spacing effects. In Roediger H. L. III (Ed.), *Learning and memory: A comprehensive reference, 2: Cognitive psychology of memory*, 65-78. Oxford: Elsevier.

- Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Taboada, A., Davis, M. (2004). Increasing reading comprehension and engagement through concept-oriented reading instruction. *Journal of Educational Psychology*, 96, 403-423.
- Hake, R. (1998). Interactive-engagement vs. traditional methods: A six-thousand student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66, 1, 64–74.
- Halpern, D.F. (2003a). *Thought & knowledge: An introduction to critical thinking* (4th Ed.). New Jersey: Laurence Erlbaum Associates.
- Halpern, D. F. (2003b). The “how” and “why” of critical thinking assessment. In D. Fasko (Ed.), *Critical thinking and reasoning: Current research, theory and practice*. Cresskill, NJ: Hampton Press.
- Halpern, D.F., (2006). Is intelligence critical thinking? Why we need a new definition of intelligence. In Kyllonen, P.C., Roberts, R.D. & Stankov, L. (Eds.), *Extending intelligence: enhancement and new constructs*, 293-310. New York: Taylor & Francis Group.
- Halpern, D.F. (2010). *The Halpern critical thinking assessment: Manual*. Vienna: Schuhfried.
- Hancock, D.R. (2002). Influencing graduate students' classroom achievement, homework habits and motivation to learn with verbal praise. *Educational Research*, 44, 1, 83-95.
- Harrell, M. (2004). The improvement of critical thinking skills. Paper 126, Tech. Rep. CMU-PHIL-158. Carnegie Mellon University, Department of Philosophy. <http://repository.cmu.edu/philosophy/126>.

- Harrell, M. (2005). Using argument diagramming software in the classroom. *Teaching Philosophy*, 28, 2, 163-177.
- Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student Learning: A meta-analysis. *Review of Educational Research*, 66, 2, 99-136.
- Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Sciences*, 8, 6, 280-285.
- Higher Education Quality Council, (1996). *What are graduates? Clarifying the attributes of "graduateness"*. London: HEQC.
- Hitchcock, D. (2004). The effectiveness of computer-assisted instruction in critical thinking. *Informal Logic*, 24, 3, 183-218.
- Hofer, B.K. (2004). Epistemological understanding as a metacognitive process: Thinking aloud during online searching. *Educational Psychologist*, 39, 1, 43-55.
- Hogan, M.J., (2006). Against didacticism: a psychologist's view. *Educational Research and Reviews*, 1, 7, 206-212.
- Hogan, M.J., & Stein, Z. (2010). Structuring thought: An examination of four methods. In D.A. Contreras (Ed.), *Psychology of Thinking*, 65-96. New York: Nova.
- Holmes, J., & Clizbe, E. (1997). Facing the 21st century. *Business Education Forum*, 52, 1, 33-35.
- Horn, R. E. (1999). *Can Computers Think?* Bainbrigde, WA: MacroVU, Inc.
- Horn, R. E. (2003). Infrastructure for navigating interdisciplinary debates: Critical decisions for representing argumentation. In Kirschner, P., Buckingham-Shum, S., Carr, C. (Eds.), *Visualizing Argumentation: Software Tools for*

- Collaborative and Educational Sense-Making*, 165-184, London: Springer-Verlag.
- Horn, R. E., & Weber, R. P. (2007). *New tools for resolving wicked problems: Mess mapping and resolution mapping processes*. Bainbrigde Island, WA: MacroVU, Inc. and Strategy Kinetics, LLC.
- Huffaker, D.A., & Calvert, S.L. (2003). The new science of learning: Active learning, metacognition and transfer of knowledge in e-learning applications. *Journal of Educational Computing Research*, 29, 3, 325-334.
- Huffman, K., Vernoy, M., Williams, B., & Vernoy, J. (1991). *Psychology in action*. New York: John Wiley and Sons.
- Hugenholtz, N.I.R., de Croon, E.M., Smits, P.B., van Dijk, F.J.H. & Nieuwenhuijsen, K. (2008). Effectiveness of e-learning in continuing medical education for occupational physicians. *Occupational Medicine*, 58, 370–372.
- Huitt, W. (2011). Bloom et al.'s taxonomy of the cognitive domain. *Educational Psychology Interactive*. Valdosta, GA: Valdosta State University. Retrieved 08/08/2011, from <http://www.edpsycinteractive.org/topics/cognition/bloom.html>
- Hwang, G.J., Shi, Y.R., & Chu, H.C. (2011). A concept map approach to developing collaborative mindtools for context-aware ubiquitous learning. *British Journal of Educational Technology*, 42, 5, 778-789.
- Jacobs, S. S. (1995). Technical characteristics and some correlates of the California Critical Thinking Skills Test Forms A and B. *Higher Education Research*, 36, 89-108.
- Jensen, L.L. (1998). *The role of need for cognition in the development of reflective judgment*. Unpublished PhD thesis, University of Denver, Colorado, USA.

- Jiang, Y., Olson, I.R., & Chun, M.M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, 26, 683–702.
- Johnson, C.E., Hurtubise, L.C., Castrop, J., French, G., Groner, J., Ladinsky, M., McLaughlin, D., Plachta L., & Mahan, J.D. (2004). Learning management systems: Technology to measure the medical knowledge competency of the ACGME. *Medical Education*, 38, 599-608.
- Johnson, D.W., Johnson, R.T., & Stanne, M.S. (2000). Cooperative learning methods: A meta-analysis. Retrieved 21/06/2011, from <http://www.cooperation.org/pages/cl-methods.html>.
- Jonassen, D., & Tessmer, M. (1996). An outcomes based taxonomy for instructional systems design, evaluation and research. *Training Research Journal*, 2, 11-46.
- Jukes, I., & McCain, T. (2002). *Minds in play: Computer game design as a context of children's learning*. New Jersey: Erlbaum.
- King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults*. San Francisco: Jossey Bass.
- King, P.M., & Kitchener, K.S., (2002). The reflective judgment model: Twenty years of epistemic cognition. In B.K. Hofer and P.R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing*, 37-61. New Jersey, Lawrence Erlbaum Associates.
- King, P. M., & Kitchener, K. S. (2004). Reflective judgment: Theory and research on the development of epistemic assumptions through adulthood. *Educational Psychologist*, 39, 1, 5-15.

- King, P.M., Wood, P.K., & Mines, R.A. (1990). Critical thinking among college and graduate students. *The Review of Higher Education, 13*, 2, 167-186.
- Kintsch, W., & van Dijk, T.A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*, 363-394.
- Kitchener, K.S., & King, P.M. (1981). Reflective judgment: Concepts of justifications and their relation to age and gender. *Journal of Applied Developmental Psychology, 2*, 2, 89-116.
- Klem, A.M., & Connell, J.P. (2004). Relationships matter: Linking teacher support to student engagement and achievement. *Journal of School Health, 74*, 7, 262-273.
- Kosslyn, S.M. (1980). *Image and mind*. Cambridge, Massachusetts: Harvard University Press.
- Kotovsky, H.A., Hayes, K., & Simon, J.R. (1985). Why are some problems hard? Evidence from the Tower of Hanoi. *Cognitive Psychology, 17*, 22, 248-29.
- Krathwohl, D.R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice, 41*, 4, 212-218.
- Kreitzer, A. E., & Madaus, G.F. (1994). Empirical investigations of the hierarchical structure of the taxonomy. In L. W. Anderson & L. A. Sosniak (Eds.) , *Bloom's Taxonomy: A Forty-Year Retrospective, Ninety-third Yearbook of the National Society for the Study of Education*, 64-81. Chicago: University of Chicago Press.
- Kropp, R.P., Stoker, H.W., & Bashaw, W.L. (1966). The validation of the taxonomy of educational objectives. *Journal of Experimental Education, 34*, 3, 69-76.

- Ku, K.Y.L. (2009). Assessing students' critical thinking performance: Urging for measurements using multi-response format. *Thinking Skills and Creativity*, 4, 1, 70- 76.
- Ku, K. Y. L., & Ho, I.T. (2010a). Dispositional factors predicting Chinese students' critical thinking performance. *Personality and Individual Differences*, 48, 54-58.
- Ku, K.Y.L. & Ho, I.T. (2010b). Metacognitive strategies that enhance critical thinking. *Metacognition & Learning*, 5, 251-267.
- Kuhn, D. (1991). *The skills of argument*. Cambridge: Cambridge University Press.
- Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher*, 28, 2, 16-25.
- Kuhn, D., Goh, W., Iordanou, K., & Shaenfield, D. (2008). Arguing on the computer: A microgenetic study of developing argument skills in a computer-supported environment. *Child Development*, 79, 5, 1310-1328.
- Kurfiss, J.G. (1988). *Critical thinking: Theory, research, practice and possibilities. Higher Education Report No.2*. Washington DC: Association for Study for Higher Education.
- Larkin, J.H., McDermott, J., Simon, D.P., & Simon, H.A. (1980). Models of competence in solving physics problems. *Cognitive Science*, 4, 317-345.
- Larkin, J., & Simon, H. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11, 65–99.
- Laws, P., Sokoloff, D., & Thornton, R. (1999). Promoting active learning using the results of physics education research. *UniServe Science News*, 13, 14-19.
- LeBlanc, J. (1998). *Thinking clearly: A guide to critical reasoning*. New York: W.W. Norton.

- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.
- Ma, A.W.W. (2009). Computer supported collaborative learning and higher order thinking skills: A case study of textile studies. *The Interdisciplinary Journal of e-Learning and Learning Objects*, *5*, 145-167.
- MacPherson, K. (1999). The development of critical thinking skills in undergraduate supervisory management unites: Efficacy of student peer assessment. *Assessment & Evaluation in Higher Education*, *24*, 3, 273-284.
- Marzano, R.J. (1998). *A theory-based meta-analysis of research on instruction*. Aurora, CO: Mid-Continent Regional Educational Laboratory. Retrieved 26/10/2007, from http://www.mcrel.org/pdf/instruction/5982rr_instructionmeta_analysis.pdf.
- Marzano, R.J. (2001). *Designing a new taxonomy of educational objectives*. Thousand Oaks, CA: Corwin Press.
- Mason, L. H., Snyder, K. H., Sukhram, D. P., & Kedem, Y. (2006). TWA + PLANS strategies for expository reading and writing: Effects for nine fourth-grade students. *Exceptional Children*, *73*, 69-89.
- Maybery, M.T., Bain, J.D., & Halford, G.S. (1986). Information-processing demands of transitive inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 4, 600-613.
- Mayer, R. E. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology*, *81*, 2, 240-246.
- Mayer, R.E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, *32*, 1, 1-19.

- Mayer, R.E. (2003). The promise of multimedia learning: using the same instructional design methods across different media. *Learning and Instruction, 13*, 125-139.
- Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist, 59*, 1, 14-19.
- Mayer, R.E. (2005). Cognitive theory of multimedia learning. In R.E. Mayer (Ed.), *The Cambridge handbook of multimedia learning*, 31-48. New York: Cambridge University Press.
- Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology, 83*, 484-490.
- Mayer, R. E., & Anderson, R. B. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology, 84*, 444-452.
- Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology, 82*, 715-726.
- Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*, 289-401.
- Mayer, R.E., Steinhoff, K., Bower, G., & Mars, R. (1995). A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. *Educational Technology Research and Development, 43*, 1, 31-44.

- Means, M. L., and Voss, J. E. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability and knowledge levels. *Cognition & Instruction, 4, 2*, 139-178.
- Meltzoff, J., (1998). *Critical thinking about research: Psychology and related fields*. Washington, DC: American Psychological Association.
- Meyer, B.J.F., Brandt, D.M., & G.J. Bluth. (1980). Use of top-level structure in text: Key for reading comprehension of ninth-grade students. *Reading Research Quarterly, 16, 1*, 72-103.
- Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review, 63*, 814-97.
- Monk, P. (2001). Mapping the future of argument. *Australian Financial Review, 16*, 8-9.
- Moreno, R., & Park, B. (2010). Cognitive load theory: Historical development and relation to other theories. In J.L. Plass, R. Moreno & R. Brünken (Eds.), *Cognitive load theory, 29-47*. New York: Cambridge University Press.
- Moseley, D., Baumfield, V., Elliot, J., Gregson, M., Higgins, S., Miller, J., & Newton, D. (2005). *Frameworks for thinking: A handbook for teaching and learning*. Cambridge: Cambridge University Press.
- Mullen, B. (1987). Self-attention theory: The effects of group composition on the individual. In B. Mullen & G. Goethals (Eds.), *Theories of group behaviour, 125-146*. New York: Springer-Verlag.

- Munch, J.M., Boller, G.W., & Swasy, J.L. (1993). The effects of argument structure and affective tagging on product attitude formation. *Journal of Consumer Research*, 20, 2, 294-302.
- Murray, D.J. (1968). Repeated recall in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 7, 2, 358-365.
- Myers, J.L. (1974). *Memory for prose material*. Amherst, Mass: University of Massachusetts.
- National Academy of Sciences, National Academy of Engineering, Institute of Medicine, (2005). *Rising above the gathering storm: Energising and employing America for a brighter economic future*. Committee on Prospering in the Global Economy for the 21st Century, Washington, DC.
- Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. New York: W.H. Freeman
- Newbern, D., Dansereau, D.F., Patterson, M.E., & Wallace, D.S. (1994). Toward a science of cooperation. Paper presented at the *Annual meeting of the American Educational Research Association*, New Orleans, April.
- Norris, S. P. (1994). The meaning of critical thinking test performance: The effects of abilities and dispositions on scores. *Critical thinking: Current research, theory, and practice*. Dordrecht, The Netherlands: Kluwer.
- Novak, J. D., & Cañas, A. J. (2006). The origins of the concept mapping tool and the continuing evolution of the tool. *Information Visualization*, 5, 3, 175–184.
- Oliver, K. (2009). An investigation of concept mapping to improve the reading comprehension of science texts. *Journal of Science Education and Technology*, 18, 5, 402-414.

- Osman, M.E., & Hannafin, M.J. (1992). Metacognition research and theory: Analysis and implications for instructional design. *Educational Technology Research and Development*, 40, 2, 83-99.
- Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: recent developments. *Educational Psychologist*, 38, 1, 1-4.
- Paivio, A. (1971). *Imagery and verbal processes*. Hillsdale, N.J.: Erlbaum.
- Paivio, A. (1986). *Mental representations: A dual-coding approach*. New York: Oxford University Press.
- Pascarella, E. T., & Terenzini, P. T. (1991). *How College Affects Students: Findings and Insights from Twenty Years of Research*. San Francisco: Jossey-Bass.
- Paul, R. (1987). Dialogical thinking: Critical thought essential to the acquisition of rational knowledge and passions. In J. Baron and R.J. Sternberg (Eds.), *Teaching Thinking Skills: Theory and Practice*, 127-148. New York: W.H. Freeman.
- Paul, R. (1993). *Critical thinking: What every person needs to survive in a rapidly changing world*. Rohnert Park, CA: Foundation for Critical Thinking.
- Paul, R., Elder, L., & Bartell, T. (1997). *California teacher preparation for instruction in critical thinking: Research findings and policy recommendations*. Sacramento, CA: State of California Commission on Teacher Credentialing.
- Perkins, D. N., Jay, E., & Tishman, S. (1993). Beyond abilities: A dispositional theory of thinking. *Merrill Palmer Quarterly*, 39, 1-1.
- Perkins, D. N., & Ritchhart, R. (2004). When is good thinking? In D. Y. Dai & R. J. Sternberg (Eds.), *Motivation, Emotion, and Cognition: Integrative*

- Perspectives on Intellectual Functioning and Development*, 351-384. Mahwah, NJ: Erlbaum.
- Perry, N. W., Huss, M.T., McAuliff, B.D. and Galas, J.M., (1996). An active-learning approach to teaching the undergraduate psychology and law course. *Teaching of Psychology*, 23, 2, 76–81.
- Petty, G. (1997). *How to be better at creativity*. London: Kogan Page.
- Phan, H.P. (2008). Predicting change in epistemological beliefs, reflective thinking and learning styles: A longitudinal study. *British Journal of Educational Psychology*, 78, 1, 75-93.
- Piaget, J. (1952). *The origins of intelligence in children*. New York: W.W. Norton & Co.
- Pintrich, P.R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P.R. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation*, 451- 502. San Diego, CA: Academic Press.
- Pintrich, P., & DeGroot, E. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82, 33-40.
- Pintrich, P.R., Smith, D.A.F., Garcia, T., & McKeachie, W.J. (1991). *A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Ann Arbor, MI: The University of Michigan, National Center for Research to Improve Post-secondary Teaching and Learning.
- Pollock, E., Chandler, P., & Sweller, J. (2002) Assimilating complex information. *Learning & Instruction*, 12, 61-86.

- Presseisen, B.Z. (2001). Thinking skills: meanings and models revisited. In A.L. Costa (Ed.), *Developing Minds: A Resource Book for Teaching Thinking*, 47-53. Alexandria, VA: ASCD Publications.
- Profeto-McGrath, J., (2003). The relationship of critical thinking skills and critical thinking dispositions of baccalaureate nursing students. *Journal of Advanced Nursing*, 43, 6, 569-577.
- Redish, E., Saul, J., & Steinberg, R. (1997). On the effectiveness of active-engagement microcomputer-based laboratories. *American Journal of Physics*, 65, 1, 45.
- Reed, C., Walton, D., & Macagno, F. (2007). Argument diagramming in logic, law and artificial intelligence. *The Knowledge Engineering Review*, 22, 1, 87–109.
- Reed, J.H., & Kromrey, J.D. (2001). Teaching critical thinking in a community college history course: Empirical evidence from infusing Paul’s model. *College Student Journal*, 35, 2, 201-215.
- Reeves, M.F. (1990). An application of Bloom's taxonomy to the teaching of business ethics. *Journal of Business Ethics*, 9, 7, 609-616.
- Rimiene, V. (2002). Assessing and developing students’ critical thinking. *Psychology Learning & Teaching*, 2, 1, 17-22.
- Robbins, S., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, 130, 2, 261-288.
- Robinson, D. H., & Kiewra, K. A. (1995). Visual argument: Graphic organizers are superior to outlines in improving learning from text. *Journal of Educational Psychology*, 87, 3, 455–467.
- Romiszowski, A. J. (1981). *Designing instructional systems*. New York: Nichols.

- Rosenthal, R. (2006). Applying psychological research on interpersonal expectations and covert communication in the classrooms, clinics, corporations and courtrooms. In S.I. Donaldson, D.E. Berger & K. Pezdek (Eds.), *Applied Psychology: New Frontiers and Rewarding Careers*, 107-118. Mahwah, NJ: Erlbaum.
- Rosenthal, R., & Jacobson, L. (1966). Teachers' expectancies: Determinants Of pupils' IQ gains. *Psychological Reports*, 19, 115-118.
- Roth, W.M., & Roychoudhury, A. (1994). Science discourse through collaborative concept mapping: New perspectives for the teacher. *International Journal of Science Education*, 16, 437-455.
- Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology*, 82, 498-504.
- Schommer, M. (1993). Epistemological development and academic performance among secondary students. *Journal of Educational Psychology*, 85, 1-6.
- Schommer, M., Calvert, C., Gariglietti, G., & Bajaj, A. (1997). The development of epistemological beliefs among secondary students: A longitudinal study. *Journal of Educational Psychology*, 89, 37-40.
- Schommer, M., Crouse, A., & Rhodes, N. (1992). Epistemological beliefs and mathematical text comprehension: Believing it's simple doesn't make it so. *Journal of Educational Psychology*, 84, 435-443.
- Schommer-Aikins, M. (2002). An evolving theoretical framework for an epistemological belief system. In B.K. Hofer & P.R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing*, 103-118. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

- Schommer-Aikins, M., & Hutter R. (2002). Epistemological beliefs and thinking about everyday controversial issues. *Journal of Psychology, 1*, 5-20.
- Schraw, G., Dunkle, M.E., & Bendixen, L.D. (1995). Cognitive processes in well-defined and ill-defined problem-solving. *Applied Cognitive Psychology, 9*, 523-538.
- Scott, J.N., Market, R.J., & Dunn, M.M. (1998). Critical thinking: Change during medical school and relationship to performance in clinical clerkships. *Medical Education, 32*, 14-18.
- Scott, T. (2003). Bloom's taxonomy applied to testing in computer science classes. *Journal of Computing Sciences in Colleges, 19*, 1, 267-274.
- Scriven, M. (1976). *Reasoning*. Point Reyes: Edgepress.
- Sherrard, M., & Czaja, R. (1999). Extending two cognitive processing scales: Need for cognition and need for evaluation for use in a health intervention. *European Advances in Consumer Research, 4*, 135-142.
- Slavin, R.E. (1995). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology, 21*, 43-69.
- Smith, S.M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review, 8*, 2, 203-220.
- Solon, T. (2007). Generic critical thinking infusion and course content learning in Introductory Psychology. *Journal of Instructional Psychology, 34*, 2, 95-109.
- Sternberg, R.J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. Cambridge: Cambridge University Press.

- Suliman, W.A. (2006). Critical thinking and learning styles of students in conventional and accelerated programmes. *International Nursing Review*, 53, 1, 73-79.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12, 257-285.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4, 295-312.
- Sweller, J. (1999) Instructional design in technical areas. *Australian Education Review No. 43*. Victoria: Acer Press.
- Sweller, J. (2005). The redundancy principle. In R.E. Mayer (Ed.), *The Cambridge handbook of multimedia Learning*, 159-167. New York: Cambridge University Press.
- Sweller, J. (2010). Cognitive load theory: Recent theoretical advances. In J.L. Plass, R. Moreno & R. Brünken (Eds.), *Cognitive Load Theory*, 29-47. New York: Cambridge University Press.
- Sweller, J., & Chandler, P. (1991). Evidence for cognitive load theory. *Cognition & Instruction*, 8, 4, 351-362.
- Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology: General*, 119, 2, 76-192.
- Taube, K.T. (1997). Critical thinking ability and disposition as factors of performance on a written critical thinking test. *Journal of General Education*, 46, 129-164.
- Taylor, B.M. (1982). Text structure and children's comprehension and memory for expository material. *Journal of Educational Psychology*, 74, 323-340.

- Taylor, B. M., & Beach, R. W. (1984). The effects of text structure instruction on middle-grade students' comprehension and production of expository text. *Reading Research Quarterly, 19*, 134-146.
- Thomson, A. (2009). *Critical Reasoning: A practical introduction (3rd ed.)*. UK: Routledge.
- Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied, 3*, 4, 257 -287.
- Toplak, M.E., & Stanovich, K.E. (2002). The Domain Specificity and Generality of Disjunctive Reasoning: Searching for a Generalizable Critical Thinking Skill. *Journal of Educational Psychology, 94*, 1, 197-209.
- Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- Tulving, E. (1984). Precis of elements of episodic memory. *Behavioral and Brain Sciences, 7*, 223–268.
- Tulving, E., & Thompson, D.M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352-373.
- Twardy, C.R. (2004). Argument maps improve critical thinking. *Teaching Philosophy, 27*, 2, 95-116.
- Twigg, C. (2002). Quality, cost and access: The case for redesign. In M. S. Pittinsky (Ed.), *The Wired Tower*, 111-143. Upper Saddle River, NJ: Prentice-Hall.
- U.S. National Research Council (2002). *Technology and assessment: Thinking ahead*. Washington, DC: Board on Testing and Assessment Center for Education.
- Valenzuela, J., Nieto, A.M., & Saiz, C. (2011). Critical Thinking Motivational Scale: a contribution to the study of relationship between critical thinking and motivation. *Journal of Research in Educational Psychology, 9*, 2, 823-848.

van den Braak, S.W., van Oostendorp, H., Prakken, H., & Vreeswijk, G.A.W. (2006).

A critical review of argument visualization tools: Do users become better reasoners? In *Workshop notes of the ECAI-06: 6th Workshop on Computational Models of Natural Argument*, Riva del Garda, Italy, August 28-29.

van Eemeren, F.H., Grootendorst, R., Henkemans, F.S., Blair, J.A., Johnson, R.H., Krabbe, E.C.W., Planitin, C., Walton, D.N., Willard, C.A., Woods, J., & Zarefsky, D. (1996). *Fundamentals of argumentation theory: a handbook of historical backgrounds and contemporary developments*. New Jersey: Lawrence Erlbaum Associates.

van Gelder, T.J. (2000). Learning to reason: A Reason!Able approach. In C. Davis, T. J. van Gelder & R. Wales (Eds.), *Cognitive Science in Australia, 2000: Proceedings of the Fifth Australasian Cognitive Science Society Conference*. Adelaide: Causal.

van Gelder, T. J. (2001). How to improve critical thinking using educational technology. In G. Kennedy, M. Keppell, C. McNaught & T. Petrovic (Eds.), *Meeting at the Crossroads: Proceedings of the 18th Annual Conference of the Australian Society for Computers in Learning in Tertiary Education*, 539–548. Melbourne: Biomedical Multimedia Unit, University of Melbourne.

van Gelder, T.J. (2002). Argument mapping with Reason!Able. *APA Newsletter: Philosophy & Computers*, 2, 1, 85-90.

van Gelder, T.J. (2003). Enhancing deliberation through computer supported argument mapping. In P. A. Kirschner, S. Buckingham-Shum, & C. Carr (Eds.), *Visualizing argumentation: Software tools for collaborative and educational sense-making*, 97–115, London: Springer-Verlag.

- van Gelder, T.J. (2007). The rationale for Rationale™. *Law, Probability & Risk*, 6, 23-42.
- van Gelder, T.J., Bissett, M., & Cumming, G. (2004). Enhancing expertise in informal reasoning. *Canadian Journal of Experimental Psychology* 58, 142-52.
- van Gelder, T.J., & Rizzo, A. (2001). Reason!Able across curriculum. In *2001: Is IT an Odyssey in Learning? Proceedings of the 2001 Conference of ICT in Education*, Victoria, Australia.
- van Merriënboer, J.J.G., & Ayres, P. (2005). Research on cognitive load theory and its design implications for e-learning, educational technology, research and development. *Educational Technology Research and Development*, 53, 3, 5-13.
- van Merriënboer, J.J.G., Kirschner, P.A., & Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. *Educational Psychology*, 38, 5-13.
- Varian, H., & Lyman, P. (2003). *How much information?* Berkeley, CA: School of Information Management & Systems, UC Berkeley.
- Vicari, S., Pasqualetti, P., Marotta, L., & Carlesimo, G.A. (1999). Word-list learning in normally developing children: Effects of semantic organization and retention interval. *Italian Journal of Neurological Sciences*, 20, 119-128.
- Walton, D. N. (2006). *Fundamentals of critical argumentation*. Cambridge: Cambridge University Press.
- Wang, L. (1993). The differential aptitude test: A review and critique. *Annual Meeting of the Southwest Educational Research Association*, Austin, TX.
- Wankat, P. (2002). *The effective, efficient professor: Teaching, scholarship and service*. Boston: Allyn and Bacon.

- Watson, G., & Glaser, E.M. (1980). *Watson-Glaser critical thinking appraisal*. New York: Psychological Corporation.
- Watson, G., & Glaser, E. M. (1994). *Watson-Glaser critical thinking appraisal: Form S manual*. San Antonio: TX: Psychological Corporation.
- Wechsler, D. (1974). *WISC-R Wechsler intelligence scale for child revised – Manual*. San Antonio, TX: The Psychological Corporation.
- Wegerif, R. (2002). *Literature review in thinking skills, technology and learning: Report 2*. Bristol: NESTA Futurelab.
- Wegerif, R., & Dawes, L. (2004). *Thinking and learning with ICT: Raising achievement in the primary classroom*. London: Routledge.
- Weinert, F.E. (1987). Introduction and overview: Metacognition and motivation as determinants of effective learning and understanding. In F.E. Weinert & R.H. Kluwe (Eds.), *Metacognition, motivation and understanding*, 1-18. Hillsdale, NJ: Erlbaum.
- Wentzel, K.R. (2002). Are effective teachers like good parents? Teaching styles and student adjustment in early adolescence. *Child Development*, 73, 1, 287-301.
- Whately, R. (1836). *Elements of Logic*. London: Fellowes.
- Wigmore, J.H. (1913). *The principles of judicial proof: as given by logic, psychology and general experience and illustrated in judicial trials*. Boston: Little, Brown & Co.
- Wilkinson J.M. (1996). *Nursing process: A critical thinking approach (2nd ed.)*. Menlo Park: Addison Wesley Nursing.
- Willingham, D.T. (2002). How we learn – Allocating student study time: Massed versus distributed practice. *American Educator*, 26, 2, 37-39.

- Wilson, B.A., & Baddeley, A.D. (1988). Semantic, episodic and autobiographical memory in a post-meningitic amnesic patient. *Brain Cognition*, 8, 31–46.
- Wittrock, M.C. (1986). Students' thought processes. In M.C. Wittrock (Ed.), *Handbook of research on teaching (3rd Ed.)*. New York: Macmillan.
- Wood, D.J., Bruner, J.S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17, 89-100.
- Wood, P.K., (1993). *Context and development of reflective thinking: A secondary analysis of the structure of individual differences*. Unpublished manuscript. Missouri: University of Missouri at Columbia.
- Wood, R. (1977). Multiple choice: A state of the art report. *Evaluation in Education*, 1, 191-280.
- Woodman, G.F., Vecera, S.P., & S.J. Luck. (2003). Perceptual organization influences visual working memory. *Psychonomic Bulletin & Review*, 10, 1, 80-87.
- Yin-Kum, L. (1995). Effects of text structures on recall. *Education Journal*, 23, 1, 15-133.
- Zmach, C.C., Sanders, J., Patrick, J. D., Dedeoglu, H., Charbonnet, S., Henkel, M., Zhihui, F., Lamme, L.L., & Pringle R. (2007). Infusing reading into science learning. *Educational Leadership*, 64, 4, 62-66.

APPENDIX A

Memory Test Scoring Manual

The immediate recall tests used in Experiments 1 and 2 consisted of 14 items and were scored on a scale of 0-28. The immediate and delayed recall tests used in Experiments 3 and 4 consisted of 12 items and were scored on a scale of 0-24. The tests used in Experiment 3 and 4 consisted of two items less than the previous experiments due to added time constraints in the experiments' design. The delayed recall test used in Experiment 3 was identical to its immediate recall counterpart, except that the questions were counter-balanced in order to avoid practice effects. Three different topics were used: *Computers Can Think* (Experiments 1 and 2), *Aggression is biologically caused* (Experiment 3) and *There was nothing wrong with the United States Supreme Court's decision to uphold the individual's right to bear arms* (Experiment 4).

Memory Test Scoring Manual

Each item was scored on a scale from 0-2. A score of zero was given to a response that was not attempted, was not correct, had little or no relevance to the question, or did not answer the question to the extent that lead the researcher to believe that the students understood the 'gist' (Kintsch & van Dijk, 1975) of the statement. A score of 1 point was awarded to responses which replicated the 'gist' of the correct answer or replicated a majority of the important aspects within the target response. For example, a proposition that correctly answered a question may have contained 3 key points. A response which identified one of those key points correctly along with the 'gist' of the proposition was awarded a point. Responses that consisted of fragmented sentences or just a few words may have received a single point, provided it identified at least one key point as well as the 'gist' of the answer. Similarly, responses with 2 key points from the proposition (leaving out one key point) also received one point.

To be credited with 2 points on an item, the response must at least have reported the gist, along with 2 of 3 key points. If the answer consisted of 2 key points only, both points must have been addressed. Responses that consisted of fragmented sentences or just a few words may have received 2 points, provided they addressed the key points and the 'gist' of the answer.

Scoring Examples

Examples of scoring are provided below, with questions extracted from the argument topic, *Computers Can Think*:

Question:

Computers can't reason scientifically as humans do,

Correct Answer:

But, Computers have already reasoned scientifically through software programmes e.g. Dendral has discovered how to synthesise previously unknown chemical compounds as well as entirely new rules of chemical analysis (Buchanan, 1976). This answer contains 3 key points: (1) Computers have already reasoned scientifically through software programmes e.g. (2) Dendral has discovered how to synthesise previously unknown chemical compounds (3) as well as entirely new rules of chemical analysis (Buchanan, 1976).

Zero points awarded to:

- a) But, (BLANK)
- b) But, Computers can't reason scientifically because they have a fixed vocabulary.
- c) But, nothing is intrinsically a digital computer.
- d) Computers have already reasoned scientifically

Reasons:

- a) Did not attempt the question.
- b) Offered a reason, not an objection (an objection was asked for)
- c) Irrelevant to the question.
- d) Simply restates a portion of the question.

One point awarded to:

- a) But, Computers have already reasoned scientifically through software programmes like Dendral.
- b) But, Dendral can synthesise unknown chemical compounds, so it reasons scientifically.
- c) But, Dendral can come up with new rules for chemical analysis.
- d) But, Computers have already reasoned scientifically through programmes (Buchanan, 1976).
- e) But, Dendral can scientifically reason.
- f) But, Dendral (Buchanan, 1976).

Reasons:

- a) Offers the gist, plus an example (e.g. a key point).
- b) Gives an example and justifies the 'gist'.
- c) Gives an example and a key point.
- d) Gives the 'gist' and an example.
- e) Uses the example to show the 'gist'.
- f) Though it doesn't provide the 'gist' it does provide two key points.

Two points awarded to:

a) But, Computers have already reasoned scientifically through software programmes (e.g. Dendral) has discovered how to synthesise previously unknown chemical compounds as well as entirely new rules of chemical analysis (Buchanan, 1976).

b) But, Dendral has discovered how to synthesise previously unknown chemical compounds as well as entirely new rules of chemical analysis (Buchanan, 1976).

c) But, computers have already reasoned scientifically through software programmes (e.g. Dendral) has discovered how to synthesise previously unknown chemical compounds.

d) But, Computers have already reasoned scientifically through software programmes (e.g. Dendral) has come up with entirely new rules of chemical analysis.

e) But, Computers have already reasoned scientifically through software programmes e.g. Dendral (Buchanan, 1976).

Reasons:

- a) Answered correctly verbatim.
- b) Provides a detailed answer that requires no 'gist' as it uses two examples and a reference.
- c) Provides the 'gist' and an example.
- d) Provides the 'gist' and an example.
- e) Provides the 'gist', an example and a reference.

Another example for all situations is provided below, with questions extracted from the argument topic, People are basically selfish:

Question:

Darwin's Theory of natural selection implies that an individual will sacrifice resources, even his or her life, for a close relative (e.g. a brother or offspring as their relative shares 50% of their genes).

Correct Answer:

But, just because sacrifices could be made for relatives, does not mean sacrifices will be made for unrelated individuals.

It is important to note that this answer consists of two key points, instead of 3 like the previous example. Therefore, to gain one point, only one factor has to be described; and to gain two points, both factors must be described.

Zero points given to:

a) But, (BLANK)

- b) But, Situational variables influence whether or not a person helps another.
- c) But, 50% of their genes is not enough to sacrifice resources.

Reasons:

- a) No attempt made to answer the question.
b) Irrelevant to the question.
c) Simply restates a portion of the question.

One point awarded to:

- a) But, this doesn't mean sacrifices will be made for others.
- b) But, won't do it for other individuals.
- c) But, won't sacrifice for others outside of kinship.

Reason:

They all address a relevant rebuttal, citing one key point provided in the proposition.

Two Points awarded for:

But, just because sacrifices could be made for relatives, does not mean sacrifices will be made for unrelated individuals.

But, may sacrifice for family, doesn't mean they'll do it for outsiders.

Reasons:

- a) Answered correctly verbatim.
b) Answered correctly by addressing both key points.

'Key Points'

The previous examples reveal how the tests are scored. However, in order to score them appropriately, the test marker must also recognise what constitutes a 'key point'. All questions on the memory tests contain either 2 or 3 key points. Below is an example of an answer that consists of 2 key points, which responds to the item: *Genes have been discovered that code for levels of testosterone; and testosterone influences aggression levels, because:*

- (1) Men generally have higher levels of testosterone than women, / (2) and are also more aggressive than women (Knight 1996).

The forward slash denotes where one key point finishes and the next starts. Key points consist of single idea unit. Cut-off points (i.e. the slashes) are generally placed after full-stops, before conjunctions or in the middle of a listing. Another example of a 2 key point response, which responds to the item: *By collective self-defense, if many*

Americans own guns, it is better for the general welfare of the U.S in case they are invaded by a foreign power:

But, (1) Given the power of the American military, / (2) it is absurd to believe that ground invasion of the U.S. could happen.

The same guidelines apply to 3 key point responses. For example: *Residents of homes where a gun is present are 5 times more likely to experience a suicide than residents of homes without guns (Kellermann et al., 1992), but:*

(1) Both the U.S. and Ireland still fall into the same suicide rate category (6.5 - 13 per 100,000 people), / (2) even though substantially less Irish own guns than Americans, / (3) with regards to both rate and percentage.

Another example of this 3 key point answer: *Humans have a biologically based empathy mechanism that is automatically activated when a person in need is encountered (Hoffman, 1981), because:*

(1) The visual processing part of the brain / (2) is connected to the emotional part of the brain, / (3) hence the "seeing with feeling".

All questions, answers and key point cut-off markers (i.e. forward slashes) are presented below:

Computers Can Think

1. Machines can't have emotions because,

Emotional beings require a limbic system / or the human equivalent of a limbic system.

Machines cannot love / or be loved.

2. Computers can't reason scientifically as humans do,

But computers have already reasoned scientifically through software programmes / e.g. Dendral has discovered how to synthesise previously unknown chemical compounds / as well as entirely new rules of chemical analysis (Buchanan, 1976).

3. The brain is a machine that can think,

But nothing is intrinsically a digital computer / in that the syntactic structures that define computers / are not intrinsic to the physics of the brain.

4. Formal programmes can be realised in multiple physical media because:

The program is defined solely in terms of its formal syntactic structure, / its mode of physical implementation is irrelevant.

The same formal program could be realised in a digital computer, / the human brain, a beer can / or a roll of toilet tissue.

5. Machines can exhibit free will by random selection,

But, responsibility is necessary for free will / and randomization sacrifices that responsibility.

6. Scientific reasoning requires social agreement because,

Scientific laws and data don't follow applications of an algorithm, / but are developed through a quasi-political process of negotiation.

Computers can't agree socially / as they are not members of society.

7. Deterministic philosophy theorises that humans are programmed by nature,

But, not everyone has a deterministic outlook on life, / nor is determinism accepted as fact.

8. Computers cannot introduce new theoretical terms or principles, thus eliminating originality,

But, computers can introduce new terms using automated principles / of explanatory adequacy.

9. Computers cannot have free will but,

Humans also / lack free will.

Machines can exhibit free will by random selection, / such as through selection of values or noises.

10. The brain is a machine that can think,

Because the brain's neurobiological processes / are similar to the information processes of a computer.

Aggression is biologically caused.

1. Aggression is biologically caused.

But, the human environment / influences aggressive tendencies.

2. Genes have been discovered that code for levels of testosterone; and testosterone influences aggression levels.

Because, men generally have higher levels of testosterone than women, / and are also more aggressive than women (Knight 1996).

Because female prison inmates who displayed unprovoked violence / also had very high levels of testosterone (Dabbs, 1998).

3. High blood alcohol increases aggression.

Because alcohol increases GABA in the cortex / and reduces the capacity of the frontal cortex / to inhibit midbrain aggressive impulses.

Because marital violence decreased in couples who completed behavioral marital therapy for alcoholism / and remained sober during follow-up (O'Farrell, 1995).

4. Codeine is found in prescription strength cough and cold medications, and codeine can trigger aggression (Spiga, 1990).

But prescription strength cough and cold medications make some people very sleepy and lethargic, / but not aggressive.

5. Inbreeding and selective breeding illustrate the role of genes and hereditary factors in aggression.

Because inbreeding can create unstable temperaments / that are associated with aggressive tendencies.

Because by selective breeding / aggressive and passive strains of mice can be created / (Lagerspetz, 1979).

6. Parents can influence levels of aggression in their children.

Because in extreme cases of child abuse, parents who beat their children / often turn out to have been victims of child abuse themselves (Parke, 1975).

Because experiments have demonstrated that children tend to imitate their parents' behaviour. / For example, after watching parents beat a playroom doll with a stick, children often did the same / (Bandura, 1977).

7. Television violence increases aggression in the viewer.

Because, a Canadian town with no television until the late 80's / saw the incidence of playground aggression double / within months. (Williams, 1986).

But, Feshbach & Singer (1971) found that being subjected to consistent violent television programmes over a six-month period / didn't increase levels of aggression in teenagers.

Gun Control: There was nothing wrong with the United States Supreme Court's decision to uphold the individual's right to bear arms.

1. Victims who defended themselves with guns were less likely to report being injured than those who either defended themselves by other means or took no self-protective measures at all (Roth, 1994).

But, Zimring (1991) reported that the use of a firearm to resist a violent assault actually / increases the victim's risk of injury and death.

2. Thousands of children die annually in gun accidents.

But, the often repeated claim that 12 children per day die from gun violence / includes "children" up to 20 years of age, / the great majority of whom are young adult males who die in gang-related violence.

But, gun accidents involving children / are actually at record lows.

But, more children die each year in accidents involving bikes, / space heaters / or drownings.

3. Guns are a main cause for many unnecessary deaths.

But states that allow registered citizens to carry concealed weapons / have lower crime rates than those that don't (Lampo, 2000).

But, there is no correlation between gun control laws and deaths by gun across a spectrum of nations and cultures. / For example, Israel and Switzerland allow gun licenses, widespread carrying of concealed firearms, / and yet, they have rates of homicide that are much lower than those of the U.S., despite rates of home firearm ownership that are at least as high (Kellermann, 1993).

4. Keeping a gun in the home is a sound method of self-defense against intruders.

But research has shown that a gun kept in the home is 43 times more likely to kill / a member of the household, or friend, / than an intruder (Kellermann & Reay, 1986).

Because victims who defended themselves with guns / were less likely to report being injured / than those who either defended themselves by other means or took no self-protective measures at all (Roth, 1994).

5. Gun control could prohibit hunting, a traditional, recreational activity.

But, you do not need a gun to / hunt.

6. The National Rifle Association (of America) argue that if law-abiding citizens have guns, they are safer from criminals, in that they bring crime rates down.

But in countries where guns are greatly restricted, such as Ireland, Great Britain or Japan, / deaths from guns are very rare, / especially compared to the United States.

7. Residents of homes where a gun is present are 5 times more likely to experience a suicide than residents of homes without guns (Kellermann et al., 1992).

But, both the U.S. and Ireland still fall into the same suicide rate category (6.5-13 per 100,000 people), / even though substantially less Irish own guns than Americans, / with regards to both rate and percentage.

8. By collective self-defense, if many Americans own guns, it is better for the general welfare of the U.S in case they are invaded by a foreign power.

But, given the power of the American military, / it is absurd to believe that ground invasion of the U.S. could happen.

APPENDIX B

Study Materials & Tests Used in Study 1

1. Computers Can Think

1.1 Small Argument Map

1.2 Large Argument Map

1.3 Small Monochrome Argument Map

1.4 Large Monochrome Argument Map

1.5 Small Text

1.6 Large Text

1.7. Recall Test

1.8 Comprehension Test

2. Aggression is Biologically Caused

2.1 Small Argument Map

2.2 Large Argument Map

2.3. Small Text

2.4 Large Text

2.5 Immediate Recall Test

2.6 Delayed Recall Test

3. Gun Control

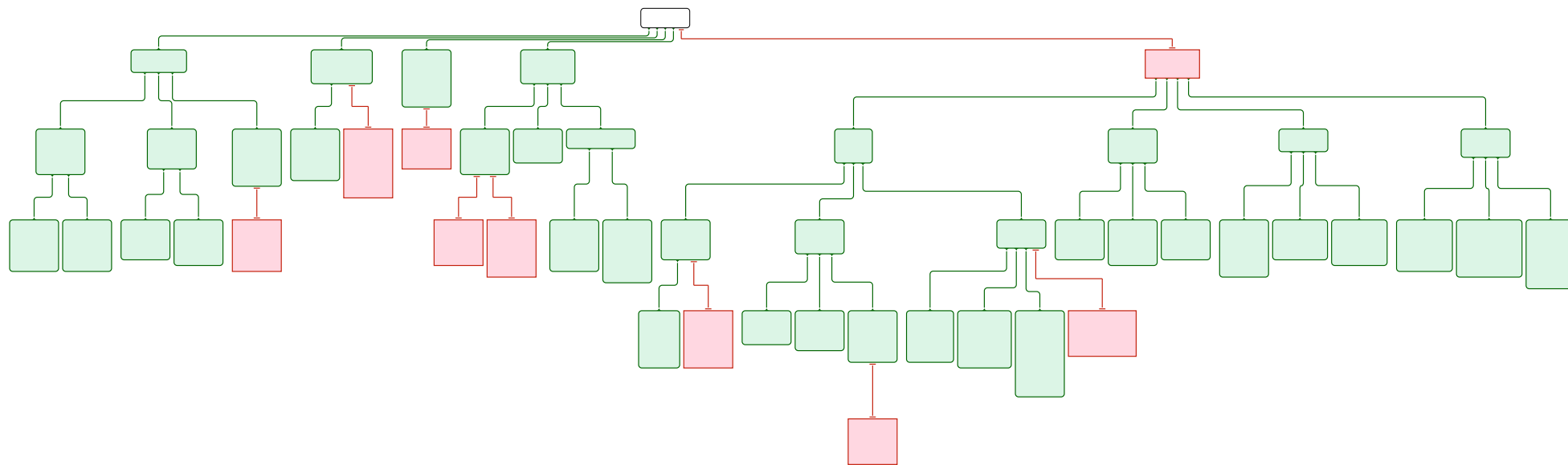
3.1 Text for Active Learning

3.2 Active Learning Argument Map

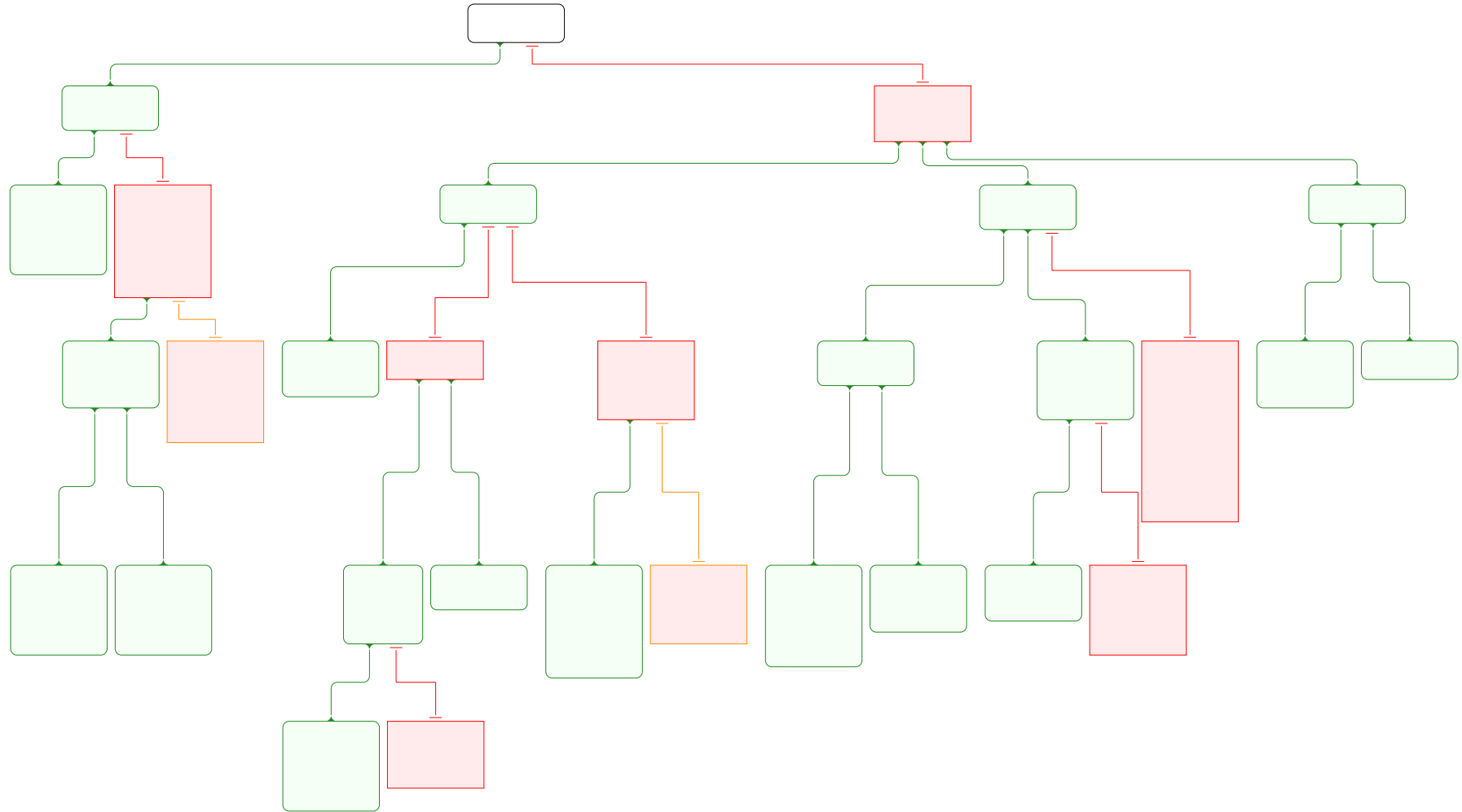
3.3 Active Learning Outline

3.4 Recall Test

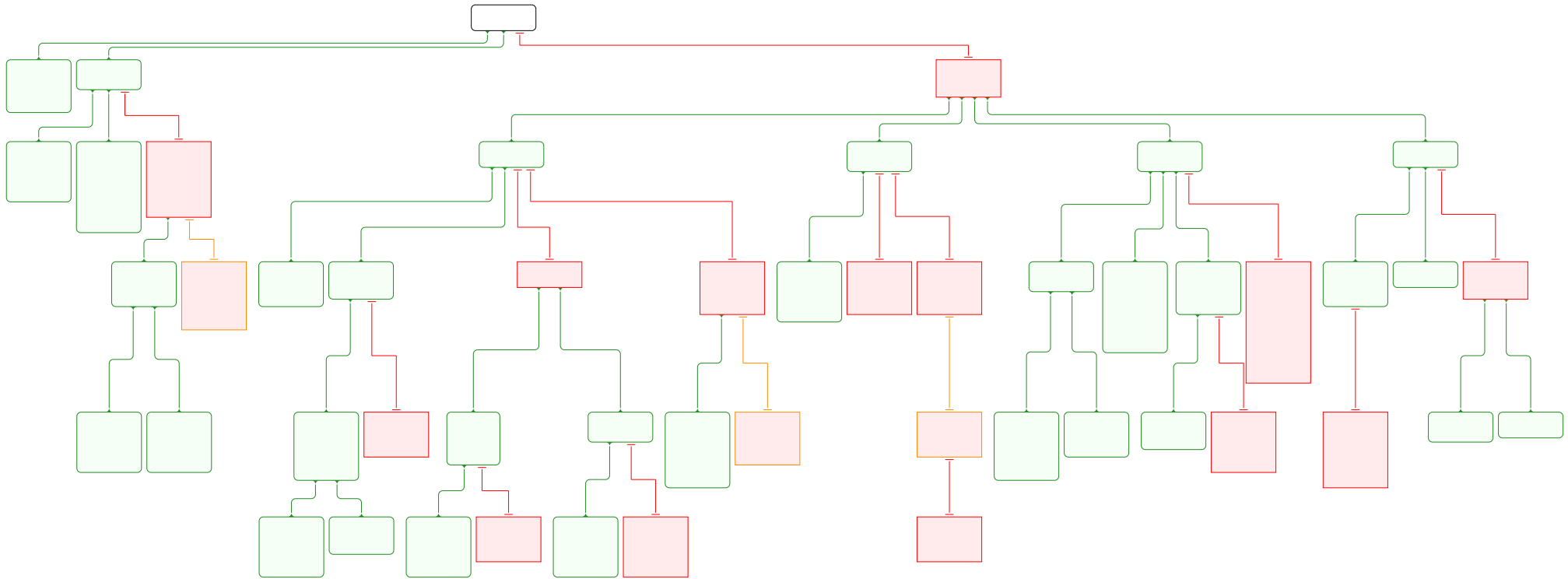
1.2 Large Argument Map



1.3 Small Monochrome Argument Map



1.4 Large Monochrome Argument Map



1.5 Small Text: *Computers Can Think*

The question of artificial intelligence has occupied both technologists and philosophers over the past century. In 1950, Alan Turing declared that it is possible for machines to think. He argued that in fact, the brain is a machine that can think because the brain's neurobiological processes are similar to the information processes of a computer.

Others have argued that this is not the case, because nothing is intrinsically a digital computer in that the syntactic structures that define computers are not intrinsic to the physics of the brain. This is because formal programmes can be realised in multiple physical media, e.g. this formal programme can be realised in a digital computer, the human brain, a beer can or a roll of toilet tissue. The programme itself is defined solely in terms of its formal structure- its mode of physical implementation is irrelevant. However, programmes are not universally realised based on implemented input, that is, computers behave according to the input given to them.

There are also various arguments against Turing's belief that computers will one day be able to think. Such reasons are that computers cannot reason scientifically as humans do, nor can they have emotions. First and foremost, computers cannot have free will. Computers can only exhibit the free will of their programmer. But at the same time, humans also lack free will, because free will is an illusion of experience. Deterministic philosophy theorises that humans are programmed by nature. One who accepts determinism believes that nature has programmed you to behave in certain ways in certain contexts. However, not everyone has a deterministic outlook on life, nor is determinism accepted as fact. Another objection to the free will argument is that machines can exhibit free will by random selection, such as through selection of values or noises. Free will arises from random selection of alternatives in situations where there is no preference for either, or any choice. However, responsibility is necessary for free will and randomization sacrifices that responsibility.

Another argument against the notion of artificial intelligence is that machines cannot have emotions. This is because emotional beings require a limbic system or the human equivalent of a limbic system. Plus, machines cannot love or be loved.

One final argument for why computers cannot think is because they cannot reason scientifically as humans do. Scientific reasoning requires social agreement and computers cannot agree socially as they are not members of society. Scientific laws and data do not follow applications of an algorithm but are developed through a quasi-political process of negotiation. However, computers have already reasoned scientifically through software programmes, e.g. *Dendral* has discovered how to synthesise previously unknown chemical compounds as well as entirely new rules of chemical analysis (Buchanan, 1976). However, computers cannot introduce new theoretical terms or principles. Therefore, originality is eliminated because computers have a fixed vocabulary and conceptual apparatus. However, as was previously mentioned, computers can introduce new terms using automated principles of explanatory adequacy. In conclusion, one can easily see that, though the notion of artificial intelligence is a possibility, various arguments suggest it is unlikely.

1.6 Large Text: *Computers Can Think*

The question of artificial intelligence has occupied both technologists and philosophers over the past century. In 1950, Alan Turing declared that it is possible for machines to think. He argued that a computational system can possess all the important elements of thinking and understanding. The brain is a machine that can think and the brain's neurobiological processes are similar to the information processes of a computer. Once we have a sufficient understanding of the laws of physics and the structure of the brain, we will be able to precisely simulate the operations of the brain with those of a computer.

Others have argued that this is not the case, because nothing is intrinsically a digital computer, in that the syntactic structures that define computers are not intrinsic to the physics of the brain. This is because formal programmes can be realised in multiple physical media, e.g. this formal programme can be realised in a digital computer, the human brain, a beer can or a roll of toilet tissue. The programme itself is defined solely in terms of its formal structure- its mode of physical implementation is irrelevant. However, programmes are not universally realised based on implemented input, that is, computers behave according to the input given to them.

There are also various arguments against Turing's belief that computers will one day be able to think. Such reasons are that computers cannot reason scientifically as humans do, they cannot have emotions, and so on. First and foremost, computers cannot have free will. An agent is free only if it can change its goals, which computers are unable to do. Computers can only exhibit the free will of their programmer. For example, programmed robots cannot have psychological states. They can only have the psychological state of their programmer. Therefore, their will is that of their programmer. Conversely, humans cannot be reprogrammed in the arbitrary way that robots can be. Another way of saying this is that a robot behaves in the same way a phonograph plays a record, they are programmed to behave in a certain way. However, some computers can program themselves, e.g. automatic program systems (APs).

At the same time, humans also lack free will, because free will is an illusion of experience. Choices are made based on previously held schemata (not will), of which the person is not consciously aware. However, choices are not made randomly, but are based upon both an individual's emotion and logic which is a function of one's own motivation, or will. Deterministic philosophy theorises that humans are programmed by nature. One who accepts determinism believes that nature has programmed you to behave in certain ways in certain contexts. However, not everyone has a deterministic outlook on life, nor is determinism accepted as fact. Another objection to the free will argument is that machines can exhibit free will by random selection, such as through selection of values or noises. Free will arises from random selection of alternatives in situations where there is no preference for either, or any choice. However, responsibility is necessary for free will and randomization sacrifices that responsibility.

Another line of argument against artificial intelligence is that God prohibits computers from thinking. Because only entities with immortal souls, given to us by God, can think as the soul permits us to make choices on moral grounds. However, the view that only humans have souls is as arbitrary as the view that men have souls and women do not. Similarly, God's existence is debatable- as there is no universally accepted proof for or against the existence of God, even though a majority of the

human population believe, or have faith in a divine force of some kind. Regardless of this, faith, itself, is a type of cognition that humans possess and which computers do not.

Another argument against the notion of artificial intelligence is that machines cannot have emotions. Simply, machines cannot love or be loved. Furthermore, emotional beings require a limbic system or the human equivalent of a limbic system. But, artificial minds mimic animal evolution, thus the development of artificial limbic systems should be attainable in the next 20-50 years (Stonier, 1992). Another point to consider is that feelings are informational signals in a cognitive system, because emotions are cognitive schemata and emotion, itself, is a type of information processing.

One final argument for why computers cannot think is because they cannot reason scientifically as humans do. Scientific reasoning requires social agreement and computers cannot agree socially as they are not members of society. Scientific laws and data do not follow applications of an algorithm but are developed through a quasi-political process of negotiation. Also, computers cannot adequately evaluate hypotheses as the computer would need a criterion preference to choose between hypotheses that account for data equally well. Nor can computers introduce new theoretical terms or principles. Therefore, originality is eliminated, because computers have a fixed vocabulary and conceptual apparatus. However, some have argued computers can introduce new terms using automated principles of explanatory adequacy. Along those lines, computers have already reasoned scientifically through software programmes, e.g. *Dendral* has discovered how to synthesise previously unknown chemical compounds as well as entirely new rules of chemical analysis (Buchanan, 1976). In conclusion, one can easily see that, though the notion of artificial intelligence is a possibility, various arguments suggest it is unlikely.

1.7. Recall Test

Student ID _____

Gender _____

Please select the study material you used.

30 BIT – Coloured Map

30 BIT – Black-and- White Map

1 Page Text

50 BIT – Coloured Map

50 BIT – Black-and- White Map

2 Page Text

Complete each sentence, word for word, from what you remember from studying the material given to you, as best you can.

1. Machines can't have emotions,

Because _____

Because _____

2. Computers can't reason scientifically as humans do,

But _____

3. The brain is a machine that can think,

But _____

4. Formal programmes can be realised in multiple physical media,

Because _____

Because _____

5. Machines can exhibit free will by random selection,

But _____

6. Scientific reasoning requires social agreement,

Because _____

Because _____

7. Deterministic philosophy theorises that humans are programmed by nature,

But _____

8. Computers cannot introduce new theoretical terms or principles, thus eliminating originality,

But _____

9. Computers cannot have free will,

But _____

But _____

10. The brain is a machine that can think,

Because _____

1.8 Comprehension Test

Student ID _____

Gender _____

Please select the study material you used.

30 BIT – Coloured Map

30 BIT – Black-and- White Map

1 Page Text

50 BIT – Coloured Map

50 BIT – Black-and- White Map

2 Page Text

Please read each question carefully. Please indicate whether each statement is a reason or an objection to the central claim from the material you have just read: “Computers can think”.

1. One who accepts determinism believes that nature has programmed you to behave in certain ways in certain contexts.

Reason

Objection

2. Machines cannot love or be loved.

Reason

Objection

3. Computers can introduce new terms using automated principles of explanatory adequacy.

Reason

Objection

4. Free will is an illusion of experience.

Reason

Objection

5. A program is defined solely in terms of its formal syntactic structure; its mode of physical implementation is irrelevant.

Reason

Objection

6. A formal program can be realised in a digital computer, the human brain, toilet tissue or a beer can.

Reason

Objection

7. Computers cannot agree socially as they are not members of society.

Reason

Objection

8. Computers have a fixed vocabulary and conceptual apparatus.

Reason

Objection

9. Not everyone has a deterministic outlook on life.

Reason Objection

10. Randomization sacrifices responsibility.

Reason Objection

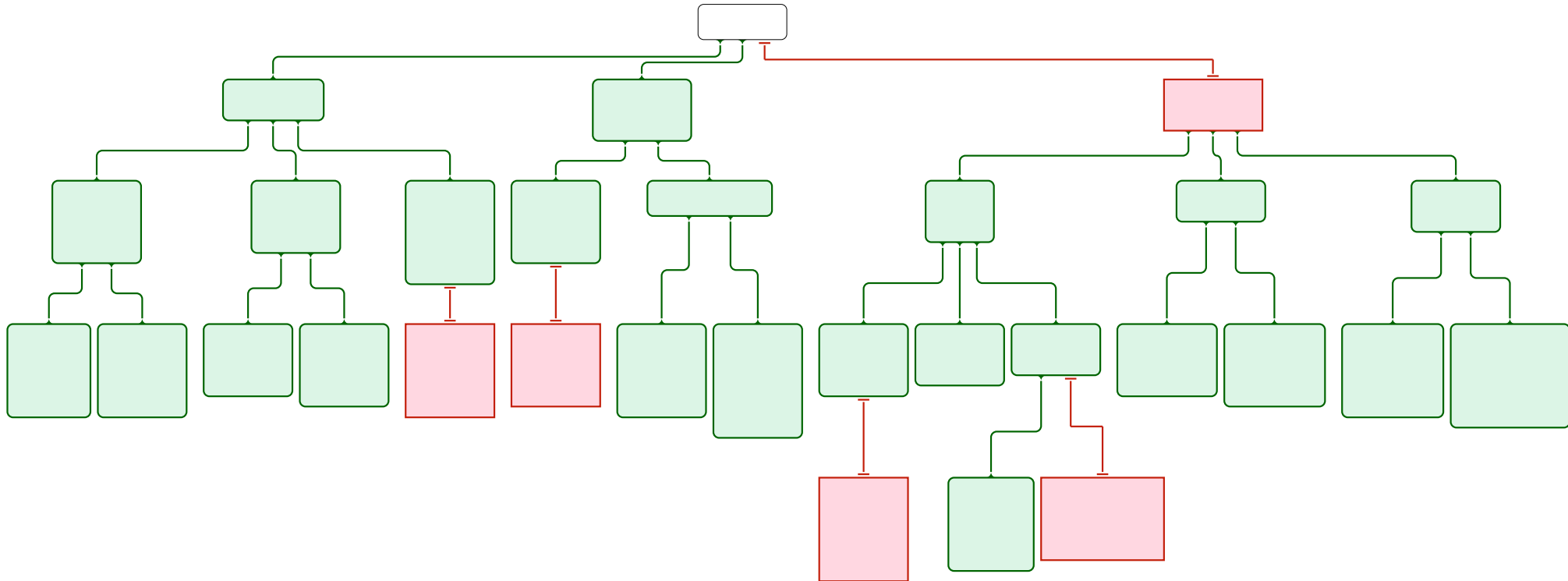
11. Emotional beings require a limbic system.

Reason Objection

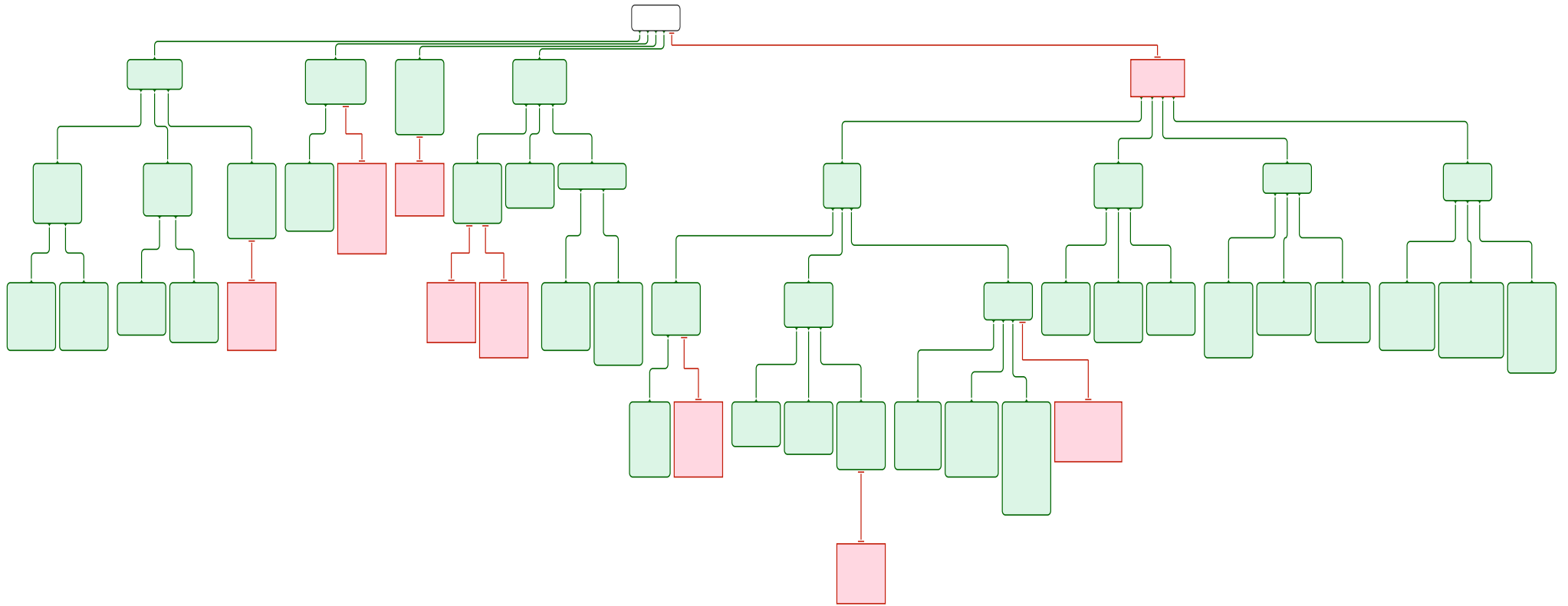
12. Scientific laws and data don't follow applications of an algorithm but are developed through a quasi-political process of negotiation.

Reason Objection

2.1 Small Argument Map



2.2 Large Argument Map



2.3. Small Text: *Aggression is Biologically Caused*

It is often thought that aggression is biologically caused. This is because genetic and hereditary factors play a major role in aggression. In addition, alterations in human biochemistry also have an effect on aggression. But, it is also thought that the human environment influences aggressive tendencies. This essay will express both sides of the argument.

Genetic and hereditary factors play a major role in aggression because genes have been discovered that code for levels of testosterone; and testosterone influences aggression levels. For example, men generally have higher levels of testosterone than women and are also more aggressive than women (Knight 1996). Also, female prison inmates who displayed unprovoked violence also had very high levels of testosterone (Dabbs, 1998). Similarly, inbreeding and selective breeding illustrate the role of genes and hereditary factors in aggression because inbreeding can create unstable temperaments that are associated with aggressive tendencies and by selective breeding, aggressive and passive strains of mice can be created (Lagerspetz, 1979). Another reason why genetic and hereditary factors play a major role in aggression is because individual differences in temperament, e.g., reactivity and frustration responses, are inherited and observed in newborn babies. But, environmental stressors, such as exposure to alcohol in the womb, influence reactivity and frustration responses in newborns.

Another reason why it is believed that aggression is biologically caused is that alterations in human biochemistry also have an effect on aggression. High blood alcohol increases aggression because alcohol increases GABA in the cortex and reduces the capacity of the frontal cortex to inhibit midbrain aggressive impulses. Also, marital violence decreased in couples who completed behavioural marital therapy for alcoholism and remained sober during follow up (O'Farrell, 1995). Similarly, codeine is found in prescription strength cough and cold medications, and codeine can trigger aggression (Spiga, 1990). But, prescription strength cough and cold medications make some people very sleepy and lethargic, but not aggressive.

However, it is also thought that the human environment influences aggressive tendencies because parents, groups, entertainment and media all influence aggressive behaviour. Parents can influence levels of aggression in their children because experiments have demonstrated that children tend to imitate their parents' behaviour. For example, after watching parents beat a playroom doll with a stick, children often did the same (Bandura, 1977). Also, in extreme cases of child abuse, parents who beat their children often turn out to have been victims of child abuse themselves (Parke, 1975). Similarly, groups can influence aggression because when people feel less responsibility for their behaviour (e.g. in a mob setting), they are more likely to act aggressively. Also, group attitudes often polarize (e.g. become more extreme and aggressive) when individuals with similar attitudes come together.

Another reason it is thought that the human environment influences aggressive tendencies is because of entertainment and media influences. For example, watching sporting events can increase aggression in spectators. Watching pornography also increases aggressive male behaviour towards females. But, in Denmark, a study conducted for 10 years after the introduction of legal pornography in the country showed a gradual decrease in sexual assaults. It is also thought that television violence increases aggression in the viewer because a Canadian town with no television until the late 80's saw the incidence of playground aggression double within months (Williams, 1986). But, Feshbach & Singer (1971) found that being subjected to consistent violent television programmes over a six-month period didn't increase levels of aggression in teenagers.

2.4 Large Text: *Aggression is Biologically Caused*

It is often thought that aggression is biologically caused. This is because genetic and hereditary factors play a major role in aggression. In addition, alterations in human biochemistry also have an effect on aggression. Similarly, ethological studies and Freudian theory back up the notion that aggression is biologically caused. But, it is also thought that the human environment influences aggressive tendencies. This essay will express both sides of the argument.

Genetic and hereditary factors play a major role in aggression because genes have been discovered that code for levels of testosterone; and testosterone influences aggression levels. For example, men generally have higher levels of testosterone than women and are also more aggressive than women (Knight 1996). Also, female prison inmates who displayed unprovoked violence also had very high levels of testosterone (Dabbs, 1998). Similarly, inbreeding and selective breeding illustrate the role of genes and hereditary factors in aggression because inbreeding can create unstable temperaments that are associated with aggressive tendencies and by selective breeding, aggressive and passive strains of mice can be created (Lagerspetz, 1979). Another reason why genetic and hereditary factors play a major role in aggression is because individual differences in temperament, e.g., reactivity and frustration responses, are inherited and observed in newborn babies. But, environmental stressors, such as exposure to alcohol in the womb, influence reactivity and frustration responses in newborns.

Another reason why it is believed that aggression is biologically caused is that alterations in human biochemistry also have an effect on aggression. For example, people suffering from lead poisoning display increased levels of aggression. High blood alcohol also increases aggression because alcohol increases GABA in the cortex and reduces the capacity of the frontal cortex to inhibit midbrain aggressive impulses. Also, marital violence decreased in couples who completed behavioural marital therapy for alcoholism and remained sober during follow up (O'Farrell, 1995). Similarly, codeine is found in prescription strength cough and cold medications, and codeine can trigger aggression (Spiga, 1990). But, prescription strength cough and cold medications make some people very sleepy and lethargic, but not aggressive. Also, people who take cough and cold medication may act aggressively or angry because they are ill and not necessarily because of the medication.

Ethologists have shown that aggression in members of a species can aid survival and enhance fitness because when accompanied by rivalry amongst males for mating opportunities, aggression tends to perpetuate the genes of more vigorous animals. But, aggression doesn't enhance fitness in all contexts, e.g. male apes may use aggression to further their mating opportunities, but aggression doesn't necessarily help male humans in the same way.

Freud (1917) claimed that humans have a self-destructive urge known as the 'death instinct', which is redirected towards others in the form of aggression. But, Freud's argument that aggression is a redirected instinct is simply theoretical and lacks empirical evidence.

However, it is also thought that the human environment influences aggressive tendencies because parents, groups, entertainment and media all influence aggressive behaviour. Parents can influence levels of aggression in their children because experiments have demonstrated that children tend to imitate their parents' behaviour. For example, after watching parents beat a playroom doll with a stick, children often did the same (Bandura, 1977). Parents who do not use consistent disciplinary rules or help their children to self-monitor their behaviour and emotions, often have children who resort to aggressive behaviour in adolescence. Also, in extreme cases of child abuse, parents who beat their children often turn out to have been victims of child abuse themselves (Parke, 1975).

Similarly, groups can influence aggression because when people feel less responsibility for their behaviour (e.g. in a mob setting), they are more likely to act aggressively. People can be led to behave cruelly in order to conform to group norms, provided that a legitimate authority is willing to assume responsibility for their actions. Also, group attitudes often polarize (e.g. become more extreme and aggressive) when individuals with similar attitudes come together.

Another reason it is thought that the human environment influences aggressive tendencies is because of entertainment and media influences. For example, watching sporting events can increase aggression in spectators. Sports with high levels of body-contact are associated with increased aggression and spectators model their behaviour on the behaviours they observe (Bandura, 1989). Brazilian football supporters had increased testosterone levels while watching their team play in the 1994 world cup (Fielden, 1994), but, maybe the football supporters had high levels of testosterone naturally, regardless of watching football.

Watching pornography also increases aggressive male behaviour towards females, for example, the sales rate of pornographic magazines in different U.S. states is positively correlated with rape rates (Baron, 1984). But, in Denmark, a study conducted for 10 years after the introduction of legal pornography in the country showed a gradual decrease in sexual assaults.

It is also thought that television violence increases aggression in the viewer because a Canadian town with no television until the late 80's saw the incidence of playground aggression double within months (Williams, 1986). Longitudinal studies find long-term viewing of violence on television to be associated with an increase in males' violent behaviour as adults (Lefkowitz, 1977). Steur, Applefield & Smith (1971) found that young children who watched violent cartoons were more likely to kick, choke and push playmates than were those young children who watched the same cartoons, but with the violence removed. But, Feshbach & Singer (1971) found that being subjected to consistent violent television programmes over a six-month period didn't increase levels of aggression in teenagers.

Also, aggression and violence are more common in some cultures than others (Bellesiles, 1999). South Africa and the Philippines have significantly higher homicide rates than the U.S. (United Nations, 1999). Aggression and violence are triggered by famine, crowding and drought (Turnbull, 1972) and crime rates tend to be higher in countries with a considerable gap between the rich and the poor (Triandis, 1994).

2.5 Immediate Recall Test

Student ID _____

Gender _____

Please select the study material you used.

**30 BIT – Argument Map
1 Page Text**

**50 BIT – Argument Map
2 Page Text**

Complete each sentence, word for word, from what you remember from studying the material given to you, as best you can. Only one statement is necessary per ‘But’ or ‘Because’

1. Genes have been discovered that code for levels of testosterone; and testosterone influences aggression levels.

Because _____

Because _____

2. Codeine is found in prescription strength cough and cold medications and codeine can trigger aggression (Spiga, 1990).

But _____

3. Parents can influence levels of aggression in their children.

Because _____

Because _____

4. Aggression is biologically caused.

But _____

5. High blood alcohol increases aggression.

Because _____

Because _____

6. Television violence increases aggression in the viewer.

Because _____

But _____

7. Inbreeding and selective breeding illustrate the role of genes and hereditary factors in aggression.

Because _____

Because _____

2.6 Delayed Recall Test

Student ID _____

Gender _____

Please select the study material you used.

a) 30 BIT – Argument Map

b) 50 BIT – Argument Map

c) 1 Page Text

d) 2 Page Text

Complete each sentence, word for word, from what you remember from studying the material given to you, as best you can. Only one statement is necessary per ‘But’ or ‘Because’

1. Aggression is biologically caused.

But _____

2. Genes have been discovered that code for levels of testosterone; and testosterone influences aggression levels.

Because _____

Because _____

3. High blood alcohol increases aggression.

Because _____

Because _____

4. Codeine is found in prescription strength cough and cold medications and codeine can trigger aggression (Spiga, 1990).

But _____

5. Inbreeding and selective breeding illustrate the role of genes and hereditary factors in aggression.

Because _____

Because _____

6. Parents can influence levels of aggression in their children.

Because _____

Because _____

7. Television violence increases aggression in the viewer.

Because _____

But _____

3.1 Text for Active Learning: *Gun Control*

There was nothing wrong with the United States Supreme Court's decision to uphold the individual's right to bear arms because the second amendment in the U.S. constitution provides that "the right of the people to keep and bear arms shall not be infringed." But, the right to bear arms can be taken away lawfully if that right interferes with public safety (Quilici v. Village of Morton Grove, U.S. Court of Appeals, Seventh Circuit, 1982). Similarly, the definition of a firearm, or arms, is unclear, (i.e. is it a handgun, a rifle, or even a bazooka), thus making it difficult to decipher what is legal. Also, laws in America apply to America only and do not necessarily convey the views of other countries in the rest of the world.

It was right to uphold the individual's right to bear arms because gun control could prohibit hunting, a traditional, recreational activity. But, you do not need a gun to hunt. Guns can be used in self-defense because women need guns as 'the great equaliser', as self-defense against attackers, i.e. (physically) larger and stronger men. In addition, keeping a gun in the home is a sound method of self-defense against intruders. because victims who defended themselves with guns were less likely to report being injured than those who either defended themselves by other means or took no self-protective measures at all (Roth, 1994). But, Zimring (1991) reported that the use of a firearm to resist a violent assault actually increases the victim's risk of injury and death. Also, research has shown that a gun kept in the home is 43 times more likely to kill a member of the household, or friend, than an intruder (Kellermann & Reay, 1986).

The National Rifle Association (of America) argues that if law-abiding citizens have guns, they are safer from criminal attacks, in that they bring crime rates down. But, in countries where guns are greatly restricted, such as Ireland, Great Britain or Japan, deaths from guns are very rare, especially compared to the United States. Guns allow collective self-defense - if many Americans own guns, it will be easier to defend against an invading foreign power. But, given the power of the American military, it is absurd to believe that ground invasion of the U.S. could happen.

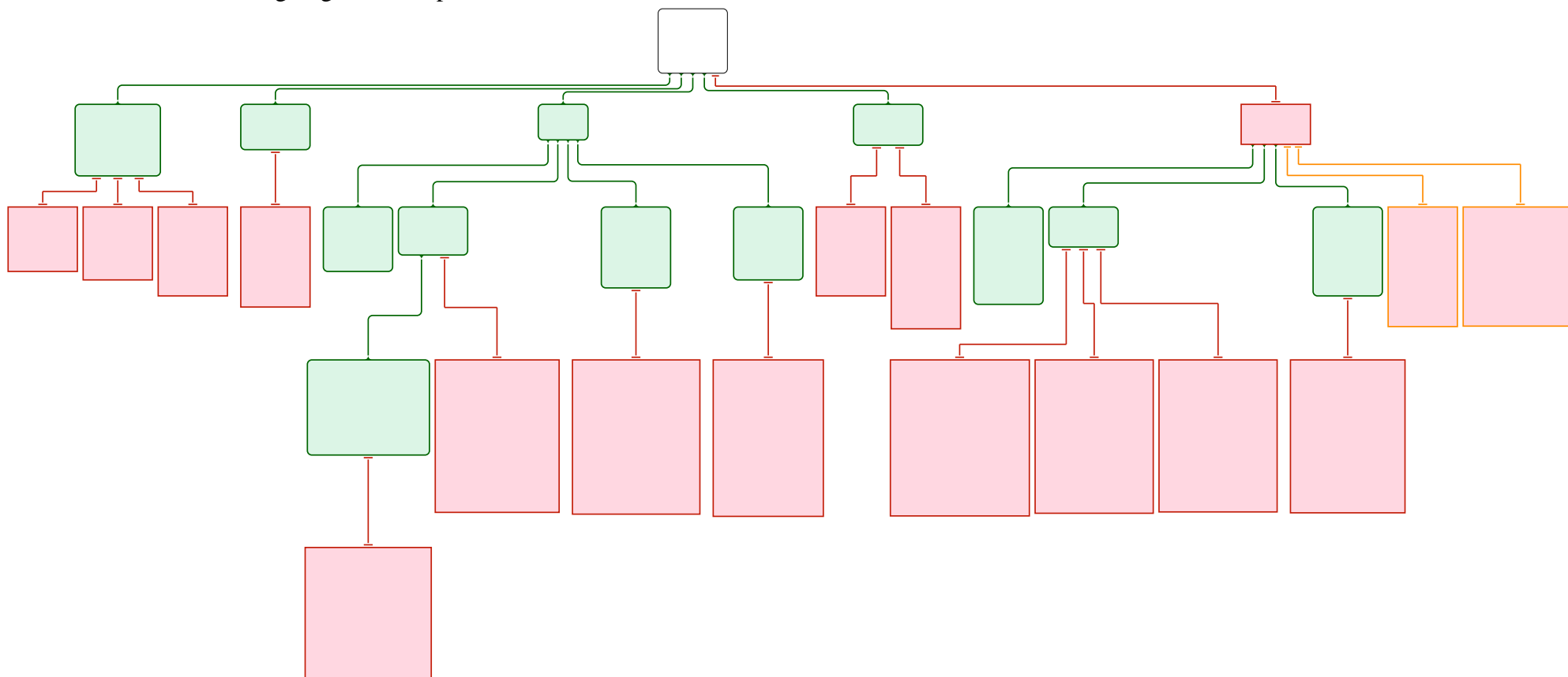
It's often believed that, "when guns are outlawed, only outlaws will have guns". But, one main reason why there are so many guns on the street in the hands of criminals is precisely because so many are sold legally and are often loaned or given to criminals. Also, FBI Crime Reports sources indicate that there are about 340,000 reported firearms thefts every year- those guns, the overwhelming amount of which were originally manufactured and purchased legally, are now in the hands of criminals.

On the other hand, the Supreme Court's decision was wrong as guns are a main cause for many unnecessary deaths. In robberies and assaults, victims are far more likely to die when the perpetrator is armed with a gun than when he or she has another weapon or is unarmed (Roth, 1994). Also, thousands of children die annually in gun accidents. But, the often repeated claim that 12 children per day die from gun violence includes "children" up to 20 years of age, the great majority of whom are young adult males who die in gang-related violence. Similarly, gun accidents involving children are actually at record lows and more children die each year in accidents involving bikes, space heaters or drownings.

Residents of homes where a gun is present are 5 times more likely to experience a suicide than residents of homes without guns (Kellermann et al., 1992). But, both the U.S. and Ireland still fall into the same suicide rate category (6.5 -13 per 100,000 people), even though substantially less Irish own guns than Americans, with regards to both rate and percentage.

However, states that allow registered citizens to carry concealed weapons have lower crime rates than those that don't (Lampo, 2000). In addition, there is no correlation between gun control laws and deaths by gun across a spectrum of nations and cultures. For example, Israel and Switzerland allow gun licenses, widespread carrying of concealed firearms, and yet, they have rates of homicide that are much lower than those of the U.S., despite rates of home firearm ownership that are at least as high (Kellermann, 1993).

3.2 Active Learning Argument Map



3.3 Active Learning Outline

There was nothing wrong with the United States Supreme Court's decision to uphold the individual's right to bear arms.

I. The second amendment in the U.S. constitution provides that "the right of the people to keep and bear arms shall not be infringed."

A. But, laws in America apply to America only and do not necessarily convey the views of other countries in the rest of the world.

B. But, the definition of a firearm, or arms, is unclear, (i.e. is it a handgun, a rifle, or even a bazooka), thus making it difficult to decipher what is legal.

C. But, the right to bear arms can be taken away lawfully if that right interferes with public safety (Quilici v. Village of Morton Grove, U.S. Court of Appeals, Seventh Circuit, 1982).

II. Gun control could prohibit hunting, a traditional, recreational activity.

A. But,

III. Guns can be used in self-defense.

A. Women need guns as 'the great equaliser', as self-defense against attackers, i.e. (physically) larger and stronger men.

B. Keeping a gun in the home is a sound method of self-defense against intruders.

1. Because

a. But,

2. But,

C. The National Rifle Association (of America) argue that if law-abiding citizens have guns, they are safer from criminal attacks, in that they bring crime rates down.

1. But,

D. Guns allow collective self-defense - if many Americans own guns, it will be easier to defend against an invading foreign power.

1. But,

IV. "When guns are outlawed, only outlaws will have guns".

A. But, one main reason why there are so many guns on the street in the hands of criminals is precisely because so many are sold legally and are often loaned or given to criminals.

B. But, FBI Crime Reports sources indicate that there are about 340,000 reported firearms thefts every year- those guns, the overwhelming amount of which were originally manufactured and purchased legally, are now in the hands of criminals.

V. But, guns are a main cause of many unnecessary deaths.

A. In robberies and assaults, victims are far more likely to die when the perpetrator is armed with a gun than when he or she has another weapon or is unarmed (Roth, 1994).

B. Thousands of children die annually in gun accidents.

1. But,

2. But,

3. But,

C. Residents of homes where a gun is present are 5 times more likely to experience a suicide than residents of homes without guns (Kellermann et al., 1992).

1. But,

D. However,

E. However,

3.4 Recall Test

Student ID _____

Name _____

Please select the study material you used.

Summary

Outline

Argument Map

Complete each sentence, word for word, from what you remember from studying the material given to you, as best you can.

1. Victims who defended themselves with guns were less likely to report being injured than those who either defended themselves by other means or took no self-protective measures at all (Roth, 1994).

But _____

2. Thousands of children die annually in gun accidents.

But _____

But _____

But _____

3. Guns are a main cause for many unnecessary deaths.

But _____

But _____

4. Keeping a gun in the home is a sound method of self-defense against intruders.

But _____

Because _____

5. Gun control could prohibit hunting, a traditional, recreational activity.

But _____

6. The National Rifle Association (of America) argue that if law-abiding citizens have guns, they are safer from criminals, in that they bring crime rates down.

But _____

7. Residents of homes where a gun is present are 5 times more likely to experience a suicide than residents of homes without guns (Kellermann et al., 1992).

But _____

8. By collective self-defense, if many Americans own guns, it is better for the general welfare of the U.S in case they are invaded by a foreign power.

But _____

APPENDIX C

Participant Consent Form & Study Information

PARTICIPANT CONSENT FORM

It is a necessary and routine practice to gain informed consent from all participants taking part in research. As a research participant, your rights are as follows:

- To be informed of the general nature of the research;
- To participate in research in a strictly voluntary manner;
- Entitlement to withdraw from the study at any time;
- To be assured that data collected from the study will be kept strictly confidential;
- To be protected from any physical or psychological harm;
- Not to be deceived in anyway that can be harmful or unnecessary;
- To be debriefed via e-mail or web-site post, where the core findings will be summarised.

Researcher

I, Chris Dwyer (researcher) agree to abide by all the guidelines and standards for conducting research with human participants as described by the Psychological Society of Ireland and other professional psychological associations.

Date:

Participant

I, _____ have been informed about the general nature of this study and agree to participate voluntarily. I have read and understand my rights as a participant; and I understand that they will be guaranteed to me.

Date:

STUDY INFORMATION FOR PARTICIPANTS

The current study examines the role of argument mapping as a learning aid. Argument mapping is a strategy in which the core propositions and relations of an argument are visually represented, thus aiding the understanding and retention of the argument.

Argument mapping has a long history, but was tedious work before improvements in computer technology allowed for the construction of maps with ease. However, advances in computer technology have not been paralleled by advances in empirical work on the cognitive benefits of reading and constructing argument maps. The current study seeks to compare the argument map design, as a learning format, with the traditional text-based format used in most common learning environments, in terms of their relative effectiveness with respect to recall, comprehension and critical thinking. Within the study, verbal and spatial reasoning, disposition towards thinking, motivation, comprehension and memory are considered and tested, as is critical thinking and reflective judgment. This work is supervised by Dr. Michael Hogan and Dr. Ian Stewart.

APPENDIX D

Customised Student Questionnaire, Exercise Handouts and DVD Recordings of Lectures & Exercises Used in Study 2

Customised Student Questionnaire

Student ID _____

Please answer each question in full.

1. If you signed up to our Critical Thinking Seminar Series in first year, but did not finish it, or did not take it at all, can you please describe the reason(s) for this?

2. If you were to suggest ways we could improve our critical thinking recordings/materials for the purpose of delivering them in future courses, what suggestions would you make? (If you were in the control group please describe optimal ways of teaching critical thinking).

3. Can you suggest novel critical thinking exercises that might be used to maintain interest and motivation in a critical thinking intervention study?

4. Can you suggest incentives or rewards we might use to keep students engaged in a critical thinking intervention study?

5. If you have any additional comments or suggestions to improve the lecture series on critical thinking for future critical thinking courses, or had any problems, please elaborate below.

If you completed the Critical Thinking Seminar Series (i.e. you completed both the pre-tests and post-tests and attended the seminars), please complete the following 10 questions.

On a scale from 1 to 7 (1 being very good; and 7 being poor), please rate the following to the side of each sentence:

1. The critical thinking course throughout.

1 2 3 4 5 6 7

2. Your ability to understand the course throughout.

1 2 3 4 5 6 7

3. The material(s) presented within the course.

1 2 3 4 5 6 7

4. The quality of the class exercises completed within the course.

1 2 3 4 5 6 7

5. The ability of the presenter throughout the course.

1 2 3 4 5 6 7

6. The quality of the presentations throughout the course.

1 2 3 4 5 6 7

7. Your overall motivation to learn during the critical thinking course.

1 2 3 4 5 6 7

8. Your overall satisfaction with your learning throughout the critical thinking course.

1 2 3 4 5 6 7

9. The prospect that you will be able to apply your learning elsewhere.

1 2 3 4 5 6 7

Exercise Handouts

Handout 1:

Hypnosis improves memory

A popular view is that individuals have extraordinary abilities while under hypnosis, such as supernormal memory. When hypnosis helps a person to recall information that he or she did not previously recall, this is termed 'hypnotic hypermnesia'. Many clinicians and their clients also believe that hypnosis can improve memory (e.g. Wolberg, 1982).

The notion that hypnosis can improve recall has received support from documented cases in which law enforcement agents have obtained useful evidence from eyewitnesses who, prior to hypnosis, had difficulty recalling details of the crimes they witnessed. One example of the success of hypnosis in refreshing the memory of a crime comes from the case of *People vs. Woods et al.* (cited in Smith, 1983), better known as the Chowchilla kidnapping case. In Chowchilla, California, kidnappers abducted 26 children and their bus driver and secluded them in an underground, rectangular enclosure. The bus driver and two of the older children were able to escape by digging through to the surface and then contacted the police. Although the bus driver had tried to memorize the license plate numbers of the abductors' vehicles at the time of the abduction, he was unable to recall the plate numbers to police. The police gave him hypnotic instructions that he should imagine sitting in his favorite easy chair to watch the crime unfold on television. He then recalled two license plate numbers, one of which turned out to be off by only one number. His testimony enabled police to speed up their investigation and apprehend the kidnappers. In a review of the literature, Smith (1983) cited other cases in which hypnosis has later improved memory of events and crimes that were observed under un hypnotized (awake) conditions but were not well remembered prior to hypnosis.

Hypnosis has been used since the time of Freud as a method for helping people with psychological problems to recall traumatic or anxiety-provoking events from their pasts, especially from their childhood. Similarly, it has long been thought that through hypnosis subjects could regress to an earlier age when given a suggestion to imagine that they were back at an earlier time. In many examples from stage and laboratory, hypnotized adults who are given suggestions to return to an earlier age have shown changes in their speech and facial expressions appropriate to that younger age. When researchers have checked such people's memories of events from previous times, they found that age-regressed subjects did not accurately remember details (Nash, Drake, Wiley, Khalsa, & Lynn, 1986). In a review of the literature on hypnotic age regression, Nash (1987) concluded that there was little evidence that hypnotically age regressed subjects actually returned to a mental state like their original state or that their recall improved. In fact, through hypnotic age regression, subjects may recall incidents, even former lives, that they actually may have constructed based on other experiences. Spanos, Menary, Gabora, Dubreuil, and Dewhirst (1991) found that subjects who were hypnotized, age-regressed, and given suggestions to recall former lives often did so; however the lives they recalled were frequently based on historical figures with whom they were familiar. Nevertheless, the subjects often recalled these lives inaccurately.

Some early research showed that hypnosis did enhance memory for information from a person's past. According to a literature review by Relinger (1984),

however, later psychological laboratory experiments concerning people's memory for nonsense syllables and other items of low meaningfulness failed to show any benefits from hypnosis.

After some of these successes in refreshing memory with hypnosis, police departments began to use hypnosis as an investigative tool in the 1960s. One result was a renewed interest among psychologists in researching hypnotic hypermnesia. In order to explain inconsistent findings concerning hypnosis and recall, psychologists began to investigate conditions that might produce hypnotic hypermnesia. Relinger (1984), in his review, argued that hypnotic hypermnesia tended to occur in studies of memory for meaningful material but not for material of low meaningfulness. A review by Erdelyi (1994) also observed this trend; however an experiment by Mingay (1985) found that hypnotized subjects were not more accurate than unhypnotized subjects in recalling meaningful visual material.

Other studies tried to make the laboratory situation more like the setting of a crime and its investigation. Because crimes are often witnessed during states of heightened arousal, subjects' arousal was heightened to make the experiment more like the natural situation. Subjects were sometimes also deceived into believing that actual crimes had been committed. In general, these experiments have not resulted in subjects' showing better memory than non-hypnotized subjects, according to Smith's (1983) review of these studies.

Other research has shown that what looks like a hypermnesic effect in fact may not be due to hypnosis. Erdelyi and Becker (1974) showed that with repeated attempts at recall, even without hypnosis, subjects remember things that they did not previously recall. Because many early studies examined memory initially without hypnosis and then under hypnosis, improved recall attributed to hypnosis actually may have been due simply to repeated attempts at recall. Repeated attempts at recall thus may explain situations in which a witness first unsuccessfully attempts to recall details of a crime and then is later able to recall the details after being hypnotized. In a recent review of the literature, Erdelyi (1994) concluded that the repeated attempts to recall initially unrecalled information account for any increase in recall after hypnosis, rather than hypnosis per se being responsible.

Some well-controlled studies, however, have shown that hypnosis improves recall. For example, Dhanens and Lundy (1975) conducted an experiment in which they compared different groups of subjects who were all highly susceptible to hypnosis. These different treatment groups were tested under various memory testing conditions on their ability to recall a prose passage immediately after learning it and then a week later. Unhypnotized subjects were asked to recall the prose passage after receiving no additional instructions, after instructions to relax, after hypnotic regression instructions, or after motivational instructions. Still other treatment groups were hypnotized and then asked to recall the prose passage after receiving either the same regression or the same motivational instructions as their unhypnotized counterparts. In the most important results of the experiment, the researchers found that the hypnotized groups, who also got motivational instructions, showed the most gain in their recall of the passage, suggesting a hypnotic hypermnesia effect.

Two other problems have been found with the use of hypnosis to refresh memory. First, some researchers, like Laurence and Perry (1983), have shown experimentally that a hypnotized person may be given a false suggestion while trying to recall some event and will incorporate this false suggestion into memory, thus creating a pseudomemory. Then, outside of hypnosis, these subjects will recall these pseudomemories as if they were true. Sheehan, Statham, and Jamieson (1991) created

pseudomemories about events viewed in videotapes and found that these events persisted at least two weeks after the viewing. It is not clear what caused the pseudomemories. Spanos and McLean (1985) produced pseudomemories in all 11 highly hypnotizable subjects they tested, but under special reporting conditions almost all subjects acknowledged the memories were imagined; these results suggested that the failure to recall was due to bias from the reporting procedure, not actual distortion of the memories. According to a literature review by Spanos, Burgess, and Burgess (1994), people who form pseudomemories of being abducted by aliens often do so within the context of hypnotic and structured interviews that create demands for the reporting of such experiences, which people then come to interpret as actual memories of abduction.

The formation of pseudomemories can create problems for those using hypnosis to obtain information. Coons (1988) studied the case of a woman interviewed by police who used a particularly suggestive interrogative technique. This led to the woman's confessing to a crime she did not commit and developing symptoms of multiple personality disorder created by the suggestions.

A second problem with using hypnosis to help recall memories is hypnotized subjects' overconfidence in their memories. Although subjects in the Sheehan et al. (1991) study who recalled pseudomemories were not more confident of their memories than were those who did not recall pseudomemories, subjects in other studies often have been more confident of pseudomemories formed during hypnosis. For a review, see the Council on Scientific Affairs, American Medical Association (1985).

In summary, hypnosis has sometimes been found to be effective in helping police to solve crimes by helping witnesses to recall more information. Although hypnotically age-regressed individuals appear to return to an earlier time, their recall of details from that time has been found to be inaccurate. Studies that have shown hypnosis to improve recall have examined the recall of meaningful information. Most studies seeking to simulate realistic natural conditions have failed to show hypnotic hypermnesia. Other studies have suggested that improvement in recall may have been due to repeated attempts at recall rather than to hypnosis per se. Finally, false memories have been created within the context of hypnosis; moreover subjects often report that they are confident of the accuracy of false or inaccurate memories. In an extensive review of the literature, a panel of experts on hypnosis and memory concluded that hypnosis does not improve recall of meaningless information and when hypnosis does affect recall of meaningful information it may increase the recall of both accurate and inaccurate information (Council on Scientific Affairs, American Medical Association, 1985).

Handout 2 (adapted from Cottrell, 2005):

'Stealing is always wrong.' Discuss with reference to unpaid downloading of music from the internet.

There are many different forms of stealing, from theft of property, muggings and burglaries, to theft of ideas through plagiarism. Although there are legal sanctions against many forms of stealing, the issue of moral and social sanctions has always been more complex. For example, Robin Hood, who stole from the rich to give to the poor, is held up as a great British hero. Piaskin (1986) suggests that ethical issues are not simply questions of right and wrong but should be regarded as 'dilemmas'. In this essay I shall use the example of downloading music from the internet to highlight these complexities but, contrary to the view held by Piaskin, to argue that in this case, stealing is always wrong.

In recent years, there have been a number of high profile cases against people who have shared music files for free on the internet. Prior to the development of the internet, music was similarly shared via home taping. Lee (2006) argues that although home taping is technically illegal, no one pursues this as perpetrators cannot be caught. Because it is possible to catch internet file sharers, Lee argues that they are being unfairly punished. Whilst there may be a practical basis to this argument – it is easier to catch downloaders than home tapers – this does not mean that one behaviour should be considered acceptable and the other should not. This kind of argument is a rationalisation, used to make unacceptable actions appear acceptable.

Indeed, this point is made by Cuttle (2007). Cuttle, a legal expert, states that 'piracy of software, video games and music is stealing' and makes it clear that all such copying is illegal. Given that there is a legal argument against both home taping and internet downloading, it appears reasonable to assume that both should be considered as wrong. However, it is important to explore the moral arguments in order to evaluate whether such behaviours should also be considered 'wrong' from an ethical perspective.

Research by Mixim, Moss and Plummer (1934), as well as later studies inspired by Mixim et al., suggest that most people do maintain an ethical sense of right and wrong even in areas where stealing appears to be more socially acceptable. Their findings suggested that people's ethical sense wanes when payment methods are difficult but they do not forget what is ethically right. Ebo, Markham and Malik (2004) examined the effect on internet downloading of easier payment schemes. During the study there was a dramatic decrease in illegal downloads with the majority of users choosing to make use of the easy payment scheme. This indicates that the majority of people in the study acknowledged that to download music for free, in effect stealing it, was wrong.

A different ethical perspective is suggested by those authors who support unpaid downloading, especially those who use ethical and artistic arguments to counter economic arguments. A number of authors such as 'Carla' (2006), an internet downloader, assert that the main argument against downloading comes from record companies who are primarily concerned with their own profits. Economic arguments are treated by such writers as if they are intrinsically weaker than artistic ones. 'Carla' develops this argument to suggest that true artists are driven by a desire to have their music heard by others and welcome the 'service' provided by file sharers. Hibbs (2006), a member of the public, also argues that file sharing is a kindness between friends. These kinds of arguments can sound convincing as they make downloading appear to be altruistic, and altruism appears to have the ethical advantage over the rush for profits. On the other hand, it could be argued that this is altruism at someone else's expense. The economics of free downloading do not help less well known artists, so not paying for downloads of their work is unethical.

Furthermore, those who defend downloading often act as if they know best the 'real' wishes and interests of artists. Carla, for example, refers to 'true artists', without defining what a 'true artist' is,

or providing evidence to show what such 'true' artists would want. Authors such as 'Carla' and Hibbs do not provide evidence to show that artists regard free downloading as being more in their interests than the actions taken by businesses. As music sales are usually of direct financial benefit to artists, many artists may also disagree with free downloading.

Moreover, Cuttle (2007) asserts that arguments such as Carla's and Hibbs's are invalid in free market terms. Publishers have a right to charge the highest price that they are able to obtain, and consumers can choose whether or not to purchase. In that case, business is not in the wrong to charge whatever price the market will sustain. However, there are other economic, and indeed artistic, arguments against Carla's and Hibbs's positions. Such authors assume that objections to downloading come mainly from large corporations who can be dismissed as 'greedy'. Kahliney (2006) argues that small, independent companies and recording artists are most likely to suffer the effects of downloading as their overall reliance on sales is greater. Given that sales for independent artists tend to be low anyway, falling sales could mean the collapse of small labels. Whilst artists could still have their music heard via free downloads, their position is unlikely to remain financially viable for long. Ironically, this increases the likelihood of a music industry populated by the type of 'bland' or 'middle of the road' acts that Carla complains would exist without internet downloading: they will be the only artists that can guarantee reasonable sales.

In conclusion, I have demonstrated in this essay that there are arguments to support the view that all stealing can be regarded as 'wrong'. This holds true even in relation to complex areas such as internet downloading, where social behaviours may appear to support the view that downloading without paying is acceptable. Indeed, in the case of unpaid downloading, there are legal and ethical, economic and artistic arguments to support the view that stealing from the industry is wrong. There are counter arguments, such as that downloading offers a service to music and small artists, but there is little evidence to support such views or to suggest that they represent the view of the majority. On the contrary, when given accessible, affordable payment options, most people chose not to steal, thereby acknowledging that free downloading is wrong. Although moral positions can easily be influenced by practical circumstances such as how easy it is to pay, research suggests people maintain an ethical sense that stealing is always wrong.

DVD Lecture Recordings

See Hardbound Version

APPENDIX E

e-Learning Lectures, Exercises and an Example of Feedback Provided to Participants in Study 3

e-Learning Lectures & Exercises

All e-learning lectures and exercises can be found at:

www.nuigalway.ie/psy/sub/CDwyer_CT_Course

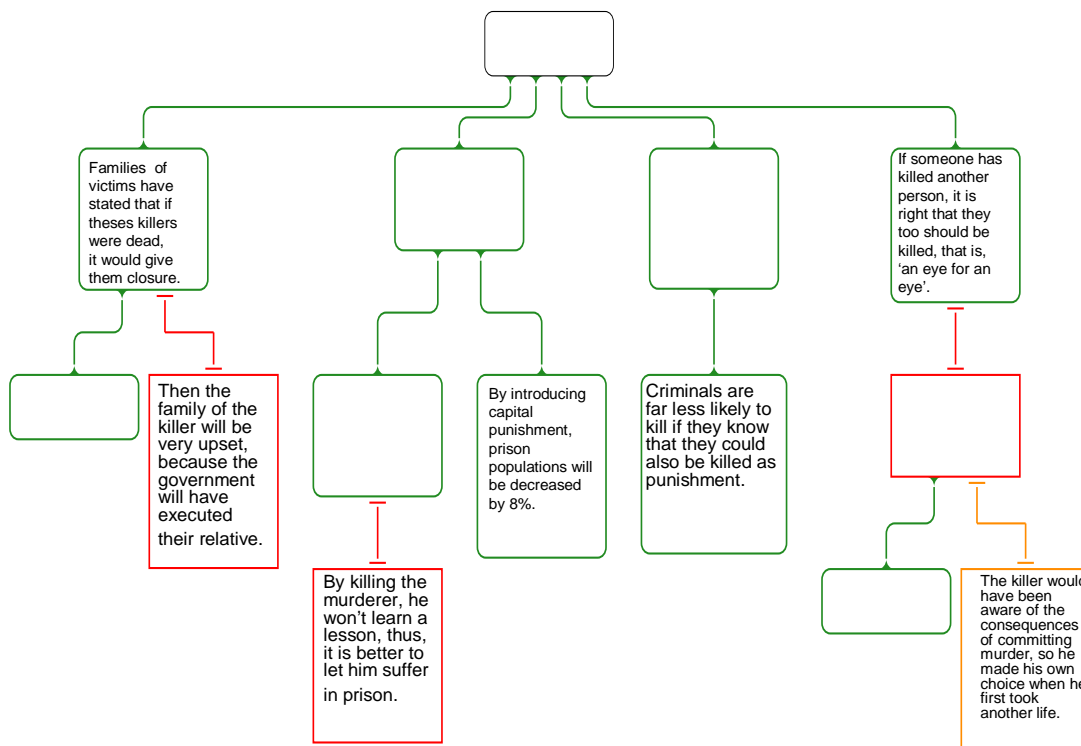
Example Feedback

Feedback from Lecture 3.1:

Thank you very much to all of you who did the exercises. Below you will find some feedback on exercises from Lecture 3.1. In the future, I'd be very grateful if you could save your PowerPoint file or Rationale file to your computer and then attach that file in an email to me. This is the easiest way for me to view your work. Thanks again!

Exercise: "Ireland should adopt Capital Punishment"

The answers for this exercise are in the argument map below. Please compare and contrast your argument map with this one. Also, please take note of where you may have differed in your placement of some propositions and why you decided on that location.



You were also asked to answer a few questions based on this argument map.

Question 1 asked:

Does the author sufficiently support their claims? Are the author's claims relevant? Does the author attempt to refute their own arguments (i.e., disconfirm their belief)?

Some of you answered that:

The author did not sufficiently support his/her claims because most of the propositions used were based on personal opinion (i.e. there was an insufficient amount of evidence to suggest that Ireland should adopt capital punishment). The

author did attempt to refute their own claims, however, did so poorly in that he/or she only used personal opinion or ‘common belief’ statements.

The author sufficiently supported their claims as they were sufficiently backed up by other supports. These claims are all relevant to the argument, specifically the central claim. The author does not attempt to disconfirm his beliefs because he sticks to his guns that capital punishment should be adopted.

The truth of the matter is that the author *did not sufficiently* support his/her claims. Of the 8 reasons he provided, only 3 were based on either expert opinion, statistics from research or research data.

All the arguments made were relevant to the central claim.

The author *did attempt* to refute his/her claims (i.e. disconfirmed their own belief), as on 3 occasions, some form of objection to the reasoning was presented. However, the objections used were not of the highest quality.

Question 2 asked:

Are there other arguments you would include?

Some of you answered that:

Some argument should be made in terms of when the death penalty should/would be used, such as in cases of mental problems or a conviction of manslaughter.

Some argument should consider the nature of the crime, such as how the murder took place –details should be considered.

Everyone has a right to life, even murderers.

Law abiding citizens might grow to fear the government as they would have now have more control over you.

Please think about these ideas and claims and also think about how you could possibly integrate them into the argument map. In addition, think about how you might support or object to these new propositions.

Question 3 asked:

Does any proposition or any set of propositions suggest to you that the author is biased in any way?

Some of you answered:

The author is biased because he/she presents more reasons for why we should adopt capital punishment than for not adopting capital punishment.

The author was not biased because though he/she did present more reasons in favour of capital punishment, they were mostly based on personal opinion and were adequately objected to.

The correct answer is that the author is biased. However, some of you argued that it is because the author stated that 'Ireland should adopt capital punishment', thus making it a biased argument. This is not true, because the author may have made the same claim and then simply presented 5 objections across the top line of reasoning (as opposed to 4 supports). Remember, there is more to determining bias than simply assimilating what the central claim is; what is more important is how the author attempts to justify or refute this claim. The reason why this argument is biased is because the author only presents credible evidence (in 3 cases) to support the claim. In the cases where the author makes a claim and objects to it, both the reasons and objections are based on personal or common belief. This is done to disguise the author's bias. In the cases where the author presents credible evidence, there are no objections.

APPENDIX F

Programme of Questions used to Guide the Focus Group Interview in Study 3

- 1 Could you tell me a little bit about the quality of the course, specifically the course materials, exercises and recordings?
- 2 Please tell me a little bit about the exercises and your reaction to those exercises with regards to the training of analysis, evaluation and inference skills.
- 3 Do you think that the exercises helped you achieve the aim of applying analysis, evaluation and inference?
- 4 Do you think the exercises would have been better if you received one per session, or were the two exercises per session necessary?
- 5 Please tell me about the quality of the feedback you received on the exercises as well as your reaction to the feedback.
- 6 Please tell me about your overall learning experience and try to describe the nature of this experience, particularly with regards to learning about critical thinking and argument mapping.
- 7 Please tell me about your experiences of thinking about thinking in this course and about whether or not it was a slow, fast or gradual process.
- 8 Over the duration of the course, from beginning to end, how do you feel you changed in the way you think about thinking?
- 9 Did you find argument mapping affected your ability to think critically?
- 10 If you were given the same course, but without argument mapping, do you think you would have gotten the same amount out of it, or maybe more out of it? Less? Please tell me what you think.
- 11 Can you tell me something about the ways in which you have applied your learning from the course since its completion?

- 12 How do you see yourself applying critical thinking and argument mapping skills in the future?
- 13 If you were to suggest ways in which to improve the critical thinking course materials for the purposes of delivering them again, what specific suggestions would you make?
- 14 Can you suggest any novel critical thinking exercises that might maintain interest and motivation in a future version of this course?
- 15 Can you suggest incentives or rewards that might be used to keep students engaged in an online critical thinking course?