

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

# **Automating the Measurement of Critical Thinking in Discussion Forums**

A thesis presented in partial fulfilment of the requirements for the  
degree of  
Doctor of Philosophy (PhD)  
in Information Systems  
at Massey University, Palmerston North,  
New Zealand

Stephen Paul Corich

2011



# Abstract

---

Critical thinking is seen as an essential skill for survival in the 21st century and those involved in education agree that encouraging students to think critically is an essential requirement of the educational system. There is considerable debate among educationalists concerning the most effective way to encourage the development of critical thinking skills and little agreement on the best way to measure evidence of critical thinking.

This study examines what is meant by the term critical thinking and investigates some of the tools that have been used to measure evidence of critical thinking. The study concentrates on the measurement of critical thinking among discussion forum participants and describes the development and testing of a tool designed to automate the process of measuring critical thinking.

Since the study involved the design, development and testing of a computerised tool, an action research approach was adopted. Action research gained popularity among information systems developers around the mid 1990s. It is seen as a methodology that encourages the generation of scientific knowledge by allowing the researcher to modify intentionally a real setting and carefully evaluate the results.

The study describes the development of an automated text classification system that uses content analysis to classify discussion forum transcripts using critical thinking models developed by Garrison, Anderson, & Archer (2000), Henri (1992) and Perkins & Murphy (2006). The findings from the study suggest that automated tools have the potential to assist in the measurement of critical thinking abilities and are worthy of further investigation.



# Acknowledgements

---

Arriving at a point where I can present this thesis has taken me on a journey of discovery that would have been impossible without the encouragement and support of a number of talented and caring people. Working full-time and studying part-time takes its toll and I would not have persevered had I undertaken the journey alone.

I would like to acknowledge the support of supervisors, colleagues and family who have acted as mentors, reviewers, encouragers and supporters. Thank you all so much.

To Kinshuk, my primary supervisor, thank you for taking me on as a student and agreeing to continue with my supervision even when you left Massey University and moved to Canada to join Athabasca University. To Lynn Jeffery, my second supervisor, thank you for your willingness to review my work, provide constructive feedback and support me during and after my defence. Thanks also to my fellow students who formed the core group of the Massey University's Advanced Learning Technology Research Centre (ALTRC), your feedback during the research symposiums and the online meetings helped me shape the direction of my research.

Special thanks to my employers at the Eastern Institute of Technology in Hawke's Bay, I would have given up long ago without your encouragement and the time release you forced upon me to allow me to complete the write-up. Thanks also for the financial support which allowed me to attend international conferences and present my work for peer review. To the wonderful team in the School of Computing, you were all instrumental in keeping me on track during the highs and lows. To the students in the School of Computing who were the subjects of my research, who agreed to participate in the study and were interested in the result, thanks. I have been very fortunate to work in such a supportive environment and I appreciate the support

of all the staff and students who have travelled with me over the last seven years of my educational adventure.

Finally and perhaps most importantly I would like to acknowledge my family. To my stepmother, Penny, thank you for helping me appreciate the value of a good education. To Shirley, my loving and supportive wife; you reviewed my chapters; you encouraged me when I wanted to give up; and you stayed the distance, thank you so much. To Matthew, Abbey and Joshua, you have all made me so proud and you have always supported my endeavours, thank you for your encouragement.

## Ethics Committee Approval Statement

Ethical approval for this project was obtained from the Massey University Human Ethics Committee and the Eastern Institute of Technology, Hawke's Bay Research and Ethics Committee.

A copy of the Massey University Human Ethics Committee approval document is provided in Appendix A.

A copy of the Eastern Institute of Technology, Hawke's Bay Research and Ethics Committee approval document is provided in Appendix B.

# List of Publications

---

## Book Chapters

- Corich, S., Kinshuk., & Jeffrey, L. M. (2007). The use of discussion forums in learning communities. In R. Luppigini (Ed.), *Learning communities in online education* (pp. 87 - 108). Charlotte, NC: Information Age Publishing.
- Corich, S., Kinshuk., & Jeffrey, L. M. (2010). Automating the measurement of critical thinking for individuals participating in discussion forums. In D. Ifenthaler, P. Isaias, J. M. Spector, Kinshuk, & D. G. Sampson (Eds.), *Multiple perspectives on problem solving in the digital age* (pp. 143 - 158). New York: Springer.

## Refereed Journals

- Corich, S., Kinshuk., & Hunt, L. M. (2004). Assessing discussion forum participation: In search of quality. *International Journal of Instructional Technology and Distance Learning*, 1(12), 3 -12.
- Corich, S., Kinshuk., & Hunt, L. M. (2006). Computerised content analysis for measuring critical thinking within discussion forums. *Journal of e-Learning and Knowledge Society*, 2(1), 47 - 60.

## Refereed Conference Proceedings (Paper presentations)

- Corich, S. (2009). Using an automated tool to measure evidence of critical thinking of individuals in discussion forums. In S. Mann & M. Verhaart (Eds.) *Proceedings of the 22nd Annual Conference of the National Advisory Committee on Computing Qualifications*, pp. 59 - 64, July 4 - 8. Auckland, New Zealand. Retrieved from <http://hyperdisc.unitec.ac.nz/naccq09/proceedings/pdfs/15-21.pdf>.
- Corich, S., Kinshuk., & Hunt, L. M. (2004). *Using discussion forums to support collaboration*, Paper presented at the Third Pan Commonwealth Forum on Open Learning, 4 - 8 July. Dunedin, New Zealand. Retrieved from [http://www.col.org/pcf3/Papers/PDFs/Corich\\_SStephen.pdf](http://www.col.org/pcf3/Papers/PDFs/Corich_SStephen.pdf)

- Corich, S., Kinshuk., & Hunt, L. M . (2006). *Measuring critical thinking within discussion forums using a computerised content analysis tool*. Proceedings of Fifth International Conference on Networked Learning 2006, 5 - 7 May. Lancaster University, England. Retrieved from <http://www.networkedlearningconference.org.uk/past/nlc2006/abstracts/pdfs/P07%20Corich.pdf>
- Corich, S., Kinshuk., & Jeffrey, L. M. (2007). *Changing focus from group to individual: Using an automated tool to measure evidence of critical thinking in discussion forums*. Paper presented at the IADIS International Conference CELDA, December 7 – 9. Portugal. Retrieved from [http://www.iadis.net/dl/final\\_uploads/200714L022.pdf](http://www.iadis.net/dl/final_uploads/200714L022.pdf)
- Corich, S., Kinshuk., & Jeffrey, L. M. (2009). *Automating the measurement of critical thinking for individuals participating in discussion forums*. Paper presented at the IADIS International CELDA Conference, November 20 – 22. Rome, Italy.

# Table of Contents

---

Abstract .....	i
Acknowledgements .....	iii
Ethics Committee Approval Statement.....	iv
List of Publications .....	v
Book Chapters.....	v
Refereed Journals .....	v
Refereed Conference Proceedings (Paper presentations) .....	v
Table of Contents .....	vii
List of Figures .....	xiii
List of Tables .....	xv
CHAPTER 1: INTRODUCTION .....	1
1.1 Background .....	1
1.2 Thesis Rationale .....	3
1.3 Purpose Statement .....	4
1.4 Research Questions .....	5
1.5 Benefits of the Study .....	6
1.6 Thesis Structure .....	7
1.7 Conclusion.....	8
CHAPTER 2: LITERATURE REVIEW .....	9
2.1 Critical Thinking .....	9
2.1.1 History of Critical Thinking.....	10
2.1.2 Defining Critical Thinking.....	14
2.1.3 Teaching Critical Thinking .....	21

2.1.4	Assessment of Critical Thinking .....	23
2.1.4.1	Commercially available standard tests.....	24
2.1.4.2	Teacher and Researcher Designed Assessment Tools .....	27
2.1.4.3	Student Assessment of Critical Thinking.....	28
2.2	Asynchronous Discussion Forums.....	29
2.2.1	Discussion Forums in Distance Education .....	31
2.2.2	Critical Thinking in Discussion Forums.....	34
2.3	Content Analysis .....	36
2.3.1	History of Content Analysis .....	37
2.3.2	Definition of Content Analysis.....	38
2.3.3	Types of Content Analysis .....	38
2.3.4	The Process of Quantitative Content Analysis .....	40
2.3.4.1	Coefficients of Reliability .....	41
2.3.4.2	Units of Analysis.....	43
2.4	Models to Measure Critical Thinking in Discussion Forums .....	44
2.4.1	The cognitive analytical framework .....	46
2.4.2	Group learning, deep learning and critical thinking model .....	48
2.4.3	Positive and negative indicators of critical thinking .....	49
2.4.4	Knowledge construction model.....	51
2.4.5	Social networking model.....	53
2.4.6	Community of inquiry model .....	54
2.4.7	Model Comparisons.....	59
2.4.8	Adoption of the COI model.....	62
2.5	Computerised Content Analysis .....	66
2.5.1	Data Analysis Software.....	67

2.5.2	Custom Built Analysis Software .....	68
2.5.2.1	An automated discussion forum analyser .....	68
2.5.2.2	An automated text processor .....	70
2.5.2.3	Automated collaborative learning analyser .....	71
2.5.2.4	Comparing Custom Built Analysis Software .....	71
2.6	Identifying Critical Thinking for Individual Participants .....	72
2.6.1	Individual critical thinking measurement tool.....	72
2.7	Summary .....	75
2.7.1	Conclusion.....	78
CHAPTER 3: RESEARCH METHODOLOGY .....		81
3.1	Research Methods .....	82
3.1.1	Quantitative Research .....	82
3.1.2	Qualitative Research .....	83
3.1.3	Mixed Method Research .....	84
3.2	Action Research.....	86
3.2.1	History of Action Research.....	87
3.2.2	Philosophical Foundations of Action Research .....	88
3.2.3	Definition of Action Research .....	91
3.2.4	Types of Action Research.....	93
3.3	Content Analysis .....	95
3.4	The Research Framework.....	97
3.4.1	Phase 1: Literature Review .....	100
3.4.2	Phase 1: Initial development of the artefact.....	101
3.4.3	Phase 2: Investigate models for group contribution.....	103
3.4.4	Phase 3: Investigate models for individual contribution.....	103

3.4.5	Phase 4: Model validation .....	105
3.5	Issues of reliability and validity .....	105
3.5.1	Reliability in quantitative research .....	107
3.5.2	Validity in quantitative research.....	107
3.5.3	Validity and reliability in qualitative research .....	108
3.6	Conclusion .....	109
CHAPTER 4: INITIAL DEVELOPMENT AND TESTING OF THE AUTOMATED		
TOOL .....		
4.1	Background.....	112
4.2	The ACAT System.....	113
4.3	Application of the Initial ACAT System .....	121
4.3.1	Methodology.....	121
4.3.2	Findings and discussion.....	123
4.4	Peer review.....	129
4.5	Conclusion .....	131
CHAPTER 5: USING THE ACAT SYSTEM TO MEASURE GROUP		
PARTICIPATION .....		
5.1	Background.....	133
5.2	Automatic Content Analysis.....	135
5.2.1	The ACAT System: Coding Categorisation Process.....	137
5.3	Text Categorisation Algorithms.....	138
5.3.1	The ACAT System: Text Categorisation Algorithm.....	139
5.4	The ACAT Automatic Coding Process Module .....	140
5.5	Application of the Automated ACAT System.....	147
5.5.1	Methodology.....	147

5.5.2	Findings and discussion .....	148
5.6	Peer review .....	152
5.7	Conclusion.....	154
CHAPTER 6: USING THE ACAT SYSTEM TO MEASURE INDIVIDUAL PARTICIPATION.....		
		157
6.1	Background .....	158
6.2	Validation of the Garrison et al. (2000) Model .....	159
6.3	Second Application of the ACAT Automatic Classification System.....	160
6.3.1	Methodology .....	160
6.3.2	Findings and discussion .....	162
6.3.2.1	Manual Coding for Group Participation .....	162
6.3.2.2	Automatic Coding for Group Participation.....	164
6.3.2.3	Automatic Coding of Individual Participation.....	167
6.4	Peer review .....	173
6.5	Conclusion.....	175
CHAPTER 7: VALIDATING THE ACAT SYSTEM USING THE PERKINS & MURPHY MODEL .....		
		177
7.1	Background .....	178
7.2	The Perkins and Murphy (2006) Model .....	179
7.3	The Cambridge TSA .....	181
7.4	Using the ACAT system with the Perkins and Murphy Model .....	182
7.4.1	Methodology .....	183
7.4.2	Findings and discussion .....	184
7.4	Peer Review.....	194
7.5	Conclusion.....	195

CHAPTER 8: SUMMARY, CONCLUSIONS, LIMITATIONS AND FUTURE

DIRECTIONS .....197

8.1 Background.....198

8.2 Summary of the Research Activities.....199

8.3 Conclusions.....203

8.4 Limitations of the Study .....206

8.5 Contributions .....209

8.6 Implications and Potential for Further Research and Development .....210

8.7 Concluding Remarks.....212

REFERENCES .....215

APPENDICES.....237

Appendix A: Ethics Documentation: Massey University.....239

Appendix B: Ethics Documentation: EIT Hawke's Bay.....241

Appendix C: Information Sheet: Discussion Forum Participants .....243

Appendix D: Consent Form: Discussion Forum Participants .....247

# List of Figures

---

Figure 2.1:	Community of Inquiry Model	55
Figure 2.2:	Practical Inquiry Model	56
Figure 3.1:	Research cycling and knowledge creation	90
Figure 3.2:	Lewin's Research Spiral	92
Figure 3.3:	Four phase action research framework for the study	100
Figure 4.1:	ACAT System Components	115
Figure 4.2:	ACAT Taxonomy Model Editing Screen	116
Figure 4.3:	ACAT Transcript Coding Screen	116
Figure 4.4:	ACAT Results Screen	117
Figure 4.5:	ACAT Coefficient Screen	117
Figure 5.1:	Documentation Classification Process	140
Figure 5.2:	ACAT System Stop Word & Stemming Processes	142
Figure 5.3:	ACAT System Training Text Entry Screen	144
Figure 5.4:	ACAT System Training Text Allocation Screen	145
Figure 5.5:	ACAT Automated System Conceptual Model	146
Figure 7.1:	Scattergram of Cambridge Thinking Skills Assessment score and the ACAT system weighted score	192



# List of Tables

---

Table 2.1:	Critical thinking cognitive skills and sub-skills	18
Table 2.2:	The affective dispositions of critical thinking	20
Table 2.3:	Tests covering more than one aspect of critical thinking	25
Table 2.4:	Survey of computer mediated communication content analysis studies	35
Table 2.5:	Bloom's levels of critical thinking	45
Table 2.6:	Positive and negative indicators of critical thinking	50
Table 2.7:	Models used to measure critical thinking in discussion forums	59
Table 2.8:	Summary of Critical Thinking Models	73
Table 2.9:	Model for identifying engagement in critical thinking	74
Table 3.1:	Emphases of quantitative, mixed, and qualitative research	85
Table 4.1:	Coding categories and indicators for Hara et al. (2000)	119
Table 4.2:	Coding categories and indicators for the COI Model	120
Table 4.3:	Number of postings using the Hara et al. Model	124
Table 4.4:	Number of postings using COI Model	125
Table 4.5:	ACAT calculations for reliability coefficients	126
Table 5.1:	Table structure of the ACAT training dictionary table	143
Table 5.2:	Posting categorisation for different training dictionary populations	149
Table 5.3:	ACAT (raw text, stopword, stemmed) and manual posting categorisations	150
Table 5.4:	Correlation between the ACAT system and the human coders	151

Table 6.1:	Number of postings using the COI model	163
Table 6.2:	Correlation between the ACAT system and the human coders	165
Table 6.3:	Posting categorisation for automatic and manual coding	166
Table 6.4:	Correlation between the ACAT system and the human coders	167
Table 6.5:	Individual participant classification (number of sentences)	169
Table 6.6:	Individual participant classification (as percentage)	170
Table 7.1:	Individual participant classification using the Perkins and Murphy (2006) model (number of sentences)	186
Table 7.2:	Individual participant classification using the Perkins and Murphy (2006) model (percentages)	188
Table 7.3:	Cambridge Thinking Skills Assessment demonstration test results and cumulative ACAT score results	190

# CHAPTER 1: INTRODUCTION

---

Intelligence is something we are born with. Thinking is a skill that must be learned.

Edward De Bono (1933 – )

This thesis is about the automation of the measurement of critical thinking among discussion forum participants. This chapter introduces the study, discusses the rationale for the research topic, the purpose of the research and introduces the research questions. The chapter concludes with a description of the structure of the thesis.

## 1.1 Background

The use of Internet related tools associated with e-learning and the recognition of the need for “critical thinkers” within society have both had a significant impact on the modern educational environment and have provided rich opportunities for academic research. One area of particular interest for researchers has been the identification of evidence of critical thinking among participants within educational discussion forums.

The concept of critical thinking has its roots in the teaching practices of Socrates, described by Plato as a teacher who encouraged his students to question whatever they are told and to seek evidence of clarity and logical consistency. Since the early 1900s academics have attempted to define critical thinking and identify the attributes of critical thinkers. Today critical thinking is recognised as an essential skill for the 21st century and critical thinking has become an important and vital topic of modern education (Schafersman, 1991).

The growth in blended delivery and e-learning among tertiary educators is well documented (Spatariu, Hartley, & Bendixen, 2004). The use of discussion forums as a means of promoting discourse among online learning participants has also been the subject of several academic studies (Henri, 1992; Gunawardena, Lowe, & Anderson, 1997; Newman, Webb & Cochrane, 1995; Garrison et al., 2000; Hara, Bonk, & Angeli, 2000). Researchers investigating the use of asynchronous discussion forums have attempted to identify evidence of critical thinking, which Bloom, Engelhart, Furst, Hill and Krathwohl (1956) suggest should be a prime objective of any form of education, including online learning. A number of models, the majority based loosely on adaptations of Bloom's taxonomy, have been documented and tested. These models have attempted to measure the extent to which knowledge is constructed through the collaborative discourse between discussion forum participants.

While there have been numerous studies to measure evidence of critical thinking among the participants in a discussion forum, almost all the studies provided a measure that described the critical thinking of the group as a collective. The literature provides very little evidence of researchers attempting to measure the critical thinking abilities of individual participants. The only study that could be found which attempted to specifically measure the critical thinking skills of individual participants in a discussion forum was published by Perkins and Murphy in 2006.

McKlin, Harmon, Evans and Jones (2002) suggest that although the transcripts from discussion forums are available in digital format, very few studies have considered the use of computerised tools to analyse the transcripts. McKlin (2004) provides one of the only examples of using a computer to automate the identification of critical thinking among discussion forum participants. The system developed by McKlin used an artificial neural network to analyse cognitive presence in online courses, measuring the critical thinking of the participative group as a collective. The literature provides no evidence of researchers developing computerised tools to automate the measurement of the critical thinking skills of individual participants in a discussion forum.

## 1.2 Thesis Rationale

The increasing recognition of the importance of developing critical thinking skills along with the rapid adoption of distance learning technologies and the apparently limited use of computing technology as a tool to measure critical thinking provides a compelling argument in support of this study.

As an information technology lecturer who has taught for a number of years in the tertiary education sector within New Zealand, the researcher has had experience with using educational technologies to support face to face learning and had used discussion forums as a means of promoting discussion between learners. Having achieved recognition for a record of high student achievement over the years the researcher started to question whether students were merely memorising and gaining surface knowledge to produce good exam results or were actually understanding and questioning what they had learnt and participating in practices that encouraged deeper learning. Following an exploration of the literature relating to constructivist approaches to education and many passionate discussions with teaching peers about the need to promote critical thinking in the classroom, the researcher went about trying to encourage learners to become more actively involved in the learning process and attempted to get them to question what was being presented in class.

While the attempts by the researcher to encourage learners to think critically about the topics discussed in the classroom did not appear to lead to improved student achievement as measured by assessment grades, it did create a learning environment where learners became more involved in the learning process and interaction between learners increased significantly. As a means of further promoting critical thinking among learners, the researcher's attention turned to investigating the possibilities of using educational technology as a mechanism to promote the development of critical thinking skills. A more in-depth exploration of the literature surrounding the use of technology to promote critical thinking led to the research presented in this thesis, which explores the use of discussion forums to promote critical thinking and the

attempts to automate the process of identifying critical thinking among discussion forum participants.

The review of the literature provided evidence that discussion forums were viewed by researchers as having the potential to promote critical thinking skills. The literature also suggested that researchers had developed models to measure evidence of critical thinking among discussion forum participants. One of the most popular approaches used by researchers to evaluate evidence of critical thinking in discussion forum postings is quantitative content analysis (QCA). Berelsen (1952) describes QCA as "a research technique for the objective, systematic, quantitative description of the manifest content of communication" (p. 519). In its simplest form, QCA involves breaking transcripts into units, assigning the units to a category and counting the number of units in each category. QCA is described by many, who have used it, as "difficult, frustrating, and time-consuming" (Rourke et al., 2001, p. 12). Agreement between coders varies considerably and very few researchers duplicate their original models to validate their findings. The literature also suggested that while there was evidence of researchers developing models to measure indicators of critical thinking among discussion forum participants there was almost no evidence of researchers using technology to automate the content analysis process.

### 1.3 Purpose Statement

The purpose of this study was to evaluate whether an automated content analysis tool could be used to identify levels of critical thinking among discussion forum participants. While the literature indicates that this has been done before to measure the critical thinking of the participative group as a collective (McKlin, 2004), this study is the first to attempt to automate the measurement of an individual participant's critical thinking abilities.

The research described in this study attempts to establish whether computers can be used to eliminate the need for human coders in the coding process, which is an important component of the identification of critical thinking.

## 1.4 Research Questions

The research presented in this thesis attempts to address the apparent gap in the literature relating to the use of technology to automate the process of identifying evidence of critical thinking among discussion forum participants.

The research presented in this study attempts to answer two important questions. The first question relates to the ability to automate the measurement of critical thinking of the participants in a discussion forum when considered as a group.

*Is it possible to automate the identification of critical thinking ability exhibited by a group as they participate in a discussion forum?*

To address this question an automated tool was built which incorporated some of the QCA based models used by previous researchers to measure critical thinking among discussion forum participants. The initial development and testing of the tool is described in chapters 4 and 5 of the thesis.

When the research question concerning the possibility of automating the identification of critical thinking ability among discussion forum participants was first proposed, the researcher mistakenly believed that the question could be answered by building a tool, testing it and then attempting to validate the results obtained utilising the tool. The building and testing of the automated tool gave rise to a second and more important question. The second question considers the critical thinking skills of individual participants as opposed to the critical thinking skills of the group.

The second question is:

*Is it possible to automate the identification of an individual participant's critical thinking ability in a discussion forum?*

The second research question concentrates on the critical thinking skills of the individual participant rather than the aggregate critical thinking skills of the group of participants.

To address the second research question a model specifically designed to measure individual critical thinking skills was applied to the automated tool. The subsequent development and testing of the tool is described in chapters 6 and 7 of this thesis. Concentrating on the individual's critical thinking abilities allowed the adoption of an alternate critical thinking measure which provided a mechanism to validate the ability of the tool to measure individual critical thinking.

## 1.5 Benefits of the Study

The research outcomes of this study will build on the existing body of knowledge relating to the measurement of critical thinking. The study will improve understanding of how automated tools can be used to help measure evidence of critical thinking, processes and how computers might be able to replace human coders in the critical thinking identification process. The tool developed and tested during this study provides a mechanism to measure evidence of critical thinking in a discussion forum and provides valuable feedback to the discussion forum administrator. Perhaps the most significant benefit of the study relates to the measurement of the critical thinking abilities of individual discussion forum participants. Automating the measurement of the critical thinking abilities of individual participants allows the individual participants to receive feedback on their level of critical thinking.

## 1.6 Thesis Structure

The thesis is divided into eight chapters; the first, as indicated in the introduction to this chapter, introduces the study, discusses the rationale for the research, lists the research questions and describes the structure of the thesis.

The second chapter presents an overview of the literature related to critical thinking and describes the attempts by educators to promote the development of critical thinking in the tertiary education sector. The chapter also investigates the use of discussion forums and, in particular, evidence of attempts by researchers to measure evidence of critical thinking among the discussion forum participants. The literature review presented in chapter 2 places the research discussed in this thesis in context and identifies apparent gaps in the literature, that this study attempts to address.

Chapter 3 describes the research framework and methodologies adopted during the study. A mixed method enquiry was adopted for the study involving a qualitative action research framework combined with quantitative content analysis. A qualitative framework involving several action research cycles was selected as primary mechanism for the research. Within each of the action research cycles, quantitative content analysis was used to categorise the discussion forum transcripts into various categories of critical thinking. The methodology chapter also considers issues of validity and reliability relating to the procedures adopted during the study.

Chapters 4 to 7 describes the four action research cycles conducted during the study. Chapter 4 describes the initial development of the automated tool, how the tool was validated and the results of the validation. Chapter 5 describes how the tool was refined to measure evidence of critical thinking of the participants as a group while chapter 6 describes how the tool was further refined to measure evidence of critical thinking of individual participants. Chapter 7 describes the last action research cycle and details how the model was validated using alternate critical thinking measurement procedures.

Chapter 8 discusses the consolidated results of the four action research cycles, and provides some analysis of the findings. The final chapter also discusses the limitations of the study and explains the boundaries within which the study can be applied. The study concludes by discussing the potential for further research and summarises the findings of the thesis.

## 1.7 Conclusion

Today more than any time in the past, the need for tertiary graduates who are able to demonstrate the ability to think critically is well documented (Nofle & Robins, 2007; Barratt, 2009). While educators and employers agree on the need to produce graduates who have critical thinking skills there is considerable debate about how well tertiary education institutes contribute to the development of these skills. The literature suggests that the use of technological tools like discussion forums can help develop critical thinking abilities and there is evidence of researchers using discussion forums and attempting to measure evidence of critical thinking among participants. There is however an apparent gap in the literature relating to the use of technology to automate the measurement of participants' critical thinking abilities. The research presented in this thesis contributes to the body of knowledge by describing a study where a tool to measure critical thinking was developed and validated in an educational setting. The study also validates a number of models developed by previous researchers and used to measure evidence of critical thinking among discussion forum participants.

# CHAPTER 2: LITERATURE REVIEW

---

"If I have seen further it is by standing on the shoulders of giants."

Isaac Newton (1679 – 1727)

The previous chapter introduced the rationale for the research and the research questions associated with this study. This chapter attempts to place the research in context by presenting an overview of critical thinking and the use of discussion forums in the educational setting. The chapter also describes a number of different approaches that have been adopted in an attempt to measure evidence of critical thinking.

The review of literature will attempt to define critical thinking and establish its importance as a necessary skill for success in the 21st century. The review will address the use of discussion forums as a tool for promoting interaction among discussion forum participants and describe previous attempts by researchers to measure evidence of critical thinking among those participants. The chapter will conclude by reviewing some of the different tools that have been used to measure evidence of critical thinking abilities.

## 2.1 Critical Thinking

Halpern (2003) suggests that while the ability to think has always been important, it is a vital skill for 21st century citizens. We live in an age where the amount of new information produced is increasing at a rate of more than 30% each year (Lyman & Varian, 2003) and the need to critically evaluate information is essential. The impact of the growth of digital media and the ease by which information can be gathered and stored has led to an information explosion. It is a time when there is said to be too much information to digest and where people are unable to locate and make use of the information that they need (Christian, Blumenthal, & Patterson, 2000). To thrive in

today's information-rich environment an individual must be able to use information in critical thinking and problem solving (Doyle, 1994).

Abrami, Bernard, Borokhovski, Wade, Surkes, Tamim and Zhang (2008) suggest that critical thinking is widely recognised as an essential skill for the knowledge age. They also state that most educationalists would agree that critical thinking is one of the most desirable goals of formal schooling. The argument that critical thinking is a fundamental aim of education is supported by Bailin and Siegel (2003) and also by Sheffler (1973) who argue that critical thinking should be of primary importance in the organisation of educational activities.

While there has been an increasing awareness of the importance of critical thinking and problem solving within the educational setting a study conducted by the Association of American Colleges and Universities in 2005 suggests that few undergraduate students can demonstrate these skills. The study revealed that 93% of college faculties consider analytical and critical thinking to be among the most essential skills students can develop. The study also indicated that while a majority of students believe college experiences prepare them to think, only 6% of graduates can actually demonstrate these essential skills. The study reinforces the findings of earlier studies that suggest that a college education has little effect on the ability of a graduate to think critically, interpret texts and formulate well reasoned arguments (Halpern, 1998; Keeley & Browne, 1986; Perkins, 1985).

### 2.1.1 History of Critical Thinking

The historical roots of critical thinking can be traced back to at least 500BC to the early Greek philosophers. The word "critical" comes from two Greek words: *kriticos*, meaning discerning judgment and *kriterion*, meaning standard (Paul & Elder, 2001, p. 369). Among the philosophers most closely associated with critical thinking was Socrates, the creator of the "Socratic Method". A method or philosophy that encourages people to rectify inconsistent and irrational thought processes, including

confused meanings, inadequate evidence, contradictory beliefs and empty rhetoric (Paul, Elder, & Bartell, 1997). Socrates established the importance of seeking evidence, closely examining reasoning and assumptions, analysing basic concepts, and tracing out implications. The thoughts of Socrates were recorded by Plato, one of his students. Plato went on to establish the Academy of Athens, said to be the first institute of higher learning in the western world. Socrates, Plato and Aristotle, a student of Plato, are said to have helped to lay the foundations of natural philosophy, science, and Western philosophy (Xenophon, 1852). Aristotle is also credited as producing the earliest known formal study of logic. From the traditions of the Greek philosophers emerged a belief that in order to gain understanding that was more than just superficial, one must reflect on an issue, ask questions and think independently, systematically and critically.

One of the most influential philosophers to adopt the logic of Aristotle and pursue systematic and critical thinking during the Middle Ages was a Roman Catholic priest, Thomas Aquinas. The ideas and writings of Thomas Aquinas, and in particular his "Summa Theologica", gave rise to Thomism, a philosophical school of thought that believes that truth is known through reason and that criticisms of ideas is a necessary stage in the development of them (Davies, 1993). Fifty years after his death, Thomas Aquinas, was declared a saint and in 1868 the First Vatican Council elevated his status to "teacher of the church."

The Renaissance gave rise to a revival of classical civilisation and learning, with a growing number of scholars starting to think critically about religion, art, society, human nature, law, and freedom. Among those scholars who encouraged freedom of intellect and emphasised the need to question and develop understanding were Thomas More (1478–1535), Niccolò Machiavelli (1469–1527), Desiderius Erasmus (1466–1536), John Colet (1467–1519), Francis Bacon (1561-1626) and René Descartes (1596-1650).

The seventeenth century, often referred to as the beginning of modern philosophy, also known as the Age of Reason or the Age of Rationalism is usually taken as starting with the works of René Descartes, who is said to have been the most influential philosopher of the period. A philosopher from the end of the period, Immanuel Kant classified his predecessors into two schools: the rationalists and the empiricists. The rationalists movement which included René Descartes, Baruch Spinoza, and Gottfried Leibniz support any view appealing to reason as a source of knowledge or justification (Lacey, 1999). Empiricism is a theory of knowledge that asserts that knowledge arises from sense experience, the empiricists built on the philosophies of their predecessor Francis Bacon. The three main empiricists were Thomas Hobbes, John Locke, and George Berkeley.

The 18th century is known as the Age of Enlightenment, it was a time where religious thought and influence gave way to enquiry based on evidence and proof (Hackett, 1992). It was a time when Newtonian science exerted its greatest influence and philosophers like John Locke (1632 - 1704), Dennis Diderot (1713 - 1784), Voltaire (1694 - 1778) and Jean-Jacques Rousseau (1712 - 1778) encouraged educational change, freedom of thought and independent decision making (Hackett, 1992).

In the 19th century, philosophers like Auguste Comte (1798 - 1857) and Herbert Spencer (1820 - 1903) extended the concept of critical thought into the realms of human sociology. Georg Hegel (1770 - 1831), an admirer of Aristotle and Kant, revolutionised Western Philosophy by developing a comprehensive philosophical framework that became known as Hegelianism. Influenced by the philosophy of Georg Hegel, Karl Marx (1818 - 1883), who is credited with being the founder of modern communism, applied critical thought to the social and economic arena. At the same time that Marx was developing his theories on socialism, Charles Darwin (1809 - 1882) was applying critical thought to the history of human culture and developing his theory of natural selection (Paul & Elder, 2001). Towards the end of the 19th century, Sigmund Freud (1856 - 1939) applied critical thought to the field of psychology, laying the foundations for modern psychotherapy. Marx and Freud were

to have a significant influence on the Frankfurt School which gave rise to the concept of "critical theory", a social theory oriented toward critiquing and changing society as a whole, in contrast to traditional theory oriented only to understanding or explaining it.

In the 20th Century William Graham Sumner (1840 - 1910) published a study of the foundations of sociology and anthropology. Sumner documented the tendency of the human mind to think sociocentrically and the parallel tendency for schools to serve the function of social indoctrination (Paul & Elder, 2001). Writing on the need for critical thinking in education Sumner (1906) stated:

Criticism is the examination and test of propositions of any kind which are offered for acceptance, in order to find out whether they correspond to reality or not. The critical faculty is a product of education and training. It is a mental habit and power. It is a prime condition of human welfare that men and women should be trained in it. (p. 632)

Another educator said to have significantly influenced the development of educational and critical thinking in the twentieth century was John Dewey (Smith, 1997). Dewey believed that education must engage with and enlarge experience and that an educator's role was to encourage students to think and reflect. The influence of Dewey is evident in the writings of Coyle, Kolb, Lindeman and Rogers, educationalists that suggest that learning should be experiential, student focused and concerned with the development of critical and analytical thinking.

The concept of critical and analytical thinking has gained popularity among modern educationalists. Robert Ennis, Stephen Norris, John McPeck, Richard Paul, Harvey Seigel and Peter and Norren Facione are some of the more recognisable authors who have encouraged the adoption of critical thinking concepts in education. Today critical thinking is recognised as one of the main goals in education (Schafersman,

1991) and yet many educational institutes fail to encourage their learners to be reflective and think critically.

### 2.1.2 Defining Critical Thinking

According to Fisher (2001), critical thinking has become somewhat of a buzz word in educational circles. Educators have moved from worrying about teaching information and content and have become more concerned with teaching "thinking skills." In the past, the emphasis for teachers was on delivering content and learning how to think was assumed to have been done indirectly or implicitly in the course of teaching content. Today questions are being asked about the effectiveness of developing critical thinking skills in this way and educational institutes are increasingly looking at ways to teach thinking skills directly (Fisher, 2001).

The literature suggests that there is a lack of consensus on how critical thinking is defined and there is some disagreement on which thinking skills can be taught and how they can be taught. Reed (1997) suggests that there is also confusion over the terminology with scholars referring to "critical thinking", "higher order thinking", "critical reflection," "reflective judgment," and "metacognition" interchangeably. Some of the confusion can be attributed to the differing viewpoints of the two disciplines that have been interested in studying critical thinking. Philosophers tend to concentrate on the nature and quality of the products of critical thinking while psychologists tend to emphasise the process of cognition (Reed, 1997). It is also interesting to note that cognitive and developmental psychology tend to be based on empirical research, while philosophy relies on logical reasoning to reach conclusions.

History indicates that people have been thinking about "critical thinking" for hundreds of years. The debate started with Socrates over 2,000 years ago and scholars have been researching how to teach critical thinking for at least the last 100 years. John Dewey, an American philosopher, psychologist and educational reformer,

regarded by many as the father of the modern day critical thinking called it reflective thinking and defined it as:

Active, persistent, and careful consideration of a belief or supposed form of knowledge in the light of the grounds which support it and the further conclusions to which it tends. (Dewey, 1909, p. 9)

Fisher (2001), when discussing Dewey's definition, suggests that the word "active" indicates a process where an individual thinks things through, raises questions and finds relevant information rather than learning passively from someone else. He also states that when using the words "persistent" and "careful" Dewey is inferring that an individual should take time to think, contrasting it with the kind of unreflective thinking when we 'jump' to a conclusion or make a 'snap' decision without thinking about it. When Dewey refers to the "grounds which support a belief and the further conclusions to which it tends", he is saying that what matters are the reasons one has for believing something and the implications of those beliefs.

Another scholar who attempted to define critical thinking was Edward Glaser, co-author with Watson of the Watson-Glaser Critical Thinking Appraisal (WGCTA), one of the most widely used tests of critical thinking abilities. Glaser defined critical thinking as:

(1) an attitude of being disposed to consider in a thoughtful way the problems and subjects that come within the range of ones' experience; (2) knowledge of the methods of logical enquiry and reasoning; and (3) some skill 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. (Glaser, 1941, p. 5).

Drawing on the Dewey definition, Glaser mentions "persistence of effort" and "knowledge in the light of the evidence that supports it and the further conclusions to which it tends." In addition he talks about "an attitude of being disposed to consider in a thoughtful way the problems" and "some skill in applying those methods", suggesting that critical thinking is more than just possessing the abilities to think critically, a person must have the skills, know how to use them and also possess a desire to apply them. Glaser (1941) also described what he believed were the abilities that one would expect from a critical thinker. These include the ability:

(a) to recognise problems, (b) to find workable means for meeting those problems, (c) to gather and marshal pertinent information, (d) to recognise unstated assumptions and values, (e) to comprehend and use language with accuracy, clarity and discrimination, (f) to interpret data, (g) to appraise evidence and evaluate statements, (h) to recognise the existence of logical relationships between propositions, (i) to draw warranted conclusions and generalisations, (j) to put to test the generalisations and conclusions at which one arrives, (k) to reconstruct one's patterns of beliefs on the basis of wider experience; and (l) to render accurate judgments about specific things and qualities in everyday life. (Glaser, 1941, p. 6).

Robert Ennis, author of the widely used Cornell Critical Thinking Test, defined critical thinking as "reasonable, reflective thinking that is focused on deciding what to believe or do." (Norris & Ennis, 1989, p. 18)

Like the earlier definitions Ennis mentions "reasonable and reflective thinking", however he adds the need for decision making, stating that critical thinkers must focus on what to believe.

During 1988 and 1989, under the sponsorship of the American Psychological Association (APA), a panel of more than forty of the world's leading critical thinking

experts gathered to discuss issues related to the promotion of critical thinking. The panel which consisted of a mixture of philosophers, educationalists and social scientists, adopted a qualitative research methodology known as the Delphi Method which required participants to share their expertise and work toward a consensual resolution of matters of opinion. The outcome of the meetings was published in 1990 in a report compiled by Peter Facione which became known as the Delphi Report. The Delphi Report defined critical thinking in terms of cognitive skills and affective dispositions. The cognitive skills were listed in the first part of the definition which states:

We understand critical thinking to be 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, 1990, p. 3)

The Delphi experts were able to define each of the cognitive skills and identify a number of sub-skills for each skill category. Table 2.1 lists the skills and sub-skills and provides a brief description of each skill category.

**Table 2.1**  
**Critical thinking cognitive skills and sub-skills**

<b>Skill</b>	<b>Description</b>	<b>Sub-skills</b>
Interpretation	To comprehend and express the meaning or significance of a wide variety of experiences, situations, data, events, judgments, conventions, beliefs, rules, procedures or criteria.	Categorisation Decoding Significance Clarifying Meaning
Analysis	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,	Examining Ideas Identifying Arguments Analyzing

	or opinions.	Arguments
Evaluation	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.	Assessing Claims Assessing Arguments
Inference	To identify and secure elements needed to draw reasonable conclusions; to form conjectures and hypotheses; to consider relevant information and to reduce the consequences flowing from data, statements, principles, evidence, judgments, beliefs, opinions, concepts, descriptions, questions, or other forms of representation.	Querying Evidence Conjecturing Alternatives Drawing Conclusions
Explanation	To state the results of one's reasoning; to justify that reasoning in terms of the evidential, conceptual, methodological, criteriological and contextual considerations upon which ones' results were based; and to present one's reasoning in the form of cogent arguments.	Stating Results Justifying Procedures Presenting Arguments
Self-Regulation	Self-consciously to monitor one's cognitive activities, the elements used in those activities, and the results educed, particularly by applying skills in analysis and evaluation to one's own inferential judgments with a view toward questioning, confirming, validating, or correcting either ones' reasoning or one's results.	Self-examination Self-correction

Source: Facione, P. A. (1990, p.6). *Critical thinking: A Statement of expert consensus for purposes of educational assessment and instructions. The Delphi Report.* Millbrae, CA: California Academic Press.

While there was consensus among Delphi panelists that the cognitive skills listed are an essential part of the definition of critical thinking, there was some debate concerning the need to include the dispositions of critical thinkers within the definition. There was agreement that having the cognitive skills did not automatically make a person a critical thinker, and that to be thought of as a critical thinker a person

must demonstrate that they can use the skills. The affective dispositions of an effective critical thinker were described in the second part of the critical thinking definition as being:

habitually inquisitive, well-informed, trustful of reason, open-minded, flexible, fair-minded in evaluation, honest in facing personal biases, prudent in making judgments, willing to reconsider, clear about issues, orderly in complex matters, diligent in seeking relevant information, reasonable in the selection of criteria, focused in inquiry, and persistent in seeking results which are as precise as the subject and the circumstances of inquiry permit. (Facione, 1990, p. 3)

The Delphi panelists expanded the list of dispositions and divided them into two categories: dispositions exhibited in approaches to living and life in general and dispositions exhibited when addressing specific issues, questions or problems. Table 2.2 details the dispositions in the two categories.

The addition of affective dispositions to the definition of critical thinking follows the lead provided by Glaser (1941), and has become an accepted practice when researchers describe what is meant by critical thinking. When identifying evidence of critical thinking researchers today often look for indicators that a person exhibits as their dispositions.

**Table 2.2**

**The affective dispositions of critical thinking**

---

Approaches to life and living in general

- Inquisitiveness with regard to a wide range of issues
- Concern to become and remain generally well-informed
- Alertness to opportunities to use CT
- Trust in the processes of reasoned inquiry
- Self-confidence in ones' own ability to reason
- Open-mindedness regarding divergent world views
- Flexibility in considering alternatives and opinions
- Understanding of the opinions of other people
- Fair-mindedness in appraising reasoning
- Honesty in facing ones' own biases, prejudices, stereotypes, egocentric or sociocentric tendencies
- Prudence in suspending, making or altering judgments
- Willingness to reconsider and revise views where honest reflection suggests that change is warranted

Approaches to specific issues, questions or problems

- Clarity in stating the question or concern
- Orderliness in working with complexity
- Diligence in seeking relevant information
- Reasonableness in selecting and applying criteria
- Care in focusing attention on the concern at hand
- Persistence though difficulties are encountered
- Precision to the degree permitted by the subject and the circumstance

---

Source: Facione, P. A. (1990, p.6). *Critical thinking: A Statement of expert consensus for purposes of educational assessment and instructions. The Delphi Report*. Millbrae, CA: California Academic Press.

Today when researchers talk about critical thinking they tend to refer to the Delphi Report definition. When considering what it means to be a critical thinker, researchers talk about cognitive skills and abilities like analysis, evaluation and self-regulation (Ennis, 2004). They also refer to evidence of critical thinking dispositions such as a willingness to suspend judgment, being open-minded, exhibiting self-confidence and having a willingness to engage in sustained critical thinking (Tishman & Andrade, 1999; Yang & Chou, 2008). Facione (2007) points out that critical

thinking skills and critical thinking dispositions are two different things. He suggests that people who have the skills and do not display the dispositions are not behaving as critical thinkers. He also suggests that those who display the dispositions but do not have the skills are failing to behave as critical thinkers. Critical thinkers are those who have the cognitive critical thinking skills and demonstrate by their behaviour, dispositions towards critical thinking.

While there is still some debate over the definition of critical thinking, most researchers agree that critical thinking is self-guided and self-disciplined thinking. Critical thinkers use intellectual tools that enable them to analyse, assess and improve their thinking. They recognise that no matter how skilled they are as thinkers, that they will always have room to improve their thinking. They also recognise that at times they will fall prey to mistakes in reasoning, prejudices, biases and self-interest that will cloud their judgement. Critical thinkers are committed to continually seek self-improvement and will try not to think simplistically about complex issues (Critical Thinking Community, n.d.).

### 2.1.3 Teaching Critical Thinking

While there is general agreement between educationalists and employers that critical thinking skills are essential for today's graduates wishing to join the workforce, educationalists often disagree about how those skills can be developed. Reed (1997) suggested that prior to the mid 1990s it was generally assumed that students attending colleges and universities would automatically develop their critical thinking skills by attending classes, listening to lectures, participating in class discussions, taking tests and completing regular course assignments. Studies indicate that this assumption may not be true, suggesting instead that improving students' thinking requires more explicit teaching of critical thinking skills (Halpern, 1998; Keeley & Browne, 1986; Perkins, 1989; Underbakke, Borg & Peterson, 1993).

The recognition that traditional education was doing little to encourage improvements in critical thinking gave rise to what Paul (1997) describes as three "waves" of critical thinking research. The first wave, based on the theory of logic, argumentation, and reasoning, saw researchers design and deliver individual courses in critical thinking or informal logic. Students commencing their tertiary studies would participate in such a course which aimed to provide the foundational intellectual skills they need to be successful in their subsequent studies. Paul (1997) suggests that there were both theoretical and pedagogical problems associated with this approach. Theoretically there was a problem associated with taking the narrow concepts of logic presented in these courses and using it as the basis for the transfer of critical thinking to the broader curriculum and the complex problems of everyday life and thought. Pedagogically little was done to restructure learning and instruction and establish how academic subjects should be taught so that students could leave school with the intellectual skills necessary to adapt to incessant and accelerating change and complexity.

The second wave of thinking research was led by a group of researchers who recognised that critical thinking cannot be developed by a single course in logic and critical thinking. They began to investigate ways of integrating critical thinking into instruction across the curriculum and across all grade levels. Paul (1997) states that while the second wave approach set out to address the integration of critical thinking across the curriculum, it did not consider if it was appropriate to apply the theory of informal logic developed for the specialist critical thinking courses to other disciplines and everyday life.

The third wave of critical thinking, research, which Paul (1997) states is in its infancy, sets out to address the problems associated with the first two waves and provides a comprehensive theory of thinking and critical thinking with a clear set of intellectual standards and an integrated set of dispositions. Paul states that such an approach needs to accommodate the role of emotion, intuition, imagination, and values in thinking and provide a comprehensive concept of logic.

Reed (1997) while reviewing the literature regarding the teaching of critical thinking identified several examples of teaching strategies. Her review provides examples of teaching strategies from each of the three waves periods. As an example from the first wave approach she mentions the use of specialised critical thinking courses and discusses how such an approach was adopted in California where teaching and assessing critical thinking skills became a state wide requirement. As a second wave example she mentions a discipline specific approach which was adopted as a way of enhancing the abilities of students to think critically and she describes a strategy that involves the incorporation of critical thinking across the curriculum. Researchers who have adopted this approach include Adler (1992), King (1990), Paul (1997) and Tishman, Perkins, & Jay (1995).

While there are a number of different teaching strategies for helping students develop their critical thinking skills, the literature suggests that there is very little evidence of empirical research being conducted to verify if one approach works better than any other.

#### 2.1.4 Assessment of Critical Thinking

Without appropriate assessment tools it is impossible to judge the effectiveness of any teaching strategies used to help students develop their critical thinking skills. Despite the differences in opinion regarding the definition of critical thinking and the number of different teaching strategies adopted, a number of assessment techniques have been developed to measure the ability of students to think critically.

Paul and Elder (2001) suggest that three main assessment approaches exist;

- commercially available standard tests;
- researcher/instructor designed assessment tools built around a particular course or research project; and

- encouraging students to assess their own thinking.

This study uses tools from the first two categories and in the concluding chapter, discusses the use of the third approach. Each approach is discussed in more detail within this chapter.

#### 2.1.4.1 Commercially available standard tests

Ennis (1993) in his paper entitled "Critical Thinking Assessment" reviewed a number of commercial critical thinking tests that were available at the time, pointing out that while many tests existed, most of them were not comprehensive. He expressed concern that although the test creators claimed to be able to measure evidence of critical thinking many of the tests were not designed with the measurement of critical thinking as the primary concern. He also expressed concern about the number of tests which for convenience's sake had used only multiple choice questions. He suggested that it was better to provide an opportunity for those being tested to provide written justification to explain the reasoning for selecting a particular multiple choice answer.

Ennis (1993) divided the critical thinking assessment tools into two categories, those that assessed multiple aspects of critical thinking and those that assessed a single aspect of critical thinking. Table 2.3, which is based on Ennis (1993. p. 183), lists the tests that assessed multiple aspects of critical thinking and identifies the name of the test, the author of the test and a summary of what the test assesses.

**Table 2.3**

**Tests Covering More Than One Aspect of Critical Thinking**

---

The California Critical Thinking Skills Test: College Level (1990) by P. Facione.  
Aimed at college students, but probably usable with advanced and gifted high school students. Incorporates interpretation, argument analysis and appraisal, deduction, mind bender puzzles, and induction (including rudimentary statistical inference).

---

Cornell Critical Thinking Test, Level X (1985) by R.H. Ennis and J. Millman.  
Aimed at Grades 4- 14. Sections on induction, credibility, observation, deduction, and assumption identification.

---

Cornell Critical Thinking Test, Level Z (1985) by R.H. Ennis and J. Millman.  
Aimed at advanced or gifted high school students, college students, and other adults. Sections on induction, credibility, prediction and experimental planning, fallacies (especially equivocation), deduction, definition, and assumption identification.

---

The Ennis-Weir Critical Thinking Essay Test (1985) by R.H. Ennis and E. Weir.  
Aimed at grades 7 through college. Also intended to be used as a teaching material. Incorporates getting the point, seeing the reasons and assumptions, stating one's point, offering good reasons, seeing other possibilities (including other possible explanations), and responding to and avoiding equivocation, irrelevance, circularity, reversal of an if-then (or other conditional) relationship, overgeneralisation, credibility problems, and the use of emotive language to persuade.

---

Judgment: Deductive Logic and Assumption Recognition (1971) by E. Shaffer and J. Steiger.  
Aimed at grades 7-12. Developed as a criterion-referenced test, but without specific standards. Includes sections on deduction, assumption identification, and credibility, and distinguishes between emotionally loaded content and other content.

---

New Jersey Test of Reasoning Skills (1983) by V. Shipman. Aimed at grades 4 through college. Incorporates the syllogism (heavily represented), assumption identification, induction, good reasons, kind and degree.

---

Ross Test of Higher Cognitive Processes (1976) by J.D. Ross and C.M. Ross.  
Aimed at grades 4-6. Sections on verbal analogies, deduction, assumption identification, word relationships, sentence sequencing, interpreting answers to questions, information sufficiency and relevance in mathematics problems, and analysis of attributes of complex stick figures.

---

---

Test of Enquiry Skills (1979) by B.J. Fraser.

Aimed at Australian grades 7-10. Sections on using reference materials (library usage, index, and table of contents); interpreting and processing information (scales, averages, percentages, proportions, charts and tables, and graphs); and (subject-specific) thinking in science (comprehension of science reading, design of experiments, conclusions, and generalisations).

---

Test of Inference Ability in Reading Comprehension (1987) by L.M Phillips and C. Patterson.

Aimed at grades 6-8. Tests for ability to infer information and interpretations from short passages. Multiple choice version (by both authors) and constructed response version (by Phillips only).

---

Watson-Glaser Critical Thinking Appraisal (1980) by G. Watson and E.M. Glaser.

Aimed at grade 9 through adulthood. Sections on induction, assumption identification, deduction, judging whether a conclusion follows beyond a reasonable doubt, and argument evaluation.

---

Source: Ennis, R.H. (1993, p. 183). Critical thinking assessment. *Theory into Practice*, 32(3), 179-186.

---

Many of the tools used to measure evidence of critical thinking that were identified by Ennis in 1993 are still in use today. Four of the most commonly used tests; the California Critical Thinking Skills Test; the Cornell Critical Thinking Tests; the Ennis-Weir Critical Thinking Essay Test; and the Watson-Glaser Critical Thinking Appraisal have been redeveloped and improved. Since 1993 a number of new tools have been developed; the most widely used is the Cambridge Thinking Skills Assessment (TSA) which was developed in 2001 as an entrance test for certain programmes offered at Cambridge University. The Cambridge TSA is now licensed around the world and is used as an admission exam by many colleges and universities. The Cambridge TSA consists of 50 multiple choice questions and is used to measure critical thinking and problem solving skills.

The Critical Thinking Foundation has also produced a number of critical thinking assessment tools, one of which is the International Critical Thinking Test developed by Linda Elder, Richard Paul and Rush Cosgrove (The Critical Thinking Community, n.d.). The test is an online tool that has three parts and consists of 100 items. The

International Critical Thinking Test is said to focus on five essential dimensions of critical thinking; the analysis of thought; the assessment of thought; the dispositions of thought; the skills and abilities of thought; and the obstacles or barriers to critical thought.

While the literature provides plentiful evidence of researchers using commercial tools to identify critical thinking skills in both the educational and the commercial setting (Abrami et al., 2008; Giancarlo, Blohm, & Urdan, 2004), there is little evidence of researchers comparing different tools to establish their validity. Some researchers suggest that many of the commercial tools involve "sequential problem solving" (Bransford & Schwartz, 1999, p. 68) and as such they fail to adequately test thinking skills employers require from their workers. Yeh (2001) suggests that many of the commercial tools test component skills and bypass the requirement to weigh and select evidence.

#### 2.1.4.2 Teacher and Researcher Designed Assessment Tools

The growing acceptance among educators of the need to develop the critical thinking skills of learners has led to an increase in the number of teachers and researchers developing their own assessment tools. Norris and Ennis (1989) provide examples of tests and techniques that could be used to help assess critical thinking programmes and to evaluate evidence of critical thinking. The majority of the teacher developed critical thinking assessment tools utilise instructional rubrics that describe varying levels of quality for specific assignments as a way of measuring quality of thinking. Andrade (2000) explains that instructional rubrics are popular because they are easy to use, make teachers' expectations clear, provide students with informative feedback, support the development of skills such as critical thinking, and support the development of understanding. Rubrics to measure critical thinking are widely available and can be downloaded from websites such as the Critical Thinking Community (n.d.), Insight Assessment (n.d.) and the Washington State University

Critical Thinking Project (n.d.). Teachers have been using rubrics to help grade student work for many years. The use of rubrics is widely accepted and they have been found to provide a framework that allows teachers to categorise activities into different levels of quality (Andrade, 2000).

A number of researchers have turned to content analysis as a way of measuring evidence of critical thinking (Henri, 1992; Gunawardena et al., 1997; Newman et al., 1995; Garrison et al., 2000; and Hara et al., 2000). The principles of content analysis and the models used are described in more detail later in this chapter.

As is the case with the commercially available critical thinking tests, few studies could be found on the comparative merits of the different tools that are available for measuring indications of critical thinking. Similarly the literature is quiet on studies that have attempted to validate instruments used to measure critical thinking. Perhaps this indicates that many of the researchers who have indicated an interest in investigating critical thinking have not taken a critical thinking approach to their work.

#### 2.1.4.3 Student Assessment of Critical Thinking

Reed (1997) suggests that teaching students to assess their own thinking is the most appropriate way to assess students' critical thinking. This sentiment is supported by Elder and Paul (1994) who have written extensively on the benefits of encouraging students to assess their own level of critical thinking. Elder and Paul state that:

Critical thinking is best understood as the ability of thinkers to take charge of their own thinking. This requires that they develop sound criteria and standards for analysing and assessing their own thinking and routinely use those criteria and standards to improve its quality. (Elder & Paul, 1994, p. 34-35).

While teaching students to assess their own thinking is probably the best approach for developing critical thinking skills, such an approach requires both the teacher and the learner to have a sound understanding of critical thinking processes and to be fully committed to enhancing the quality and level of thinking.

Once again, the literature provides little evidence of published research related to studies that have assessed the improvements in critical thinking resulting from teaching students how to measure their own levels of critical thinking.

## 2.2 Asynchronous Discussion Forums

Asynchronous discussion forums originated from the early bulletin board systems (BBS) which appeared in the 1970s and allowed users to connect and login to a remote computer system using a telephone line, modem and terminal program. The original BBS allowed users to perform functions such as uploading and downloading software and data, reading news and bulletins, and exchanging messages with other users, either through electronic mail or in public message boards (Internet Forum, 2010). With the growth of the Internet in the early to mid 1990s, the use of BBS waned and the public message boards gave way to the asynchronous discussion forums that are in use today.

Like their BBS predecessors, the early online asynchronous discussion forums provided limited services. A user could log onto the system and post messages and then await a response. Some discussion forums also allowed anonymous access, but most expected users to have a user name and account. Today discussion forums have a variety of features and are often classified according to the features that are provided. Forums can be flat or threaded. In a flat forum new messages are added to the end of the discussion and there is no relationship to the previously posted messages. A threaded forum allows users to specify that their message is a response to a previous message and they can display relationships between messages and topics. Threaded forums can be semi threaded or fully threaded. With a semi-threaded

forum a user can specify that their message is a reply to a specific message or topic. A fully threaded forum provides extra functionality by allowing users to reply to a message or reply to a reply to a message and the system can then display a thread showing the original message with the replies and within each reply the responses to the replies. The more sophisticated discussion forums provide extra features that can help reduce duplicate threads occurring; when a user creates a new thread, the system displays a list of existing threads with similar topics.

As discussion forums have evolved, the number of features available to help enrich the quality of online discussion has increased. Almost all asynchronous discussion forum systems allow forum participants to add images to their postings. Some discussion forum systems also allow documents and multimedia content to be added to a post. One of the more useful features that has been included within discussion forum software is the ability to include web site addresses within a post, allowing anyone who reads the post to link to content available on the Internet by selecting a link. Most modern discussion forum systems are able to track unread messages, so that when a user logs on to the system they are presented with messages that have been posted since their last visit. Another useful feature becoming more prevalent is the inclusion of word censoring software, which detects inflammatory content and prevents post containing such content from being displayed. Other additional features include text searching facilities, spam filtering and the ability to generate usage statistics.

Modern discussion forum systems provide functionality that allows a designated forum user to be appointed as a forum moderator. The moderator sets the tone of the forum by promoting forum interaction and encouraging forum participation (Lam, 2004). Brazell (2003) suggests that while the effective moderation of a successful forum takes a great amount of discipline and ability, it yields great success and respect. Brazell also suggests that an effective moderator is someone who is able to lead discussion and promote lively debate while at the same time monitoring inappropriate forum contributions. Mason (1997) describes the role of the moderator

in both a technical and an educational sense. At a technical level the moderator can alter or delete any message posted to the forum and is responsible for removing offensive or irrelevant material. At an educational level, the moderator guides online discussion, stimulates participation and often offers intellectual leadership. Klemm & Snell (1996) when discussing the role of the moderator state that one of the functions of a discussion forum moderator is to promote collaborative learning by encouraging participants to work together to help each other learn.

### 2.2.1 Discussion Forums in Distance Education

The expansion of distance education is transforming the higher education environment (Kriger, 2001). The number of tertiary institutes offering blended and fully online coursework is expanding rapidly and the number of Internet-based distance education courses is steadily increasing. Asynchronous tools, such as discussion forums, are being used to replicate features available in traditional face-to-face education. When used to encourage knowledge building and social reinforcement, asynchronous communication tools provide opportunities for learning communities that are not available within the confines of the traditional classroom (Moller, 1998).

Asynchronous discussion forums are convenient because they are not time or place dependent (Frey, Sass, & Alman, 2006). Students are able to respond at their own pace and they have an equal opportunity to express themselves (Palloff & Pratt, 2002; Peters, 2000). Since discussion forums provide a permanent record of interaction they provide transcripts that are easy to archive and search (Meyer, 2004; Clouse, 2003). Thomas (1999) suggests that asynchronous discussions are collaborative in nature, creating an environment that can encourage social construction of knowledge.

Asynchronous discussion forums are recognised as having the potential to support learning communities where learners actively interact and construct knowledge by sharing experiences and information (Jonassen, 1994). Educational researchers have

reported positive outcomes using threaded discussions which encourage students to accept responsibility for building knowledge by reflecting on course materials and discussing content with fellow participants (Lamy & Goodfellow, 1999).

Learning communities are said to share common goals, needs and problems and can promote solutions by sharing collective knowledge (Rovai, 2002). Online communities are described by Rheingold (1994, p. 57) as, “cultural aggregations that emerge when enough people bump into each other often enough in cyberspace.” According to Whittaker, Isaacs and O'Day (1997), online communities have a number of core attributes:

- Shared goals, interests, needs or activities,
- Repeated, active participation, with intense interactions and strong emotional ties between participants,
- Access to shared resources with policies to determine access,
- Reciprocity of information, support and services between members, and
- Shared context (social conventions, language, protocols).

Palloff and Pratt (1999) describe learning communities as places where people with mutual interests join together to examine a particular theme and learn by exchanging existing knowledge. Online learning communities exist where people sharing common goals, with appropriate access, can pose questions and respond with answers and suggested resources. Within an online learning community knowledge and meaning are actively constructed, and the members of the community enhance the acquisition of knowledge and understanding, thereby satisfying learning needs (Rigou, 2004). Members of a learning community may be students, lecturers, tutors, researchers, practitioners and domain experts.

Discussion forums are increasingly seen as one of the most powerful tools for creating online learning communities (Sergiovanni, 1999; Swan & Shea, 2005). Used appropriately, discussion forums can enable rapid dissemination of information and

can encourage feedback and the refinement of ideas among participants (Hiltz & Turoff, 1993). Discussion forums are being increasingly used to promote collaboration among a diverse variety of people from a wide range of settings and locations (Mayadas, 2001). While support for the use of discussion forums as a means of creating online communities is widely accepted, critics point out that unless they are managed effectively, they can easily become a vehicle for idle chatter and have the affect of promoting discord amongst an online community (Pallof & Pratt, 2002).

The pedagogy supporting the use of asynchronous communication tools within learning communities has its roots in constructivism and social constructivism (Knowles, Holton & Swanson, 1998; Palloff & Pratt, 1999; Squire & Johnson, 2000). In a constructivist learning environment control shifts from the instructor to the learner. One of the main principles of constructivism involves using open-ended questions which replicate realistic problem situations, enabling learners to develop skills in complex and unstructured problem-solving situations. Constructivism involves learning in a social context using group activities, collaboration and teamwork. Participants in a constructivist setting have shared goals which are negotiated between instructors and learners and between themselves and learners; in this situation the instructor has the role of facilitator or coach.

Social constructivists suggest that the optimal learning environment is one where there is a dynamic interaction between instructors and learners and tasks provide an opportunity for learners to create their own understanding by communicating with others. McMahon (1997) claims that social constructivism emphasizes the importance of culture and context in understanding what is happening in society and constructing knowledge based on this understanding. Participants in a social constructivist environment collaborate by sharing experiences and information in a way that promotes critical thinking and knowledge construction.

### 2.2.2 Critical Thinking in Discussion Forums

Rourke, Anderson, Garrison and Archer (2000) state that discussion forums create a unique and valuable environment for distance, distributed, and lifelong learning by supporting interaction among participants. They also point out that the automatically recorded and machine-readable data generated by conferencing software provides a compelling source of data for educational researchers and software developers. Fahy (2001) suggests that computer conferencing is virtually ubiquitous in distance education and that the temptation to analyse the resulting interaction for evidence of critical thinking has met with only partial success. He claims that complexity of the measurement instrument and inappropriate units of analysis are the main reasons for only achieving partial success.

Despite the growing popularity of distance education and the development of online tools such as discussion forums, there is limited empirical evidence to suggest that such tools facilitate higher level thinking (Bullen, 1998). While a number of researchers have developed models for measuring various aspects of critical thinking among discussion forum participants, the primary focus of many of the studies has not been limited to the identification of evidence of critical thinking. Many of the studies have measured critical thinking alongside other activities such as the impact of social interaction and the effect of teacher involvement in the discussion forum activities (Perkins & Murphy, 2006).

The literature suggests that quantitative content analysis (QCA) has proven to be one of the most popular techniques adopted by researchers who have developed models to analyse discussion forum transcripts and measure evidence of critical thinking among discussion forum participants. Rourke et al. (2001) provide an excellent summary of the influential content studies published between 1995 and 2000. Table 2.4 lists the studies which were identified by the authors as involving measurement of critical thinking or cognitive elements. The table describes the unit of analysis, the variables investigated, the reliability and the research technique adopted for each of the studies.

**Table 2.4****Survey of computer mediated communication content analysis studies**

<b>Study</b>	<b>Unit of Analysis</b>	<b>Variables Investigated</b>	<b>Reliability</b>	<b>Research Design</b>
Henri (1992)	Thematic Participation Interaction	Social, cognitive, metacognitive elements	Not reported	Descriptive
Newman, Webb & Cochrane (1995)	Thematic	Critical thinking	Percent agreement after discussion	Descriptive
Zhu (1997)	Thematic Interaction Participation Participant roles	Knowledge construction	Not reported	Descriptive
Bullen (1998)	Thematic Participation	Critical thinking	Not reported	Descriptive
McDonald (1998)	Thematic Participation Interaction	Group development Social, cognitive, metacognitive elements	Cohen's kappa Descriptive	
Weiss & Morrison (1998)	Thematic and Message	Critical thinking Understanding/ correcting misunderstandings Emotion	Percent agreement after discussion	Descriptive
Kanuka & Anderson (1998), Anderson & Kanuka (1997)	Thematic	Collaborative knowledge construction	Not reported	Descriptive

Fahy, Crawford, Ally, Cookson, Keller & Prosser (2000)	Sentence Interaction Participation	Critical thinking	Percent agreement	Descriptive
Garrison, Anderson,& Archer (2000)	Message	Critical Thinking	Cohen's kappa	Descriptive
Hara, Bonk, & Angeli (2000)	Paragraph Participation Interaction	Social, cognitive, metacognitive elements	Percent agreement Coder stability	Descriptive

Source: Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (2001, p. 10).  
Methodological issues in the content analysis of computer conference transcripts.  
*International Journal of Artificial Intelligence in Education*, 12(1), 8-22.

Others who have also adopted QCA as a mechanism for identifying evidence of critical thinking within discussion forums include: Gunawardena et al. (1997); Clulow, Clulow, & Brace-Govan (2003); McKlin et al. (2002); Jeong (2003); Meyer (2004); Perkins & Murphy (2006); and Frey et al. (2006).

### 2.3 Content Analysis

Since QCA is the most widely adopted technique for measuring evidence of critical thinking it is important to understand the origins of content analysis and the processes involved in undertaking a study using content analysis to categorise different levels of activity.

Krippendorff (2004) describes content analysis as one of the most important research techniques in the social sciences. He also explains that "content analysis entails a systematic reading of a body of texts, images, and symbolic matter, not necessarily from a perspective of an author or user" (Krippendorff, 2004, p. 3). White and Marsh (2006) describe content analysis as a highly flexible research method that has been widely used in library and information science (LIS) studies with varying research

goals and objectives. Content analysis is used in qualitative, quantitative, and sometimes mixed modes of research frameworks and employs a wide range of analytical techniques.

### 2.3.1 History of Content Analysis

While the term did not appear in English until 1941, historic evidence of content analysis can be found in 18th century Sweden when the Lutheran church conducted a careful and doctrinal analysis of a hymn book (Dovring, 1954). The mass production and increased availability of printed texts and in particular, growth of newsprint in the early 20th century led to a growing acceptance of content analysis as a research tool and the development of "quantitative newspaper analysis" (Krippendorff, 2004). The emergence of behavioural and social sciences in the period between 1930 and 1940 gave rise to a growing use and acceptance of survey research and with it an increased acceptance of content analysis as a research methodology (Wilson & Rayson, 1993). In the 1950s it gained greater prominence when it was used to study mass communications (White & Marsh, 2006). During this period content analysis moved from being a tool used to look for patterns in newspapers and became increasingly accepted by the scientific community as a tool that could be used to measure public opinion. Since then it has increased its prominence in many fields, including anthropology, library and information studies, management, political science and sociology. With the advent of computers in the 1960s, researchers started to look towards computer technology as a means of doing automatically what had previously be done by hand. This led to the development of the General Enquirer at Harvard University, which despite the development of other computer assisted content analysis programs, has maintained its position as the most sophisticated and automated content analysis tool (Wilson & Rayson, 1993). Today content analysis is still a popular tool in the social sciences.

### 2.3.2 Definition of Content Analysis

Like most research methodologies, multiple definitions for content analysis can be found. One of the earliest definitions of content analysis is attributed to Bernard Berelsen who defined content analysis as "a research technique for objective, systematic and quantitative description of manifest content of communication" (Berelsen, 1952, p. 18). Klaus Krippendorff, Professor of Communication at the University of Pennsylvania Annenberg School for Communication argues that content analysis can be qualitative as well as quantitative. He defines content analysis as "a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use" (Krippendorff, 2004, p. 18). Krippendorff also suggests that the techniques of content analysis are expected to be reliable and that research techniques should result in findings that are replicable. Krippendorff also makes special mention of the reference to other meaningful matter explaining that content analysis can be used to analyse a variety of content including text, images, videos, expressions, tools and signage.

White & Marsh (2006) stress the importance of inference in content analysis, stating that the researcher must use analytical constructs, or rules of inference to move from the text being analysed to formulate answers to the research questions being investigated. They suggest that the analytical constructs could come from existing theories or practices, experience or knowledge of experts and previous research.

### 2.3.3 Types of Content Analysis

The literature suggests that there are two general categories of content analysis: conceptual analysis and relational analysis (Palmquist, Carley, & Dale, 1997). Conceptual analysis concentrates on establishing the existence and frequency of concepts in a text while relational analysis examines the relationships among concepts in a text.

### Conceptual analysis

Conceptual analysis is the more traditional approach to content analysis. Also known as thematic analysis, conceptual analysis begins with the identification of a research question and the selection of the samples, followed by the coding of text into content categories (Palmquist et al., 1997). The emphasis is on quantifying and tallying, looking at the occurrence of selected terms in a text or texts. Conceptual content analysis usually involves the use of a specialised dictionary or a series of contextual rules against which units of the text are classified.

### Relational analysis

Like conceptual analysis, relational analysis begins by identifying the presence of concepts in a text or set of texts. Relational analysis, however goes beyond presence by exploring the relationships between the concepts identified. Also known as semantic analysis, the focus of relational analysis is to look for semantic, or meaningful, relationships (Palmquist et al., 1997). Relational analysis has been adopted by researchers in a number of different areas. Palmquist et al. (1997) describe three subcategories of relational analysis:

- Affect extraction: An approach which provides an emotional evaluation of concepts explicit in a text.
- Proximity analysis: An approach which is concerned with the co-occurrence of explicit concepts in the text.
- Cognitive mapping: An approach that builds on the two previous approaches, by attempting to represent the two relationships visually for comparison to create a model of the overall meaning of the text.

When discussing categories of content analysis researchers also distinguish between quantitative and qualitative content analysis. When discussing quantitative and qualitative content analysis proponents often emphasise the differences, but as Krippendorff (2004) points out, text is always qualitative to begin with and that the

two methods have a number of common elements. Discussing the similarities he points out that "Both sample text, in the sense of selecting what is relevant; both unitise text, in the sense of distinguishing words or propositions and using quotes and examples; both contextualise what they are reading in light of what they know about the circumstances surrounding the texts; and both have specific research questions in mind" (Krippendorff, 2004, p. 67).

White & Marsh (2006) state that quantitative analysis flows from a positivist research tradition and as such adopts a deductive approach. The objective of quantitative content analysis is to test hypothesis and not to develop them. Quantitative content analysts employ a range of techniques, including counting, rating, ranking, logging, and other means of reducing their data to numerical form. The quantitative approach has its roots in the scientific method and is the basis of the conceptual analysis approach to content analysis.

The qualitative analysis approach to content analysis is similar to that described in the relational analysis category. It stems from a humanistic tradition and is inductive in nature. While qualitative content analysis may yield a testable hypothesis, the results of the analysis of the data may lead to the identification of new ideas and concepts and the subsequent investigation of modified or new hypotheses (White & Marsh, 2006).

#### 2.3.4 The Process of Quantitative Content Analysis

QCA is one of the most widely used tools for analysing the content of discussion forum transcripts (Bullen, 1997; De Wever, Schellens, Valcke, & Van Keer, 2006; Fahy, Crawford, & Ally, 2001; Garrison et al., 2001). Rourke et al. (2001) explain that the process of QCA can be broken down into four essential steps. The first step is to identify the representative sample; which in the case of an asynchronous discussion forum involves obtaining a text file of the transcript. The second step involves the creation of a model which will allow the transcript to be categorised and training

coders how to use the model. An alternate strategy would be to adopt an existing model and train coders to code against the predefined categories. The third step involves coding the transcript components into the categories defined by the model. The final step involves comparing the results for different coders to establish reliability.

#### 2.3.4.1 Coefficients of Reliability

Berelsen (1952) stated that content analysis is an objective technique, susceptible to the subjectivity and bias of the coders. QCA is described by many, who have used it, as “difficult, frustrating, and time-consuming” (Rourke et al., 2001, p. 12). Agreement between coders varies considerably and very few researchers duplicate their original models to validate their findings. To ensure the objectivity in content analysis, multiple coders are employed to code the same transcripts and their coding decisions are compared. This comparison of coding results is referred to as inter-rater reliability, which is defined by Rourke et al. (2001) as the extent to which different coders, each coding the same content, come to the same coding decisions.

The simplest and most commonly used method of calculating inter-rater reliability is the percent agreement statistic, which gives a measure of the degree of agreement between raters. Two of the most popular techniques for calculating the percent agreement QCA are Holsti's coefficient of reliability and Cohen's kappa.

Holsti's (1969) coefficient of reliability (CR) provides a simple formula for calculating percent agreement statistic:

$$CR = 2m / ( n1 + n2 )$$

where m = the number of coding decisions upon which the two coders agree, n1 = number of coding decisions made by rater 1, and n2 = number of coding decisions made by rater 2.

The calculation yields a reliability coefficient between 0 (totally unreliable) and 1 (totally reliable), however the formula does not take into account that the coders may have agreed by chance on their coding.

Cohen's kappa is another inter-rater reliability statistic (Cohen, 1968). It calculates the same formula as Holsti's coefficient but takes the role of chance into account. The formula for calculating Cohen's kappa is:

$$k = (F_o - F_c) / (N - F_c)$$

where N = the total number of judgements made by each coder, F<sub>o</sub> = the number of judgements on which the coders agree, and F<sub>c</sub> = the number of judgements for which agreement is expected by chance.

While there are several QCA studies which provide evidence of utilising Holsti's CR and Cohen's kappa, Heath (2005) claims that both measures fail to take into account that coders may have agreed by chance on their coding. Heath suggests using William. A. Scott's pi index or Klaus Krippendorff 's alpha, both of which attempt to take chance out of the equation. Since both of these coefficients are difficult to calculate they are rarely reported.

While choosing an appropriate tool to measure inter-rater reliability may be difficult, the level of inter-rater reliability that must be achieved has not been established (Rourke et al., 2001). Researchers appear unable to agree on acceptable levels of reliability. Riffe, Lacey and Fico (1998) suggest that a minimum level of 80%" (p. 128) is acceptable for communication studies, while Capozzoli, McSweeney and Sinha (1999) state that values greater than 75% may be taken to represent excellent agreement. Rourke et al. (2001) suggest that while there is confusion over the required levels of agreement, the "mere act of reporting these figures gives readers sufficient information to interpret the results" (p. 129).

#### 2.3.4.2 Units of Analysis

When coding transcripts, a researcher must decide the nature of the coding unit upon which the analysis will be based. Units can be syntactical, such as the word, preposition, sentence or paragraph, which are delineated by syntactical criteria. Fahy et al. (2000) used the sentence, Hara et al. (2000) used the paragraph, and the COI model used the message as the unit of analysis. As Rourke et al. (2001) suggest, fixed units such as words, sentences and messages are easily delineated, however they may not always encompass the construct being investigated.

Henri (1992) rejected the construct of using a unit that was syntactically based choosing instead to use a "unit of meaning" which is similar in form to the thematic unit. Budd, Thorp and Donohew (1967) describe the thematic unit as "a single thought unit or idea that conveys a single item of information extracted from a segment of content" (p. 34). An argument against using thematically based units which require in-depth processing to identify the unit is that it increases the opportunity for subjective ratings and low reliability (Rourke et al., 2001).

The evidence of successful adoption of content analysis as a mechanism to measure indicators of critical thinking suggest that it would be an ideal candidate for adoption in the development of an automated tool. The process of content analysis is well defined and should be able to be duplicated in a computerised model. While there is some debate concerning the most suitable unit of analysis, it would be difficult to develop an automated tool that could implement a thematic approach. The adoption of a syntactically based unit of analysis would appear to be a suitable and practical approach to take when designing and building an automated tool. The two most popular measures of inter-rater reliability are Holsti's CR and Cohen's kappa, both being suitable candidates for adoption in an automated system and both able to provide a suitable measure of correlation between the automated system results and the results obtained by human coders.

## 2.4 Models to Measure Critical Thinking in Discussion Forums

Rourke et al. (2001) identified a number of researchers who had used the methods of content analysis to develop models to measure evidence of critical thinking. Since 2001 researchers have continued to develop instruments which use content analysis to identify indicators of critical thinking (Clulow et al., 2003; McKlin et al., 2002; Jeong, 2003; Meyer, 2004; Perkins & Murphy, 2006; Frey et al., 2006). Rourke & Anderson (2004) suggest that rather than develop a new coding scheme, researchers should use schemes that have already been developed and used in previous research. They suggest that such an approach would foster replicability and the validity of the instrument and contribute to a growing catalogue of normative data (Rourke & Anderson, 2004).

The Rourke et al. (2001) study identified nineteen computer mediated communication content analysis studies. The studies considered a number of different variables, including; critical thinking, social interaction, group interaction, linguistic variation and knowledge construction. Of the nineteen studies identified by Rourke et al. (2001), only six focused on the identification of critical thinking indicators. All of the six critical thinking related studies attempted to group messages obtained from discussion forum transcripts into a number of different categories. The categories used in each of the studies defined different levels of critical thinking. While all the studies appear to use categories based on the cognitive domain of Bloom's taxonomy of learning objectives, each study adopts a slightly different approach.

The idea for the taxonomy, named after Benjamin Bloom, was formed at an informal meeting of college examiner's attending the 1948 American Psychological Association Convention in Boston. The idea was discussed at subsequent American Psychological Association conferences over a number of years and was published in an edited volume in 1956 by a committee of educators chaired by Bloom (Bloom et al., 1956). Bloom's Taxonomy divided educational objectives into three domains: Cognitive, Affective, and Psychomotor. Within each of the domains, learning at a

higher levels is dependent on having attained prerequisite knowledge and skills at a lower level. The original edited volume focused on the cognitive domain. A second volume, edited by David Krathwohl, was produced in 1964 and described the affective domain (Krathwohl, Bloom, & Masia, 1964). Although the original intention was to produce a third volume describing the psychomotor domain, the third volume was never published.

Bloom et al. (1956) describe the cognitive domain as having six categories, or levels of learning. The lowest level is knowledge, which is followed by comprehension, application, analysis, synthesis and evaluation. Table 2.5 provides definitions for each level of the Bloom's taxonomy cognitive domain and list verbs that illustrate activities associated with each level.

**Table 2.5**  
**Bloom's levels of critical thinking**

<b>Cognitive level</b>	<b>Definition</b>	<b>Illustrative verbs</b>
Knowledge	Recalling learned material	recall, underline, list, name, record, label, cluster, match, memorise, define, arrange
Comprehension	Understanding the material	understand, show, summarise, explain, describe, demonstrate, review, cite, restate, locate
Application	Using the material	apply, select, model, organise, illustrate, utilise, choose, imitate, demonstrate, use
Analysis	Breaking material down to increase understanding	analyse, compare, contrast, classify, map, characterise, divide, break down, choose, examine
Synthesis	Reshaping material into a new form	construct, speculate, design, compose, create, develop, invent, blend, propose, formulate
Evaluation	Judging the worth of material	evaluate, convince, argue, judge, criticise, rate, measure, persuade, assess, recommend

Adapted from: Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W.H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain*. New York: David McKay Co. Inc.

The use of the cognitive domain of Bloom's taxonomy is evident in the majority of the studies, that were identified by Rourke et al. (2001), as focusing on the identification of critical thinking indicators. The six studies identified by Rourke et al. (2001) as focusing on the identification of critical thinking indicators are discussed below.

### 2.4.1 The cognitive analytical framework

One of the first researchers to adopt content analysis as a way of measuring levels of cognitive activity was Henri. He developed an analytical framework which had five dimensions, one of which was the cognitive dimension (Henri, 1992). The other dimensions included in Henri's framework were: participative, social, interactive and metacognitive dimensions. Henri was particularly interested in the social activity and the interactivity of individuals in the group. When coding transcripts, Henri divided postings into statements corresponding to units of meaning.

Under the cognitive dimension Henri classified postings into one of five categories. DeWever, Schellens, Valcke, & Van Keer (2006, p.12) describe the five categories of Henri's cognitive dimension as follows:

1. Elementary clarification, where participants observe or study a problem, identifying its elements, and observing their linkages in order to come to a basic understanding.
2. In-depth clarification, where participants analyse and understand a problem, shedding light on the values, beliefs, and assumptions which underlie the statement of the problem.
3. Inference, where participants use induction and deduction, admitting or proposing an idea on the basis of its link with propositions already admitted as true.

4. Judgment, where participants make decisions, statements, appreciations, and criticisms.
5. Strategies, where participants propose coordinated actions for the application of a solution, or follow through on a choice or a decision.

Like Bloom's levels of learning, Henri's model assumed that evidence of higher level cognitive activity built upon the lower levels. Similarities between the higher levels of Bloom's classifications and Henri's classifications can be seen. The second level of Henri's model involves analysis and understanding, activities described by Bloom et al. (1956) as occurring at the analysis level. Similarly the fourth level of Henri's model, refers to making judgements, activities found in the synthesis and analysis levels of the Bloom et al. (1956) model.

When Henry (1992) published her findings, she failed to provide details of how messages were classified into each of the five categories, making it difficult for researchers to duplicate or validate her model. She also failed to provide any details on the inter-rater reliability or the code-recode reliability of her instrument and she failed to empirically test her instrument. However, her model was used in a subsequent study by Hara et al. (2000) who when coding messages obtained from 20 doctoral students participating in a discussion forum was able to achieve a 0.75 percentage agreement between two coders for the cognitive dimension of the Henri model. The cognitive dimension of the Henri (1992) model was also used by McKenzie and Murphy (2000) who were able to achieve a percentage agreement of 0.44 when coding the messages obtained from 25 students participating in an 11 week course.

Even though the Henri (1992) model has some resemblance to the cognitive domain of Bloom's taxonomy and has only five categories it is not considered a suitable model for duplication. The lack of detailed criteria for systematic classification and the adoption of a thematic approach to categorisation make this model an unsuitable candidate for automated analysis.

## 2.4.2 Group learning, deep learning and critical thinking model

In 1995 Newman et al. developed an instrument to measure indicators of group learning, deep learning and critical thinking. The model was based on Garrison's (1991) five stages of critical thinking and Henri's (1992) cognitive skills. The instrument had 10 categories: relevance, importance, novelty, outside knowledge, ambiguities, linking ideas, justification, critical assessment, practical utility, and width of understanding. For each category, a number of positive and negative indicators were formulated and a critical thinking ratio was calculated using the totals of the negative and positive indicators.

The relationship between this model and the cognitive dimension of Bloom's taxonomy is difficult to see. Like the study conducted by Henri (1992), little detail is given regarding how the messages were coded into each of the ten categories. The study also suggests that coding requires detailed subject domain expertise making multi coder activities difficult. A thematic approach was adopted for the unit of analysis with researchers looking at sentences, paragraphs or topics that contribute to a specific theme. As Marra, Moore, and Klimczak (2004) suggest this would make calculating inter-rater reliability a difficult if not impossible task.

The research reported by Newman et al. (1995) did not mention reliability of the findings and did not discuss empirical validation of the instrument. Even though the authors suggested that others should replicate their work there is little evidence in the literature to suggest that the instrument has been widely adopted.

The complexity of the model, with ten categories which were poorly defined, along with the adoption of the thematic approach to classification make this a difficult model to duplicate and it is considered to be an unsuitable candidate for automated analysis.

### 2.4.3 Positive and negative indicators of critical thinking

Bullen (1997) developed an instrument that focused on identifying evidence of critical thinking. Bullen's instrument was based on a taxonomy of critical thinking which was developed by Ennis (1987), which consisted of four categories of critical thinking skills. Within each category Bullen looked for evidence of positive and negative indicators of critical thinking. Positive indicators were assigned when a posting provided evidence of critical thinking; negative indicators were assigned for evidence of what he termed uncritical thinking. The four categories described by Bullen were: clarification; assessing evidence; making and judging inferences; using appropriate strategies and tactics.

Table 2.6 lists the positive and negative indicators used in Bullen's instrument for each of the four categories.

**Table 2.6****Positive and Negative Indicators of Critical Thinking**

<b>Category</b>	<b>Positive indicators</b>	<b>Negative indicators</b>
Clarification	Focusing on a question; analyzing arguments; asking and answering questions of clarification; and defining terms and judging definitions.	Focusing on a question unrelated to the problem; analysing arguments inappropriately; asking inappropriate or irrelevant questions; incorrectly answering questions of clarification; incorrectly defining terms; inappropriately judging definitions.
Assessing evidence	Judging the credibility of a source; making and judging observations.	Judgments and observations based on inappropriate criteria.
Making and judging inferences	Making and judging deductions; making and judging inductions; making and judging value judgments.	Making and judging inferences that do not follow the listed criteria.
Using appropriate strategies and tactics	Using models, metaphors, drawings, and symbols to simplify problems; talking through a confusing issue with another person.	The inappropriate use of strategies and tactics.

Source: Bullen, M. (1997). *A case study of participation and critical thinking in a university-level course delivered by computer conferencing*. (Unpublished doctoral dissertation). University of British Columbia, Vancouver, Canada. Retrieved from <http://bullenmark.files.wordpress.com/2009/05/dissertation.pdf>

Bullen did report on the reliability of his research; however, using three coders, there was only a 17% agreement. The low level of agreement was explained as being due to ambiguity in the indicators and disagreement between one of the coders and the other two. An interesting finding from the Bullen (1997) study was the conclusion that there was no relationship between participation and the level of critical thinking.

While Bullen went on to conduct further research using the instrument, evidence could not be found within the literature indicating that others had adopted his instrument or duplicated the study.

Like the model developed by Henri (1992), the model developed by Ennis (1987) and adopted by Bullen (1997) bears some similarities to the higher levels of the cognitive domain of Bloom's taxonomy. However the adoption of positive and negative indicators and the documented ambiguity between indicators adds to the complexity of the Bullen model. That combined with the adoption of the thematic approach to classification make this a difficult model to duplicate and it is considered to be an unsuitable candidate for automated analysis.

#### 2.4.4 Knowledge construction model

Gunawardena et al. (1997) developed an instrument designed to examine the social construction of knowledge in a discussion forum. Based on grounded theory the tool used the phases of a discussion to determine the level of knowledge constructed. The researchers refer to the work of Henri (1992) and Newman et al. (1995) suggesting that these models had provided a useful starting point for further research. The knowledge construction model (Gunawardena et al., 1997) identified five phases of knowledge construction:

- Phase one: sharing and comparing of information, which comprises observations, opinions, statements of agreement, examples, clarifications, and identifications of problems.
- Phase 2: the discovery and exploration of dissonance or inconsistency among ideas, concepts, or statements.
- Phase 3: the negotiation of meaning and/or co-construction of knowledge, which includes negotiation, identifications of areas of agreement, and proposing new co-constructions on topics where conflict exists.

- Phase 4: the testing and modification of proposed syntheses or co-construction against existing cognitive schema, experiences, and literature.
- Phase 5: referencing statements of agreement and application of newly-constructed meaning; encompassing summaries of agreements, applications of new knowledge, and metacognitive statements revealing new knowledge construction.

Like Henri (1992), Gunawardena et al. (1997) also used the unit of meaning as the unit of analysis. They mentioned that the messages were coded independently by two researchers but they did not report any inter-rater reliability coefficients. The coding scheme presented by Gunawardena et al. (1997) was subsequently adopted and tested by Veerman & Veldhuis-Diermanse (2001) who coded 1428 messages produced by 230 undergraduate students and obtained a percentage agreement of 0.69 between two coders. Marra et al. (2004) also used the Gunawardena et al. (1997) model and using Krippendorff's alpha obtained values between 0.50, and 0.93. Schellens & Valcke (2005) also used the Gunawardena et al. (1997) model., they coded 11095 messages and obtained a percentage agreement (Cronbach  $\alpha$ ) between two coders of between 0.42 and 0.99.

Like the cognitive analytical framework model (Henri, 1992) and the positive and negative indicators of critical thinking model (Bullen, 1997), the knowledge construction model (Gunawardena et al., 1997) bears some similarities to the higher levels of the cognitive domain of Bloom's taxonomy. This model however concentrates on constructivist principles of shared knowledge construction and relies on interaction between the discussion forum participants. The Bullen (1997) study used the message as the unit of analysis. Since messages can vary in size, from a short sentence to a complex paragraph, the adoption of the positive and negative indicators of critical thinking model (Bullen, 1997) would make the referencing to the data dictionaries likely to be used in an automated system problematic. Since the positive and negative indicators of critical thinking model (Bullen, 1997) relies on

participant interaction and resulting knowledge construction, the model would not be suitable for identifying evidence of critical thinking of individual participants.

#### 2.4.5 Social networking model

Fahy, Crawford, & Ally (2001) conducted research using a model based on the definition of interaction provided by Gunawardena et al. (1997). Applying the concept of social networking, they focused on structural and interactional exchange of messages observed in transcripts. Their model was also derived from the work of Zhu (1997) and had five categories:

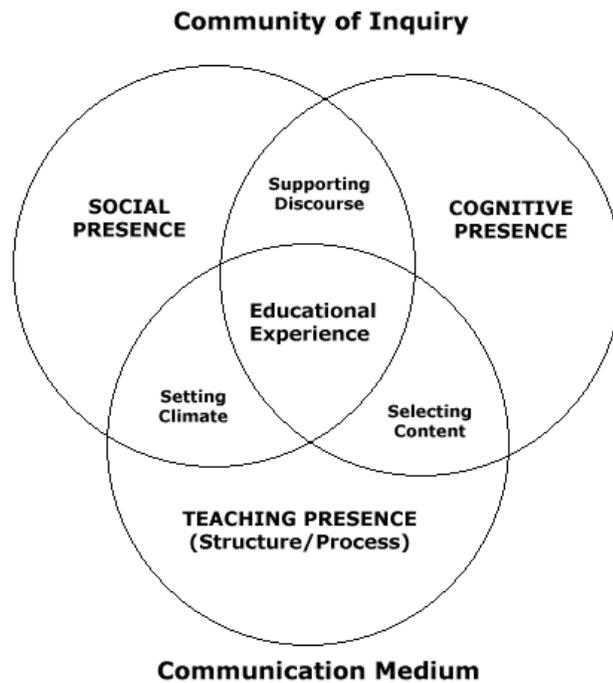
- Questioning, which included vertical questions where it is assumed a right answer exists; horizontal questioning where there might not be a right answer.
- Statements and supports, where the main intention is to impart facts or information without inviting dialogue or response.
- Reflections, where the speaker expresses thoughts, judgments, opinions or information which are personal and are usually guarded or private.
- Scaffolding and engaging, where the intention of the speaker is to initiate, continue or acknowledge interpersonal interaction.
- References, where the speaker provides references and quotations or provides paraphrases of other sources.

Fahy et al. (2001) adopted the sentence as the unit of analysis, however when specifying the results the researchers reported the number of words (53671) rather than the number of sentences. They reported on three studies each involving thirteen participants and using three independent coders reported inter-rater reliability coefficients ranging from 60 to 86% agreement. While Fahy went on to describe further research using the instrument (Fahy, 2002a, 2002b, 2005; Fahy & Alley, 2005) evidence of other researchers adopting the instrument could not be found.

Like the knowledge construction model (Gunawardena et al., 1997), the social networking model (Fahy et al., 2001) concentrates on constructivist principles of shared knowledge construction and relies on the interaction between the discussion forum participants. While the similarities to the cognitive domain of Bloom's taxonomy are less obvious than the previous models, the social networking model (Fahy et al., 2001) concentrates on the analysis, synthesis and evaluation levels of Bloom's taxonomy. Like the group learning, deep learning and critical thinking model (Newman et al., 1995), the social networking model (Fahy et al., 2001) uses categories which are poorly defined and have not had the benefit of further refinement by subsequent researchers. The poorly defined categories make the model a poor candidate for automated classification. Since the social networking model (Fahy et al., 2001) relies on participant interaction, the model would not be suitable for identifying evidence of critical thinking of individual participants.

#### 2.4.6 Community of inquiry model

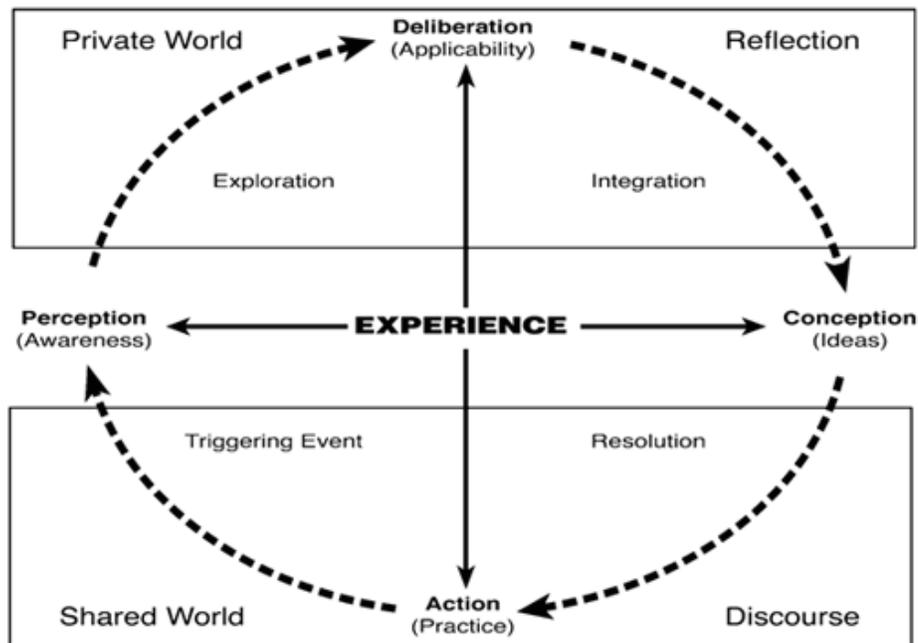
One of the most widely accepted models, which was specifically designed to guide the use of computer conferencing to support critical thinking in higher education is the Garrison et al. (2000) Community of Inquiry (COI) model. The model (see Figure 2.1) attempts to identify the elements that are crucial for a successful higher educational experience. In the model, the authors propose that a worthwhile educational experience is composed of teachers and learners and that learning occurs within a community through the interaction of three core components: cognitive presence, teaching presence, and social presence (Rourke et al., 2001).



**Figure 2.1: Community of Inquiry Model**

(from Garrison, Anderson & Archer, 2000, p. 88)

Cognitive presence is defined by Garrison et al. (2000, p. 89) as “the extent to which participants of any particular configuration of a community of inquiry are able to construct meaning through sustained communication.” Teaching presence includes facilitating the construction of learning, providing guidance relating to subject materials. Social presence is defined as the “ability of participants in a community of inquiry to project themselves socially and emotionally, as real people (i.e. their full personality), through the medium of communication being used” (Garrison et al., 2000, p. 89). The authors explain that the element in their model which is most basic to success in higher education is cognitive presence. They also suggest that "cognitive presence is a vital element in critical thinking, a process and outcome that is frequently presented as the ostensible goal of all higher education" (p. 89). Garrison et al. (2000) assert that critical thinking takes place in the element of cognitive presence and is operationalised through the Practical Inquiry model (see Figure 2.2).



**Figure 2.2: Practical Inquiry Model**

(from Garrison, Anderson & Archer, 2000, p. 104)

The Practical Inquiry model has its roots in the concepts of practical inquiry first introduced by John Dewey (1933) who suggested that practical inquiry is grounded in experience but also includes imagination and reflection leading back to experience and practice. The practical inquiry model defines four phases essential to describing and understanding cognitive presence in an educational context (Garrison et al., 2000).

The four phases are summarised below:

- Phase 1 reflects the initial phase of critical inquiry and is considered the triggering event.
- Phase 2 is exploration where learners are required to perceive or grasp the nature of a problem, and then move to a fuller exploration of relevant information.

- Phase 3, the integration phase is characterized by constructing meaning from the ideas generated in the exploratory phase.
- Phase 4 is the resolution phase which is the process of critically assessing the concepts and deductively testing their validity.

Using the practical inquiry model, Garrison et al. (2000) described how messages contained in discussion forum transcripts were coded into one of the four categories: a triggering event, an exploration event, an integration event, or a resolution event. Two empirical studies were conducted involving eleven students, two student moderators and one instructor. The levels of inter-rater reliability reported were between 45% and 84% using Holsti's CR and between 35% and 74 % using Cohen's kappa.

Following its introduction in 2000, the COI model became widely accepted and several researchers adopted the model, conducting studies aimed at measuring evidence of critical thinking among discussion forum participants (Kanuka & Garrison, 2004; Vaughan & Garrison, 2005; Garrison & Cleveland-Innes, 2005; Schrire, 2006; Kanuka, Rourke & Laflamme, 2007; Stein, Wanstreet, Glazer, Engle, Harris, Johnston, Simons, &, Trinko, 2007; Shea & Bidjerano, 2009a, 2009b; DeLeng, Dolmans, Jöbssis, Muijtjens, & van der Vleuten, 2009; Archibald, 2010; Akyol & Garrison, 2011). The studies that have been published describing the use of the COI model to measure evidence of critical thinking have all suggested that the model provides an effective tool to measure evidence of critical thinking of the participants of a discussion forum when considered as a group. There is no evidence to suggest that the model has flaws, however none of the published studies have made any effort to empirically validate the model by testing against other known tests of critical thinking.

Garrison (2003) describes the COI model and expanded the cognitive presence component of the model. The cognitive presence component is described as being

concerned with the processes of reflection, discourse and the initiation, construction and confirmation of meaningful learning outcomes. He explains that in an asynchronous online learning environment, reflection and collaboration shape cognitive presence. To participate in a discussion forum a participant has to use written communication, and participants should be encouraged to revise and refine their comments when making contributions. The forum facilitator or teacher needs to model reflective practice, providing concrete examples of how to approach subject matter in a way that constructs personal learning. The Garrison (2003) paper adds to the findings of Newman et al. (1995) who when comparing online facilitation with face to face facilitation suggests that online students are more convergent than their face-to-face counterparts and exhibit a more serious approach to discussions. In contrast, the face-to-face students are described as being more creative and produce a higher volume of interaction.

In 2003, Katrina Meyer conducted a study comparing the experiences of students involved in a face-to-face (in class) discussions with those involved in a threaded discussion, she also evaluated the threaded discussions looking for evidence of higher-order thinking (Meyer, 2003). Using the COI model she coded the postings of twenty-two students participating in a discussion forum and compared the results with those described by Garrison et al. (2000). In 2004, Meyer reported on a study that compared the COI model to a model based on Bloom's taxonomy of educational objectives. In her 2004 study, Meyer found that both models indicated similar level of higher order thinking and both identified evidence of a significant amount of higher order thinking. The COI model indicated more than 52% of messages classified in the top two layers while the Bloom's taxonomy model indicated 54% in the levels four to six of the taxonomy. Meyer's studies add credibility to the COI model and her comparisons with the Bloom's taxonomy model help validate the COI model. When concluding her study, Meyer (2003) expressed a couple of concerns. The first concern related to the need to be mindful that the COI model focuses on the ebb and flow of discussion postings between participants and as such it provides a measure of critical thinking as a group effort. This suggests that individual postings should not be considered as a reflection of the individual students level of thought. Her second

concern relates to the constricting nature of the model being used. Whatever the model being used, Meyer warns that the postings and analyses may come to relate more to the model and what it measures rather than to the actual thought expressed.

### 2.4.7 Model Comparisons

Table 2.7 summarises the features of each of the six models described in the previous section and identifies the units of analysis for each. The table also indicates the correlation tests the and identifies subsequent studies that have been conducted.

**Table 2.7**  
**Models used to measure critical thinking in discussion forums**

<b>Model</b>	<b>Categories</b>	<b>Unit of analysis</b>	<b>Inter-rater reliability</b>	<b>Subsequent studies</b>
Henri (1992)	Elementary clarification In-depth clarification Inference Judgment Strategies	Unit of meaning	Not reported	Hara, Bonk, & Angeli (2000) McKenzie & Murphy (2000)
Newman, Web, & Cochrane (1995)	Relevance Importance Novelty Outside knowledge Ambiguities Linking ideas Justification Critical assessment Practical utility Width of understanding.	Thematic approach	Percentage agreement	No evidence of further use

Bullen (1997)	Clarification Assessing evidence Making and judging inferences Using appropriate strategies and tactics	Thematic approach	Three coders report 17% agreement	Bullen (1998)
Gunawardena, Lowe, & Anderson (1997)	Sharing and comparing of information Exploration of inconsistency among ideas, concepts, or statements Co-construction of knowledge Testing and modification of proposed synthesis against existing cognitive schema Referencing statements of agreement and revealing new knowledge construction	Unit of meaning	Not reported	Veerman & Veldhuis-Diermanse (2001) Marra, Moore & Klimczak (2004) Schellens & Valcke (2005)
Fahy, Crawford & Ally (2001)	Questioning Statements and supports Reflections Scaffolding and engaging References	Sentence	Reliability from 60% - 80% agreement	Fahy (2002a, 2002b, 2004) Fahy & Alley (2005)

Garrison, Anderson and Archer's (2000) Community of Inquiry (COI) model	Triggering	Messages	Holst's CR	Kanuka & Garrison (2004)
	Explanation		45% - 84%	
	Integration	Cohen's kappa	35% - 74%	Vaughan & Garrison (2005)
	Resolution			Garrison & Cleveland- Innes (2005)
				Schrire (2006)
				Kanuka et al. (2007)
				Steine et al. (2007)
				Shea & Bidjerano (2009a).

Two of the models measured only critical thinking (Bullen, 1997; Newman et al., 1995), the remaining four models considered a number of different variables including: teaching presence, social presence, cognitive presence, deep and surface learning and group interactions. The most complex model was that developed by Newman et al. (1995), which had ten different levels of critical thinking. It is also interesting to note that this was the only model that provided no evidence of being adopted for further studies. Four of the six studies described measures of correlation between coders and only three provided a numerical indicator of the level of agreement. In terms of the number of subsequent studies, the Garrison et al.(2000) COI model is the most popular model, probably due in part to its simplicity and the level of detail provided by the authors in publications describing the model and its use. When considering which model would be best suited to automated analysis, the COI model was the only model that incorporated a relatively simple classification scheme, that was well defined and had been tested by a number of different researchers over a period of several years. It should be noted however, that the while the COI model has a good track record of measuring evidence of critical thinking, the measurement it provides is a measure of the combined critical thinking effort of the group of participants and that doubts have been expressed by Meyer (2003) of its ability to measure an individual participants critical thinking abilities.

The following section provides an overview of the findings of researchers who used the COI model. The approaches taken range from using the model to promote discussion on how to build a community of learning to using the model to validate its effectiveness as an indicator of critical thinking.

#### 2.4.8 Adoption of the COI model

A study aimed at advancing the understanding of how to facilitate higher levels of learning in an asynchronous text-based Internet communications technology was conducted by Kanuka and Garrison (2004). The study provided guidance on how to create an environment that promotes critical thinking within the online discussion. The study used the focus group discussion to determine if it was possible to use discussion forums to create a community of inquiry and promote higher levels of learning. The focus group, which consisted of ten academics from a prominent research university in Canada, based its discussion on the COI model and made a number of conclusions regarding how higher levels of learning could be promoted within a discussion forum. The group concluded that the group facilitator must take an active role in guiding discussion and that participants should be encouraged to engage in collaborative activities that promoted critical thinking. The group also concluded that participants should be asked to reflect on what they and fellow participants post to the forum and that they should be encouraged to demonstrate knowledge construction.

In 2005, a study compared measurements of critical thinking in an online environment with those in an equivalent face-to-face environment. The Vaughan and Garrison (2005) study attempted to understand how a blended learning approach can support cognitive presence. This study suggests that evidence of critical thinking can be identified in both face to face environment and online discussion environments using the COI model and stress the importance of facilitator presence in the both environments to ensure that participants engage in the resolution/application phase. Using the COI model they compared a number of face to face discussions with a

similar number of online discussions. The study involved recording a number of face to face discussions and obtaining the transcripts obtained from a number of discussion forum exercises involving the same group of participants. Both the recorded face to face discussions and the transcripts of the online discussions were coded against the COI coding scheme to identify evidence of cognitive presence. The unit of coding was the message and the coding was performed by two trained graduate students. Inter-rater reliability figures of 0.84 using Cohen's kappa and 0.86 using Holsti's CR were obtained for the online discussions and 0.92 (Cohen's) and 0.92 (Holsti's) were obtained for the face to face discussion. The results indicated that the exploration phase was dominant in both the face to face and the online environments and that triggering was slightly higher in the face to face environment. The results also indicated that the resolution/application phase was almost non-existent in both environments and the integration phase was more prominent in the online environment. Vaughan and Garrison (2005) suggested that teaching presence in the face-to-face environment promoted higher levels of critical thinking and that an increased emphasis should be placed on teaching presence within a blended learning environment to ensure that participants achieve resolution in the inquiry cycle.

A study aimed at assessing the depth of online learning occurring within a series of discussion forum activities was reported by Garrison and Cleveland-Innes (2005). The results indicated that the way students approached their study was strongly influenced by the design and teaching approach; that teaching presence contributed to the adoption of a deep approach to learning; that interaction by itself did not promote a deep approach to learning. The study concentrated on the relationship between participant interaction and critical discourse and the importance of both in maintaining a community of inquiry. Seventy-five online course participants completed a questionnaire before and after participating in a series of online discussions to measure changes in how they chose to strategise their learning in a particular learning setting.

Schrire (2006) conducted a study aimed at measuring evidence of cognitive presence in a series of online discussions. In comparison with the other studies that have used

the COI model, this was the first to compare a number of different models and its findings provide evidence that helps validate the COI model. The study compared the measurements obtained using the COI model with a model based on Bloom's taxonomy and a model based on the SOLO taxonomy proposed by Biggs and Collis (1982). The study used the discussion forum transcripts of three discussion forums involving doctoral level students at an American university. Coding was undertaken by a minimum of two coders and inter-rater reliability figures ranging from 0.55 to 0.74 using Holsti's CR were reported. All three models indicated similar levels of higher order thinking, however the author indicated that no single theory or definition of instrument could satisfactorily represent the complexity of cognitive presence.

A study which considered the influence of five groups of communication activities on the quality of students' contributions to online discussion was conducted by Kanuka et al. (2007). The activities were the nominal group technique, debate, invited expert, WebQuest and reflective deliberation. The study identified evidence of critical thinking for each of the five activities using the quantitative content analysis and the COI model. The postings of nineteen students in an undergraduate university course were coded against the COI model and the results of using each of the five instructional activities were compared. The unit of coding was the message; two coders reported an aggregate inter-rater reliability of 0.57 using Cohen's kappa. The study assumed that the COI model was able to provide an good indication of cognitive presence and reported that out of the five instructional activities considered, the WebQuest and debate activities provided evidence of the highest levels of critical thinking.

The COI model was also used to investigate the process by which shared understanding develops in chat and discussion forum learning spaces in a study conducted by Stein et al. (2007). The study suggests that teaching and social presence contributes to cognitive presence and that cognitive presence changes over time. The study also looked at how the pattern of conversation in synchronous discussion supports cognitive presence and how cognitive presence changes over time. The unit

of coding was the message and three coders coded the transcripts from a group of five participants. The inter rater reliability was calculated using Krippendorff's (2004) alpha ( $\alpha$ ) and figures of between 0.83 and 0.89 for the chat sessions and 0.67 and 0.96 for the discussion forums were obtained.

A study involving 5,000 online learners and aimed at validating the COI framework was conducted by Shea & Bidjerano (2009a). The study presents some interesting findings regarding the impact of teaching and social presence on cognitive presence and confirms that the COI model is a useful tool for investigating evidence of cognitive presence in discussion forum environments. The study investigated the relationship between all three core components of the COI model, with an emphasis on the cognitive presence component. The instrument used in the study was student survey based on the COI model, which asked online learners a series of questions related to teaching presence, social presence and cognitive presence. The survey was distributed to all students participating in a multi-institutional state-wide educational network in the United States and more than 5,000 responses were received. The responses were analysed using cluster analysis.

Most of the studies that have used the COI model as the basis for further research all confirm that the model is a useful tool for identifying evidence of critical thinking. There is variation in the individual aims of the studies and not all of them involve using quantitative content analysis to identify evidence of different levels of critical thinking. Of those studies that have attempted to measure evidence of critical thinking, the sample sizes of the studies that have adopted QCA vary from five to twenty-two participants and none of the studies measured the critical thinking skills of individuals. The only studies that made an attempt to validate the COI model by comparing it with another similar model were the studies conducted by Meyer (2004) and Schrire (2006). No studies reported the use of computers to automate the process of identifying evidence of critical thinking.

## 2.5 Computerised Content Analysis

Wegerif and Mercer (1997) argue that there are great benefits in incorporating computer based text analysis into methods for researching educational activity. Stubbs (1996) explain that computer-based text analysis is one of the fastest growing areas of linguists, with some linguists describing it as a revolution (Baker, Francis, & Tognin-Bonelli,1993). When discussing the advantages of using computers for text analysis, Wegerif and Mercer (1997) suggest that computers allow large texts to be searched quickly, they also suggest that computers are capable of complex searches and that text analysis systems can present results in variety of ways. Other advantages of using computers for text analysis include the ease in which data can be managed and recoded and the consistency in the way that coding rules can be applied and documented. While the use of computers for text analysis has proven popular over the last decade, it is important to note that researchers have expressed concerns about the risks associated with computerised text analysis. Morris (1994) explains that automated text analysis systems may fail to recognise the communicative intent of word usage, she also points out that an automated text analysis system is only as good as the coding system and classification rules under which the system operates and that automated systems might result in transforming meaningful data into numbers that are meaningless.

Even though discussion forums produce transcripts in a machine-readable format, there is little evidence of using computers to assist with the analysis of discussion forum transcripts (McKlin et al., 2002). While the literature suggests that researchers analysing discussion forum transcripts have been slow to adopt computer technology, there are numerous examples of research involving Computer Assisted Qualitative Data Analysis (CAQDAS). The term CAQDAS was coined by Ray Lee and Nigel Fielding in 1989 and refers to computerised tools that take a qualitative approach to qualitative data (Lewins & Silver, 2007). Included within the qualitative data are text, visual and multimedia forms of non-numeric information. The qualitative approach usually involves the interpretation of data through the identification and coding of

themes, concepts and processes as a way of building explanations or testing theories. Qualitative data collection techniques include in-depth interviews, focus groups and participant observation. The qualitative analysis techniques include conversation and narrative analysis, which differ from the more quantitative approach adopted by those involved in content analysis where the emphasis is on statistics of word and phrase frequencies and the occurrence of words and phrases relative to other words and phrases.

### 2.5.1 Data Analysis Software

There are a number of well known software tools that can be used to assist in the task of analysing data. Rourke et al. (2001) identified a number of the more commonly used data analysis tools, which included Wordnet, WordStat, NUD\*IST, HyperQual and General Inquirer. Lewins & Silver (2007) provide an in-depth analysis of the different computerised tools available for the qualitative and quantitative analysis of qualitative texts and provide advice on how to select the best tool for a number of different research situations. The tools reviewed by Lewins & Silver include ATLAS.ti, MAXqda, QSR NVivo, HyperRESEARCH, QDA Miner, QUALRUS, and TRANSANA.

While there are numerous examples of researchers using qualitative analysis tools like NVivo and ATLAS.ti to undertake studies involving the analysis of qualitative data sources (such as interview transcripts and newspaper articles) there is little evidence to suggest that CAQDAS are being used to assist in the analysis of discussion forum transcripts. References to studies involving CAQDAS and discussion forums include Stacey & Gerbric (2003), Gilbert & Dabbagh (2004), Williams & Lahman (2009), Ozkan (2004), Guevarra (2009), and Fujita & Teplovs (2009). All the studies involved using NVivo to categorise participant postings according to an analytical framework; only one of the studies investigated aspects of critical thinking.

## 2.5.2 Custom Built Analysis Software

Manually coding transcripts is by far the most popular technique for researchers attempting to identify evidence of critical thinking in discussion forum transcripts. A handful of researchers have used NVivo as a tool and to aid the coding process and even fewer have developed custom built tools. There are a small number of examples of researchers developing customised solutions to automate the coding and analysis process; these include McKlin et al. (2002), Brook Wu and Chen (2005), and Dönmez, Rosé, Stegmann, Weinberger and Fischer (2005). McKlin et al. (2002) described an automated tool that used neural network software to categorize messages from a discussion forum transcript. Brook Wu and Chen (2005) used keyword contribution mining to measure the quality of student work. Dönmez et al. (2005) developed a system based on the Minorthird text-learning toolkit (Cohen, 2004), which contains a large collection of configurable machine learning algorithms, any of which can be used to classify text strings.

### 2.5.2.1 An automated discussion forum analyser

Recognising the time saving benefits that could be gained by automating the transcript analysis process and wanting to attempt to reduce the subjectivity of human coding classification, McKlin et al. (2002) constructed a discussion list analysis tool that categorised online discussion lists messages into levels of cognitive effort. The purpose of the tool was to provide feedback to instructors who facilitate on-line learning, researchers studying computer-supported collaborative learning and administrators interested in correlating objective measures of the students' cognitive effort with other measures of success.

The tool which utilised artificial neural network (ANN) software to analyse discussion forum transcripts, was described in a PhD thesis submitted by Thomas McKlin at Georgia State University in 2004 (McKlin, 2004). The ANN software employed content analysis methods to categorise messages obtained from a discussion forum with 162 participants conducted as part of Georgia State University

distance education programme. The messages were classified according to the four categories of the Garrison et al. (2000) COI model.

Like the Garrison et al. (2000) study, McKlin (2002) used the message as the unit of analysis and compared the results obtained by six human coders with the results produced by the ANN software. McKlin conducted two sets of experiments and using Cohen's kappa was able to achieve reliability values of between 0.504 and 0.747 for the human coders in the first experiment and reliability values of between 0.816 and 0.879 for the human coders in the second experiment. In the first experiment he was able to achieve a reliability value of 0.519 for the ANN software when compared with the average of the human coder scores and 0.70 for the second experiment.

McKlin (2002) concluded that the ANN software could classify text with near human accuracy. McKlin also suggested that his study had demonstrated that ANN software could perform a task that had traditionally been reserved for humans alone. When discussing limitations of the study, McKlin (2002) pointed out that the tool was not tested across multiple courses and as a result its application could not be generalised across different learning domains. He also suggested that the tool should provide a way of allowing an instructor to train the model for their own context. Although his work is well cited within the literature as being an example of using computers for text analysis, there is no evidence to suggest that anyone adopted the tool developed by McKlin and no evidence that of further studies being conducted by McKlin.

While McKlin (2002) claims to have been able to build an automated system to duplicate the categorisation of critical thinking abilities using the COI model, the levels of reliability reported between the automated system and the human coders averaged little more than 0.6. McKlin (2002) reported that he adopted the message as his unit of analysis, however he did not indicate how this was implemented within his automated system. The automated system built and tested by McKlin used ANN software to analyse the discussion forum transcripts. The details describing the

algorithms used and how they were implemented within the automated system were not sufficient to enable the system to be duplicated.

### 2.5.2.2 An automated text processor

In an article in the Journal of Asynchronous Learning Networks in 2005, Yi-fang Brook Wu and Xin Chen described how they developed an automated system which used automated text processing techniques to predict the class performance of students. Brook Wu and Chen (2005) describe how they used computers to automatically grade assessments and how correlations of between 0.4 and 0.9 have been achieved when comparing human assigned and computer generated grades. They point out that, while the automated grading systems had been used to successfully grade the quality of single assessment items, such a system could be redesigned to cater for ongoing assessment items such as discussion forums. Brook Wu and Chen's (2005) model used three measures: keyword contribution, message length and message count, to produce the final score called a *performance indicator*. The results obtained using the system were encouraging, however they did point out that they did not expect the computer to entirely replace human instructors in evaluating student class performance.

The system described by Wu and Chen (2005) was not used to measure critical thinking and relied on primitive measures such as message length and message count to grade an assignment. While correlations as high as 0.9 were reported between the human grader and the automated grader, detailed figures for reliability measures were not reported. As with the previous system built by McKlin (2002), the details of the algorithms used in the automated system were not sufficiently detailed to enable the system to be duplicated.

### 2.5.2.3 Automated collaborative learning analyser

Using the MinorThird text categorisation tools developed by William Cohen at Carnegie Mellon University, Dönmez et.al. (2005) built a tool designed for streamlining the process of multi-dimensional analysis of the collaborative learning process. The tool referred to as TagHelper, used a number of algorithms to classify text into a multi-level categorisation schema. A dictionary for each category of the schema was built based on previously hand-coded transcripts. The tool then compared each new posting with those stored in the dictionary, allocating the posting to the category which had the best match. Once the computer analysis was completed, human coders could look at each posting and manually code each posting, allowing a coefficient of reliability to be calculated.

Dönmez et al. reported that they evaluated the tool using two different algorithms (non-binary classification and cascaded binary classification) and achieved an acceptable level of agreement (Cohen's kappa of 0.7 or more) with 6 out of 7 of the dimensions of the schema.

Like the two systems described in the previous sections, the tool described by Dönmez et.al. (2005) was able to duplicate the categorisation process associated with some measure of success. Like both the previous systems, the study did not provide sufficient detail about the algorithms used in the classification process to enable the system to be duplicated.

### 2.5.2.4 Comparing Custom Built Analysis Software

Of the three examples of researchers building automated custom built tools to measure evidence of critical thinking that could be identified, the tool built by McKlin was the only one that adopted and used the COI model. The tool built and described by McKlin was also the only example of a researcher building a tool specifically designed to measure evidence of critical thinking using the transcripts that were generated by discussion forum software. It should be noted that while

McKlin was able to demonstrate evidence of identifying levels of critical thinking among discussion forum participants, the measurements related to the critical thinking skills of the participants as a group and no attempt was made to measure the critical thinking skills of the individual participants. It should also be noted that McKlin assumed that the COI model provided an accurate assessment of the identification of evidence of critical thinking, and no attempt was made to validate the COI model by comparing it with results produced using alternate models.

## 2.6 Identifying Critical Thinking for Individual Participants

Despite widespread interest in the concept of critical thinking, few have focused on measuring the critical thinking skills of individual participants. While the models described so far provide a measure of the thinking skills of the group of participants, they consider it only as a collective. One study that describes the use of a model specifically designed to measure the thinking skills of individual participants was developed by Perkins and Murphy (2006).

### 2.6.1 Individual critical thinking measurement tool

The individual critical thinking measurement tool (Perkins and Murphy, 2006) was specifically designed to measure individual participation, and like the Garrison et al. (2000) model, it classified participant contributions against four critical thinking indicators. The indicators used were evidence of clarification, assessment, inference and strategy formulation, all of which are described in detail in Table 2.9. When designing their model, Perkins & Murphy (2006) examined a number of models which had been used by earlier researchers to measure evidence of critical thinking of discussion forum participants. They identified the critical thinking processes associated with each of the models and summarised their findings in a table. The table is reproduced as Table 2.8.

**Table 2.8****Summary of Critical Thinking Models**

<b>Authors</b>	<b>Norris &amp; Ennis (1989)</b>	<b>Henri (1992) Clulow, Clulow, &amp; Brace-Govan (2001)</b>	<b>Garrison, Anderson, &amp; Archer (2000)</b>	<b>Newman, Webb, &amp; Cochrane (1995)</b>	<b>Bullen (1997)</b>
Step 1	elementary clarification	elementary clarification	triggering events	clarification	clarification
Step 2	basic support	in-depth clarification	exploration	in-depth clarification	assessing evidence
Step 3	inference	inference	provisional	inference	making and judging inferences
Step 4	advanced clarification	judgement	resolution	judgement	using appropriate strategies and tactics
Step 5	strategies and tactics	strategies	—	strategy formation	—

Source: Perkins, C., & Murphy, E. (2006). Identifying and measuring individual engagement in critical thinking in online discussions: An exploratory case study. *Educational Technology & Society*, 9(1), 298-307.

Drawing on the classifications of the previous models, Perkins and Murphy developed a model which closely resembles the model developed by Bullen (1997) and consists of four categories. They described each of the four categories and provided indicators of behaviours that would be expected in each of the four categories. Table 2.9 shows the indicators used for each of the four categories associated with the Perkins and Murphy (2006) model.

**Table 2.9**

**Model for identifying engagement in critical thinking**

<b>CLARIFICATION</b>				
All aspects of stating, clarifying, describing (but not explaining) or defining the issue being discussed.				
Proposes an issue for debate.	Analyses, negotiates or discusses the meaning of the issue.	Identifies one or more underlying assumptions in a statement in the discussion.	Identifies relationships among the statements or assumptions.	Defines or criticises the definition of relevant terms.
<b>ASSESSMENT</b>				
Evaluating some aspect of the debate; making judgments on a situation, proposing evidence for an argument or for links with other issues.				
Provides or asks for reasons that proffered evidence is valid.	Provides or asks for reasons that proffered evidence is relevant.	Specifies assessment criteria, such as the credibility of the source.	Makes a value judgment on the assessment criteria or a situation or topic.	Gives evidence for choice of assessment criteria.
<b>INFERENCE</b>				
Showing connections among ideas; drawing appropriate conclusions by deduction or induction, generalising, explaining (but not describing), and hypothesising.				
Makes appropriate deductions.	Makes appropriate inferences.	Arrives at a conclusion.	Makes generalisations.	Deduces relationships among ideas.
<b>STRATEGIES</b>				
Proposing, discussing, or evaluating possible actions.				
Takes action.	Describes possible actions.	Evaluates possible actions.	Predicts outcomes of proposed actions.	

Source: Perkins, C., & Murphy, E. (2006). Identifying and measuring individual engagement in critical thinking in online discussions: An exploratory case study. *Educational Technology & Society*, 9(1), 298-307.

Perkins and Murphy applied their model to the transcripts of eight graduate students participating in an online education course. A thematic unit of analysis was chosen as the unit of analysis, and prior to the coding exercise the discussion forum transcript was broken into units of meaning. The researchers found that individual participants did exhibit different levels of critical thinking, and that on average 42% of postings were coded into the Clarification category, 42% in the Assessment category, 23% in the Inference category and 16% in the Strategies category. Since only one coder was involved in the classification exercise, no tests of reliability were conducted.

Perkins and Murphy (2006) suggested that the model should be tested using multiple coders and that it should be used in different courses with more participants. The same model was used in subsequent studies by Jacob and Sam (2008, 2010) and Murphy (2009). Unlike all the previous models which were designed to measure the critical thinking skills of the group of participants considered as a collective, the model developed by Perkins and Murphy (2006) was designed specifically to measure individual participation. Since it was the only model that was identified as being specifically designed to measure individual participation it appears to have the potential to be adopted within an automated system with the same design philosophy.

## 2.7 Summary

The literature indicates that critical thinking and its application to the educational setting is not new. From the days of the early Greek philosophers, who encouraged students to question and openly discuss, to today's educational curriculum designers, who seek to ensure that critical thinking skills are embedded within curricula, the development of critical thinking skills has been one of the fundamental aims of educationalists. While the definition of critical thinking has been and still is the subject of debate, there is some level of agreement on the cognitive skills involved in the critical thinking process and dispositions of critical thinkers.

There is general agreement that critical thinking skills are essential for survival in the 21st century workforce and that tertiary educational institutes have a responsibility to ensure that learners are encouraged to develop their critical thinking skills. There is evidence of disagreement on the best way to develop those skills and concerns have been expressed that traditional education is failing to ensure that graduates develop their critical thinking skills while studying. Despite the differences in opinion regarding the relative success of the differing approaches adopted for developing critical thinking skills, a number of different techniques have been developed to measure evidence of critical thinking. The techniques include commercially developed standard tests, assessment tools built around a course or educational programme, and the technique where learners are encouraged to assess their own thinking. While there are a number of different techniques, the literature suggests that there is little evidence of researchers attempting to verify that the different approaches achieve similar results.

The development of technology has led to the introduction of a number of tools which have provided educators opportunities to enhance their teaching practice. The development of the Internet has increased the quality and availability of distance education and has led to the adoption of tools like learning management systems and discussion forums. The availability of the discussion forum as tool to encourage group discussion has motivated a number of researchers to investigate how discussion forums can be used to promote critical thinking among participants. Content analysis has proven to be the most popular technique adopted by researchers attempting to identify evidence of critical thinking within discussion forums and a number of models have been developed based on the use of this technique. The literature suggests that while there are a number of different models there is very little evidence of the models being compared or research being undertaken to empirically validate their success.

The advances in computer technology and the availability of personal computers has led to the development of a number of software tools that can be used by researchers

to analyse the ever expanding collection of digitised text. A number of researchers have used software packages like NVivo to assist in the content analysis process and there is limited evidence to suggest that researchers have used commercially available software tools to analyse discussion forum transcripts for evidence of critical thinking. The literature also suggests that the number of researchers developing custom-built systems to measure critical thinking is small in number. Despite the evidence of computers being used to assist with the content analysis component of attempting to measure evidence of critical thinking in discussion forums there is no evidence of duplicated or sustained research and no evidence of any attempts to measure the critical thinking skills of individual participants.

This study attempts to build on the research conducted by the small number of researchers who have used computer technology to automate the process of identifying evidence of critical thinking among discussion forum participants. A process described by Palmquist et. al. (1997) as "being extremely time consuming, subject to error, often devoid of theoretical base, inherently reductive, and often simply consisting of word counts that disregard the context that produced the text" (p. 172). A process which is described as being difficult to automate or computerise. Rather than attempting to develop new models to measure critical thinking, the study uses and attempts to validate some of the already existing models.

The literature indicates that researchers have been attempting to measure critical thinking skills using a variety of techniques since the early 1970s. The techniques used include a range of commercially available tests, and a variety of teacher and researcher developed rubrics and tests. The ability of the commercial tests to measure critical thinking are widely accepted, although evidence of debate does exist concerning exactly what component of critical thinking is being measured. A number of researchers have reported success in using teacher and researcher designed tools, although there is insufficient evidence to suggest that the results obtained have been empirically validated.

A number of researchers have developed models to measure evidence of the critical thinking abilities of discussion forum participants, and most of them rely on the adoption of QCA. The models have been tested by the developers and in most cases validated using multiple coders. Unfortunately, there is little evidence to suggest that the models were subsequently adopted by others and re-tested to provide further validation of their critical thinking measurement abilities. It should also be noted that evidence cannot be found of researchers comparing the results of the different models to see if they come to the same conclusions about the critical thinking abilities of discussion forum participants.

McKlin (2004) provides one of the few examples of using computer technology to automate the identification of critical thinking among discussion forum participants. The McKlin study suggested that computerised systems have the potential to duplicate the findings of human coders when being used as an aide for content analysis. The system developed by McKlin was designed to measure evidence of critical thinking of the participants as a collective group using the COI model and was not evaluated by comparing the results achieved to those obtained using alternate models.

This study described in this thesis is unique in a number of ways: it is the first study to use an automated tool to measure evidence of critical thinking of discussion forum participants using more than one coding model. It is also the first documented study to use a commercial critical thinking measurement tool to validate the automated tool. This study is also unique, as it is the first study to use an automated tool to measure evidence of critical thinking of individual discussion forum participants.

### 2.7.1 Conclusion

This chapter reviewed the literature relating to critical thinking and the attempts to measure evidence of critical thinking. The aim of the chapter was to place the study in context and identify previous research relating to measurement of critical thinking

among discussion forum participants. The literature review identified that while there is some evidence of researchers attempting to measure indications of critical thinking among discussion forum participants, there is little evidence of the models being duplicated or validated. The review also identified a gap in the literature relating to research concerning the use of computers to automate the measurement process. Of the few documented studies that have attempted to measure critical thinking using computers, none has been duplicated and subsequently independently validated.

The research in this study uses models that have been developed by others and used in subsequent studies to measure evidence of critical thinking within discussion forums. The study attempts to fill the gap identified in the literature related to the use of computers for automating the process of critical thinking measurement. The study also provides the first documented attempt of using computer technology to identify evidence of critical thinking of individual participants. The next chapter examines the research approach adopted for this study.



# CHAPTER 3: RESEARCH METHODOLOGY

---

Put simply, action research is “learning by doing” - a group of people identify a problem, do something to resolve it, see how successful their efforts were, and if not satisfied, try again. (O'Brien,2001)

The previous chapter identified the importance of critical thinking within today’s society. While the need for developing critical thinking skills has gained emphasis among educators over the last twenty years, the literature suggests that there is little evidence to indicate the widespread use of automated tools to measure critical thinking.

The first chapter proposed two main research questions related to the measurement of critical thinking. The two main questions that are being addressed in this study are:

1. "Is it possible to automate the identification of critical thinking ability exhibited by a group as they participate in a discussion forum?"
2. "Is it possible to automate the identification of an individual participant’s critical thinking ability in a discussion forum?"

This chapter explains the research approaches adopted for this study to explore the central questions related to the automation of the identification of critical thinking in discussion forums. A mixed method enquiry was adopted for the study involving a qualitative action research framework combined with quantitative content analysis. This approach enabled the adoption of a dynamic and flexible process that supported the development and refinement of an artefact built to assist in the exploration of the research questions.

### 3.1 Research Methods

There are three broad research paradigms used in education, social science, behavioural science and more recently information systems. The three broad research paradigms are: quantitative research, qualitative research and mixed research. Quantitative research is research that relies primarily on the collection of quantitative data. Qualitative research is research that relies on the collection of qualitative data. Mixed research is research that involves the mixing of quantitative and qualitative methods.

#### 3.1.1 Quantitative Research

Quantitative researchers are most commonly associated with the positivist school of thought and believe that social observations should be treated as entities in much the same way that physical scientists treat physical phenomena (Johnson & Christensen, 2004). Quantitative research emphasises the measurement and analysis of relationships between variables rather than processes (Denzin & Lincoln, 2000). Quantitative researchers adopt the scientific method and focus on controlling variables, gathering measurable evidence and coming to conclusions or providing new explanations (Cormack, 1991). The focus of quantitative research is on measurement and proof; it is based on an ethos that relies on results that can be observed and counted. The main idea behind quantitative research is to be able to separate things easily so that they can be counted and modelled statistically, thereby removing factors that could be seen to distract from the intent of the research. Before they start measuring, quantitative researchers will usually have set up controls and have a very clear picture of what is being measured.

### 3.1.2 Qualitative Research

Qualitative researchers adopt a constructivist and interpretivist philosophical viewpoint, they may reject the positivist view and contend that multiple-constructed realities abound, also that context-free generalisations are neither desirable nor possible (Johnson & Christensen, 2004). Qualitative research emphasises the qualities of entities over processes and meanings that are not experimentally examined or measured in terms of quantity, amount, intensity, or frequency (Denzin & Lincoln, 2000). According to Denzin & Lincoln (2000) qualitative researchers stress the socially constructed nature of reality, the relationship between the researcher and what is studied, and the situational constraints that shape inquiry. Qualitative researchers seek answers to questions that stress how social experience is created and given meaning.

Creswell (1994) identifies five main types of qualitative research:

- Phenomenology is a form of qualitative research in which the researcher attempts to understand how one or more individuals have experienced a phenomenon.
- Ethnography relies on the researcher having a grounding in cultural anthropology and the meaning of a social-cultural system. It is the form of embedded qualitative research that focuses on describing the culture of a group of people.
- Case study research is a form of qualitative research that is focused on providing a detailed account of one or more cases.
- Grounded theory is a qualitative approach to generating and developing a theory from data that the researcher collects. From the data collected, key points are marked with a series of codes which are grouped into similar concepts from which categories are formed. The categories form the basis for the creation of the theory, or a reverse-engineered hypothesis.

- Biographical or historical research is a qualitative approach where the researcher collects extensive information about events that occurred in the past.

### 3.1.3 Mixed Method Research

Mixed method research recognizes that both quantitative and qualitative research are important and useful, and attempts to both draw from the strengths and minimize the weaknesses of both approaches (Johnson & Christensen, 2004). A mixed research design is a general type of research that includes both quantitative and qualitative research data, techniques and methods. The mixed method research design involves research that uses data involving both numbers and text. A mixed method uses both deductive and inductive scientific methods, has multiple forms of data collection and produces eclectic and pragmatic reports.

Johnson & Christensen (2004, p. 52) provide an excellent summary of the different emphases for each of the three research methods. The summary is provided in Table 3.1.

**Table 3.1****Emphases of Quantitative, Mixed, and Qualitative Research**

	<b>Quantitative Research</b>	<b>Mixed Research</b>	<b>Qualitative Research</b>
Scientific method	Deductive or “top-down”. The researcher tests hypotheses and theory with data	Deductive and inductive	Inductive or “bottom-up”. The researcher generates new hypotheses and grounded theory from data collected during fieldwork
View of human behaviour	Behaviour is regular and predictable	Behaviour is somewhat predictable	Behaviour is fluid, dynamic, situational, social, contextual, and personal
Most common research objectives	Description, explanation, and prediction	Multiple objectives	Description, exploration, and discovery
Focus	Narrow-angle lens, testing specific hypotheses	Multi lens focus	Wide-angle and “deep-angle” lens, examining the breadth and depth of phenomena to learn more about them
Nature of observation	Attempt to study behaviour under controlled conditions	Study behaviour in more than one context or condition	Study behaviour in natural environments. Study the context in which behaviour occurs
Nature of reality	Objective (different observers agree on what is observed)	Commonsense realism and pragmatic view of world (i.e., what works is what is “real” or true)	Subjective, personal, and socially constructed

Form of data collected	Collect quantitative data based on precise measurement using structured and validated data collection instruments (e.g., closed-ended items, rating scales, behavioural responses)	Multiple forms	Collect qualitative data (e.g., in-depth interviews, participant observation, field notes, and open-ended questions). The researcher is the primary data collection instrument
Nature of data	Variables	Mixture of variables, words, and images	Words, images, categories
Data analysis	Identify statistical relationships	Quantitative and qualitative	Search for patterns, themes, and holistic features
Results	Generalizable findings	Corroborated findings may generalize	Particularistic findings. Representation of insider (i.e., “emic”) viewpoint. Present multiple perspectives
Form of final report	Statistical report (e.g., with correlations, comparisons of means, and reporting of statistical significance of findings)	Eclectic and pragmatic	Narrative report with contextual description and direct quotations from research participants

Source: Johnson, B. R., & Christensen, L. B. (2004, p. 52). *Educational research: Quantitative, qualitative, and mixed approaches, (2nd ed.)*. New York: Pearson Education.

### 3.2 Action Research

Since the research describe in this thesis, is based upon the development of an artefact which is subject to an iterative process of testing, evaluation and redevelopment it is suited to an action research approach. This section describes the origins and philosophical foundations of action research. The section also defines action research and describes the different types of action research that can be used.

Since the middle of the twentieth century, action research has been recognised as a useful research methodology. It has been used in the social and medical sciences, and since the 1990s it has increased in importance for information systems (Baskerville, 1997).

### 3.2.1 History of Action Research

Most texts suggest that the term "action research" was first coined in 1946 by Kurt Lewin, a German social and experimental psychologist in a paper entitled "Action Research and Minority Problems". During the late 1930s and early 1940s Lewin developed a field-theory version of action research at the University of Michigan Research Centre for Group Dynamics in order to study social psychology. Independently and at the same time the Tavistock Clinic developed an operational research version of action research (Trist, 1976). The Tavistock Clinic used action research to study the effects of "shell shock" and its treatment through counselling. The clinic also used action research to develop an understanding of the effects of early separation of children from their parents following wartime evacuations. In 1946, Eric Twist, a British clinical psychologist and an admirer of Lewin's work joined the Tavistock Clinic and helped establish the Tavistock Institute where he was appointed as director. Shortly after the establishment of the Tavistock Institute, Twist and the team at the Tavistock Institute joined forces with Lewin and the team at the University of Michigan and launched the journal entitled *Human Relations*, bringing together the two schools of action research thought.

In 1949 Stephen Corey and others at Teachers College of Columbia University introduced the term action research to the educational community. Corey (1953) defined action research as the process through which practitioners study their own practice to solve their personal practical problems. The enthusiasm of researchers like Corey, Steinhouse, Elliot, Adelman, Schon, Carr and Kemmis for action research in

the educational arena has led to its acceptance as a mechanism for developing reflective practice among teachers.

Enid Mumford, a researcher who had worked at the Tavistock institute, was one of the first to introduce action research into the information systems field (Mumford & Weir, 1979). Mumford used action research techniques to increase participation in information systems design by developing the ETHICS methodology. Trevor Wood-Harper was also an advocate of participatory action research for information systems development when he developed a methodology called Multiview in the early 1980s (Wood-Harper, Antill, & Avison, 1985). Peter Checkland was another advocate who used action research to develop the soft systems development methodology (Checkland, 1981). Today participatory action research is an accepted practice among information systems developers and researchers.

### 3.2.2 Philosophical Foundations of Action Research

The use of the word paradigm in scientific research is often attributed to the science historian Thomas Kuhn. He adopted the word to refer to the set of practices that define a scientific discipline at any particular period of time. Over time, Kuhn refined his definition of the word to include the terms exemplar and normal science. In his book "The Structure of Scientific Revolutions", Kuhn (1962: p. 15) suggests that paradigms help scientific communities to confine their discipline in that they help the scientist to:

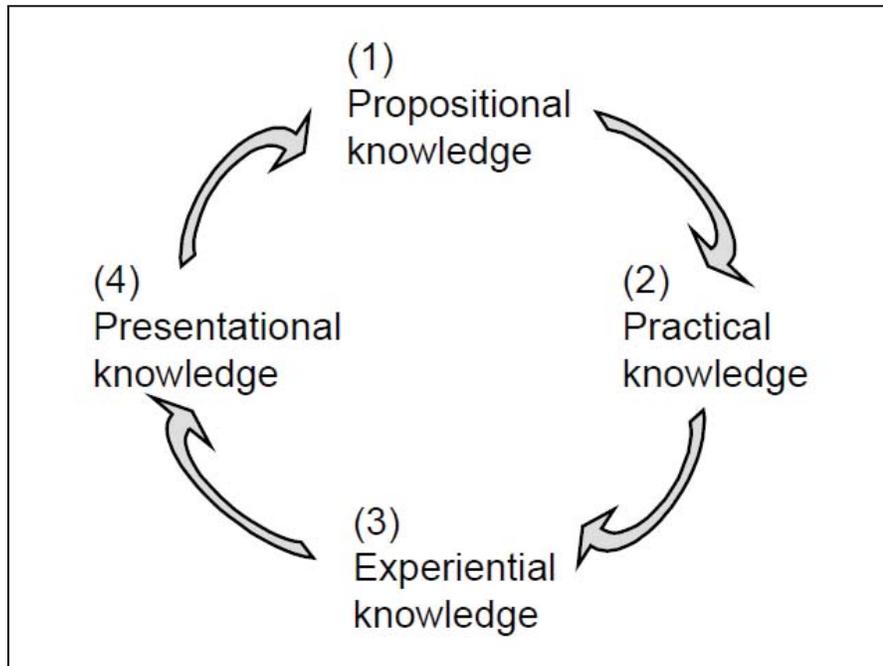
- create avenues of inquiry;
- formulate questions;
- select methods with which to examine questions;
- define areas of relevance.

The main research paradigm for the past several centuries has been that of Positivism. Positivism is a philosophy of science which holds that authentic knowledge is knowledge that is based on sense experience and positive verification. While the

positivist approach has been a recurrent theme in the history of western thought from the Ancient Greeks to the present day (Cohen, Manion, & Morrison, 2007) the concept was developed in the early 19th century by the philosopher and founding sociologist, Auguste Comte. During the early twentieth century, a group of philosophers referred to as the “Vienna Circle” introduced the concept of Logical Positivism. According to logical positivism, meaningful statements can be divided in two classes; analytic a priori and synthetic posteriori (Friedman, 1999). Analytic a priori contain statements that are true or false in virtue of their logical forms or in virtue of their meaning while synthetic posteriori contain statements whose truth or falsity can be ascertained only by means of the experience. Logic and mathematics belong to the class of analytic a priori statements and genuine scientific statements; those of physics, biology and psychology belong to the class of synthetic posteriori. The methods of logical positivism rely heavily on quantitative measures, with relationships among variables commonly shown by mathematical means. Positivism, used in scientific and applied research, has been considered by many to be the antithesis of the principles of action research (Susman & Evered 1978; Winter, 1989a, 1989b).

In the middle of the twentieth century, a new research paradigm emerged in the social sciences which attempted to break out of the constraints imposed by logical positivism (O'Brien, 2001). Known as the interpretive paradigm, it was developed mainly out of sociological inquiry, where researchers use qualitative methodological approaches such as ethnography, phenomenography and hermeneutics in order to interpret or develop understanding of participants' viewpoints (Lincoln & Guba, 2000; Kember, 2000). While sharing a number of perspectives with the interpretive paradigm, and making considerable use of its related qualitative methodologies, there are some researchers who feel that neither it nor the positivist paradigms are sufficient epistemological structures under which to place action research (Lather 1986; Morley, 1991).

In recent years a participatory paradigm has emerged. The participatory paradigm is an epistemological and ontological principle which has been adopted for participatory forms of inquiry (Heron & Reason, 1997). The participatory worldview gives primary attention to the promotion of human welfare through the conduct of socially responsible research (Maxwell & Delaney, 2004). In participatory research, the researcher is recognised as an integral part of the world he studies, co-creating with the research subjects their world and sharing the responsibility for the application of the research outcomes (Reason, 1998). According to Heron (1981) participatory inquiry is cyclic and knowledge is produced through a cycle of interaction, reflection and application.



**Figure 3.1 Research cycling and knowledge creation**

(from Breu & Peppard, 2001, p. 246)

Figure 3.1 indicates that as one progresses through the participatory research cycle four types of knowledge are produced; experiential knowledge, practical knowledge, presentational knowledge, and propositional knowledge (Heron, 1981). Through the

direct contact with the world, be it people, events, places or things, experiential knowledge is gained. Experiential knowledge is mostly tacit and difficult to surface. From the filtering of experience and its representation in stories, concepts and metaphors presentational knowledge emerges. Propositional knowledge which represents the type of knowledge we form in the conclusions from our research is expressed in the theories proposed about the world. It is “rooted in and derived from the experiential and practical knowledge of the subjects in the inquiry. Practical knowledge is ‘how- to-do’ knowledge and is expressed in skills and competencies” (Reason, 1994, p. 326). It is within the participatory research paradigm that action research best sits.

### 3.2.3 Definition of Action Research

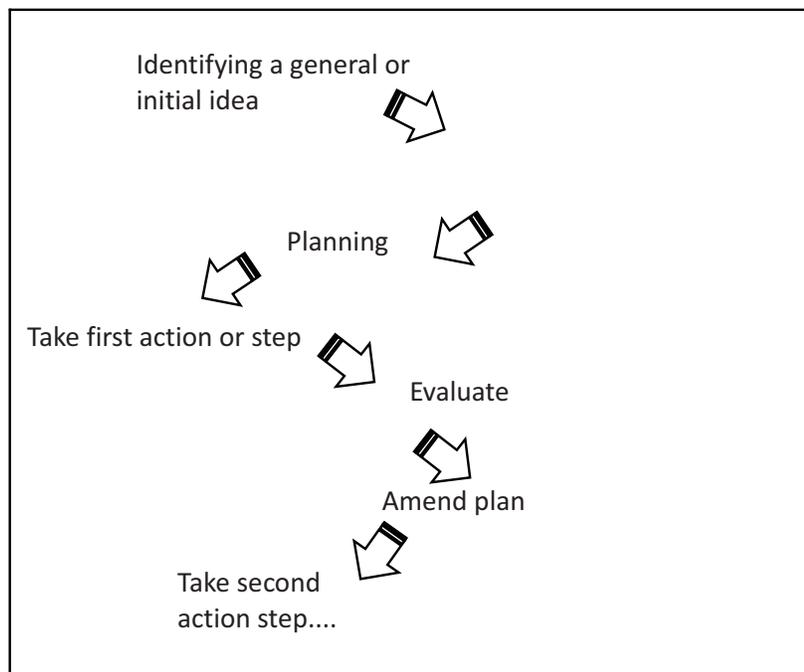
According to McCutcheon and Jurg (1990), action research is a "systemic inquiry that is collective, collaborative, self-reflective, critical and undertaken by participants in the inquiry". It is a way of learning through a set of reflective stages that helps develop and refine understanding. Tandon (1988) suggests that being a participatory form of enquiry, a key feature of action research is empowerment through participation. The practice of action research involves people reflecting on issues and processes during the research, participants as co-researchers, and entails an element of risk given that the process and outcomes are in a state of on-going change (Winter, 1989a).

While the definitions of action research within the literature differ somewhat in their basic terminology, there appears to be agreement on the main characteristics of action research. Hart and Bond (1994, p. 38) suggest that action research is:

- Educative.
- Deals with individuals as members of social groups.
- Is problem-focused, context-specific and future-orientated.
- Involves a change intervention.

- Involves a cyclic process in which research, action and evaluation are interlinked
- Aims at improvement and involvement
- Is founded on a research relationship in which those involved are participants in the change process.

The cyclic process in which research, action and process are interlinked is often referred to as the "action research spiral". Kurt Lewin (1946), first suggested that action research was an approach that involved a spiral of steps, each of which is composed of a circle of planning, action and fact-finding about the result of the action. Figure 3.2 depicts Lewin's research spiral as an iterative cycle where the researcher first identifies a topic worthy of research. The researcher then conducts a fact finding mission to clarify the topic and place it in context with previous research out of which should arise a plan of action for the first iteration of the cycle. Once the first action has been processed it is evaluated and the plan may then be amended leading on to another research cycle.



**Figure3.2: Lewin's Research Spiral**

Most action research iterates through two or more cycles. The action research process involves people reflecting on issues and processes during each cycle of the research. The participants act as co-researchers, which often creates a process where outcomes are often in a state of on-going change (Winter, 1989a). Riel (2007) suggests that over time, those involved in action research develop a deep understanding of the systems they study, developing a form of "adaptive" expertise as they work through a set of reflective stages.

### 3.2.4 Types of Action Research

A review of the literature would suggest that the classification of the various types or approaches to action research differs from author to author. Grundy (1988) discusses three modes of action research; technical, practical, and emancipatory. McCutcheon and Jurg (1990) also discuss three perspectives of action research: a positivist perspective, an interpretivist perspective, a critical science perspective. Holter and Schwartz-Barcott (1993) discuss three types of action research: that of a technical collaborative approach, a mutual collaborative approach and an enhancement approach. McKernan (1991) also lists three types of action research:

- Type 1: the scientific-technical view of problem solving;
- Type 2: practical-deliberative action research; and
- Type 3: critical-emancipatory action research.

Hughes, Ndonko, Ouedraogo, Ngum, & Popp (2004) suggest that even though the typologies are not completely consistent, there is a general correspondence. They describe three broad areas of action research: technical action research, practical action research, emancipative action research.

#### Technical action research

Technical action research includes experimental action research (Hart & Bond, 1994); it is essentially a deductive process where the goal of the research is to accumulate

predictive knowledge to refine existing theories. Projects tend to be instigated and managed by researchers who are seen as skilled experts. Technical action research uses a scientific frame of reference associated with the positivist paradigm used to refine existing knowledge in a deductive process (Hughes et al., 2004).

A technical action research project will usually be instigated by a person or group of people who would be regarded as experts or authority figures. Technical action research values traditional research methods and techniques and promotes efficient and effective practice. It promotes participation by practitioners in the process of improvement. A major thrust of technical action research is on validating and refining existing theories. In a technical action research project a researcher might use an experimental design to test a hypothesis in a real-world situation, rather than a laboratory. It is a form of applied research, and has many features in common with operations research (Harmse, Pothas, & de Wet, 2002).

#### Practical action research

The majority of action research projects are likely to be practical action research projects and are often seen in professional practice. Practical action research allows for a more flexible approach than is available in the positivist paradigm by taking a pragmatic approach to solving practical problems (Hughes et al., 2004). Practical action research is used to improve practice by developing and applying the personal knowledge and wisdom of the participants (Grundy, 1988). It adopts a more collaborative approach to problem solving with a set of people seen as owning a problem or a situation requiring improvement. It is commonly used by practitioners wanting to gain an understanding of their own practice and over time the practitioners gain skills, relying less on specialist researchers. It is popular among teachers who use practical action research to reflect upon their teaching practices.

### Emancipative action research

Emancipative action research is often used to address issues of social change and emancipation and employs a critical perspective (Hughes et al., 2004). It is referred to as emancipative because it promotes a critical consciousness and empowerment, which leads to political as well as practical action to promote change. Emancipative action research is said to be fully participative, because decision making about the action research project and the way it is conducted is made by all those who are closely affected by the action research and its outcomes. Emancipative action research seeks to liberate participants from conditions, which they experience as oppressive. Participation in decision making, raising consciousness and autonomous action are keys to this approach.

### 3.3 Content Analysis

White and Marsh (2006) describe content analysis as a highly flexible research method that has been widely used in library and information science (LIS) studies with varying research goals and objectives. Content analysis is used in qualitative, quantitative, and sometimes mixed modes of research frameworks and employs a wide range of analytical techniques.

Since content analysis can be applied to examine any piece of writing or occurrence of recorded communication it has been applied in a wide number of fields, ranging from marketing and media studies, to literature and rhetoric, ethnography and cultural studies, gender and age issues, sociology and political science, psychology and cognitive science, and many other fields of inquiry (Palmquist et al., 1997). Content analysis reflects a close relationship with socio- and psycholinguistics, and is playing an integral role in the development of artificial intelligence. Berelsen (1952) describes some of the possible uses of content analysis:

- Reveal international differences in communication content
- Detect the existence of propaganda

- Identify the intentions, focus or communication trends of an individual, group or institution
- Describe attitudinal and behavioural responses to communications
- Determine psychological or emotional state of persons or groups

Palmquist et al. (1997) suggest that content analysis offers several advantages to researchers who may consider using it. The advantages of using content analysis include:

- it looks directly at communication using transcripts and as a result gets at the central aspect of social interaction,
- it can allow for both quantitative and qualitative operations,
- it allows text to be classified and coded into specific categories,
- it can be used to interpret texts and can be applied to the development of expert systems,
- it is an unobtrusive means of analysing interactions,
- when done well it is considered a relatively "exact" research method.

While there are advantages to using content analysis in research, Palmquist et al. (1997) also suggests that there are disadvantages. These disadvantages include:

- it can be extremely time consuming,
- when used to attain high levels of interpretation it is subject to error,
- it is often said to be devoid of theoretical base,
- when dealing with complex tasks it is inherently reductive,
- it often disregards the context that produced the text,
- and it can be difficult to automate or computerise.

Despite the difficulties associated with content analysis, it has been adopted by a growing number of researchers. Neuendorf (2002) describes the expanding use of content analysis that is evident when searching scholarly articles and journals, and

she also points out that there is growing evidence of the use of content analysis as a technique for graduate theses and dissertations.

### 3.4 The Research Framework

An action research approach was adopted as the overall framework for this study which aims to explore the central questions related to the automation of the identification of critical thinking in discussion forums. Action research provides a dynamic and flexible process that is well suited to support the development and refinement of the computerised system built to assist in the exploration of the research questions. Among information system developers, action research is a popular methodology. Wood-Harper (1985), Checkland (1981) and Mumford and Weir (1979) all support the use of action research for activities that involve the development of information systems.

The action research framework approach also allowed the adoption of a quantitative content analysis research approach within each of the action research cycles. The quantitative content analysis approach was adopted within each of the action research cycles to validate the models used to measure evidence of critical thinking. Quantitative content analysis was also used to validate the results obtained using the automated tool.

The research path taken for this study involved a number of cycles, starting with a literature review to identify the current state of research in the area of measuring critical thinking in discussion forum transcripts. Having established that there appeared to be little evidence of research into the area of using automated tools in the identification of critical thinking, a research plan was developed in an attempt to answer the research questions associated with this study.

Once the literature review had been conducted, the gaps in the literature identified and the research questions formulated, a model which was designed around the use of

content analysis was identified and subsequently used to test the operation of the computerised tool. The testing of the model, the building and testing of the tool, the identification of a model suitable for measuring evidence of critical thinking for individual participants and the validation of the individual participation model were the major activities in the action research cycles.

Since an essential component of the action research process involves people reflecting on issues and processes during each action research cycle, a peer review process was built into each cycle of the research framework. Peer review is a system whereby research is scrutinised by independent experts (peers). It is said to inform decision making and assess the quality of research (Kassirer & Campion, 1994). The peer review process associated with this study involved presenting the findings of each of the action research cycles both within New Zealand and overseas.

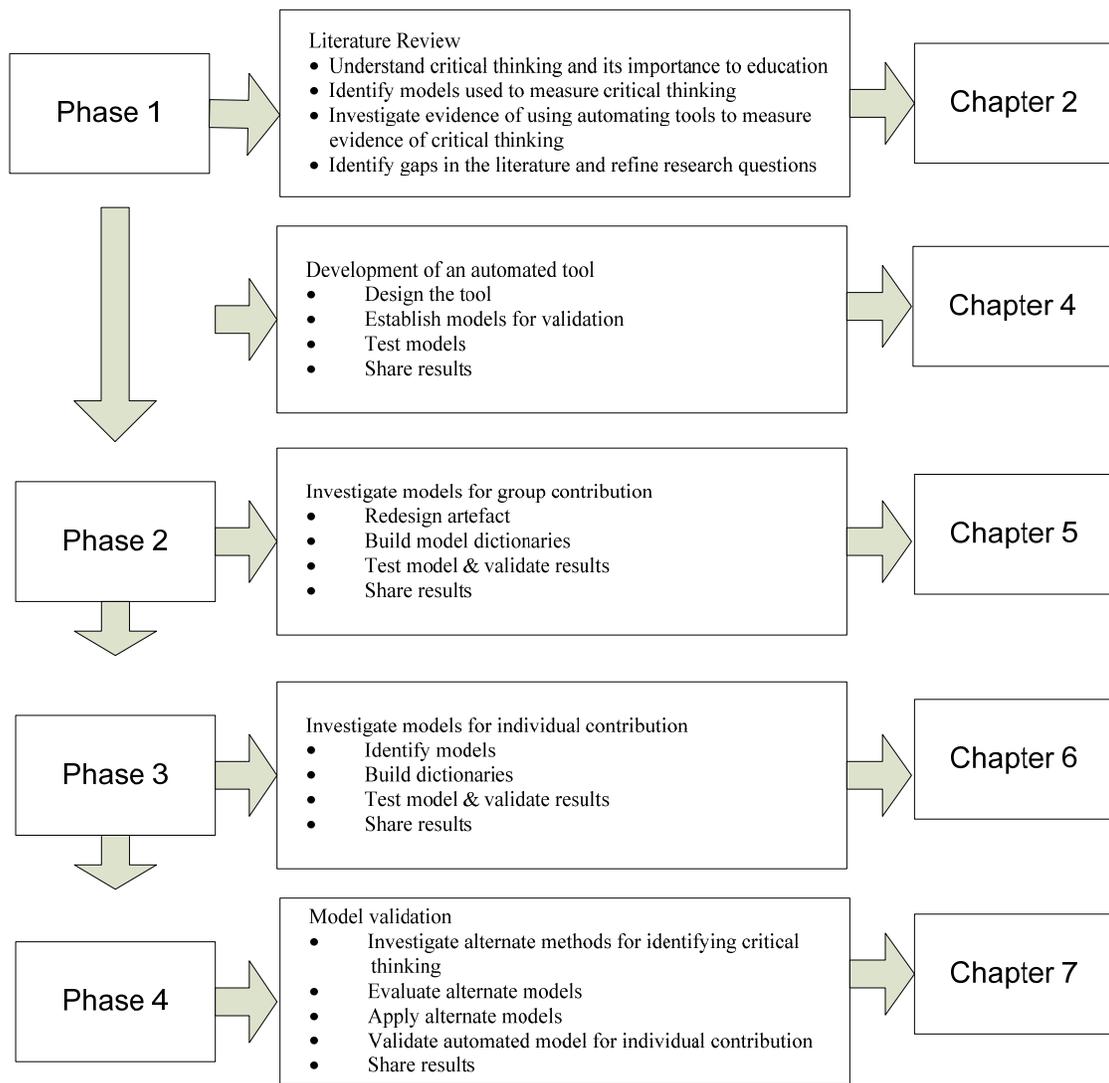
The presentations within New Zealand were made at a series of research symposia held at the Advanced Learning Technologies Research Centre (ALTRC) at Massey University in Palmerston North. The symposia were initiated to provide an opportunity for post-graduate students to share their research findings and seek advice from their peers and research supervisors. At the completion of the action step of each of the research cycles a PowerPoint presentation summarising the work to date was prepared and presented to the symposia attendees. Prior to the presentation attendees were informed that their feedback was an essential component of the research activity and that it would be used to inform the direction of future research activity. Attendees were also informed that their participation was voluntary and that the presentation and audience responses would be recorded and that a written transcript would be produced after the symposia presentation. The written transcript was used to identify common questions and suggestions which could inform future iterations of the action research cycles.

The international presentations were made at a number of different international conferences. Following each of the ALTRC symposia presentations a paper was

written describing the current state of the research. The paper was then submitted for presentation at an international conference. The international conferences were chosen from a list of suitable conferences which accepted peer reviewed papers and had a conference theme which supported presentations related to critical thinking and learning technologies. All the papers that were submitted and consequently accepted were subject to a double-blind review process and the comments made by reviewers were noted and saved electronically. Prior to the paper presentation the researcher informed the audience that the research being presented formed part of a study that was being conducted as partial fulfilment of the requirements for the degree of Doctor of Philosophy in Information Systems at Massey University, Palmerston North, New Zealand. The researcher also informed the audience that any feedback would be recorded and a written transcript produced that would be used to inform the future directions of the study.

The written transcripts resulting from the post-graduate symposia, the conference review panels and the conference presentations were analysed to identify common themes and questions. The themes and questions were then used to help inform the decision-making processes that lead to the subsequent activity component of the following action research cycle.

Figure 3.3 illustrates the four phase action research framework that was used to develop the computerised tool and attempt to answer the questions which are central to this study.



**Figure 3.3: Four phase action research framework for the study**

### 3.4.1 Phase 1: Literature Review

Before commencing the research project a literature review was conducted to place the planned research topic in the context of existing work on the subject. The process of starting an action research project with a literature review is common practice.

Bourner (1996) suggests a number of reasons for conducting a literature review which include:

- Identifying gaps in the literature.
- Attempting to avoid reinventing the wheel, which should reduce the chance of making the same mistakes as previous researchers.
- Building on the platform of existing knowledge and ideas.
- Identifying other people working in the same field.
- Identify seminal works about the topic.
- Providing an intellectual context for the research project.
- Discovering information and ideas that may be relevant to the project.
- Identifying research methods that could be appropriate to adopt.

Chapter 2 contained the literature review which looked at the definition and importance of critical thinking. It identified a number of tools and models used to measure critical thinking and discussed some of the models that have more commonly been used to identify evidence of critical thinking within discussion forum transcripts. The chapter concluded with a review of how automated tools have been used to identify evidence of critical thinking. The literature review did as Bourner (1996) suggested help inform the research and enabled the refinement of the research questions.

This chapter, chapter 3, describes the research framework and methodologies adopted during the study.

### 3.4.2 Phase 1: Initial development of the artefact

Informed of the gaps in current research by the literature review a concept model of a tool that could be used to assist with the automation of critical thinking in discussion forum transcripts was developed. As previously stated in this chapter, the tool was

developed as an instrument to help answer the research questions and establish if the identification of evidence of critical thinking could be automated.

The first iteration of the use of the tool allowed transcripts from discussion forums to be imported, parsed into measurable units and then manually coded, using content analysis methods, into categories based on two of the recognised models used to identify evidence of critical thinking in discussion forums. The models used were the Garrison et al. (2000) COI model and the model developed by Henri (1992) and refined by Hara et al. (2000). The results obtained for different human coders were then compared and coefficients of reliability between the different coders were automatically calculated. Chapter 4 details the research activity conducted during Phase 2 of the action research cycle.

The research to test the efficacy of the initial ACAT system was conducted in the second semester of 2003 and involved a discussion forum activity undertaken by a class of first year undergraduate degree students at a tertiary education institution in New Zealand. The class consisted of fifteen students, three females and twelve males, aged between 18 and 38, and of varying academic abilities. The class was involved in a data communications course and the transcript was coded using content analysis methods by two human coders.

Having evaluated the use of the tool, the findings were peer reviewed, locally at a meeting of the Advanced Learning Technologies Research Centre (ALTRC) at Massey University in Palmerston North and internationally at the Third Pan-Commonwealth Forum on Open Learning in Dunedin, New Zealand (Corich, Kinshuk & Hunt, 2004a). Informed by the feedback obtained as a result of the two presentations the tool was refined and Phase 2 of the research commenced.

### 3.4.3 Phase 2: Investigate models for group contribution

Having established that an automated tool could reduce the time involved in attempting to identify evidence of critical thinking in discussion forum transcripts, the tool was refined to enable the coding of the transcripts to be automated. The model adopted for testing the tool was again based on the COI model. A dictionary of terms relating to the four categories of the COI model was developed and imported transcripts were analysed by the tool using probability theory to place transcript units into each of the four categories. The results of those obtained by the tool were compared to the results obtained using human coders.

The research to test the efficacy of automated text classification component of the ACAT system was conducted in the second half of 2004 and involved the same transcript used in the first phase of the action research cycle. Once again a content analysis approach for coding was adopted, this time using the automated tool to code the discussion forum transcript.

Chapter 5 details the research activity conducted during Phase 3 of the action research cycle.

Once again the findings of the research were subjected to peer review, locally at an ALTRC meeting in Palmerston North and internationally at the Fifth International Conference of Networked Learning held at the University of Lancaster, England (Corich et al., 2006a). Paper reviewers, symposium and conference attendees acknowledged the apparent success of the tool in identifying evidence of critical thinking in the group setting and suggested that maybe the tool could be used to test evidence of critical thinking for individual participants.

### 3.4.4 Phase 3: Investigate models for individual contribution

Phase 2 established that an automated tool could be used to measure evidence of critical thinking among discussion forum participants in a group setting. Phase 3

attempted to collate the transcripts for individual participants so that the COI model could be applied to the postings of the individual participants. The findings suggested shortcomings when using a model designed to measure group contributions for individuals and a literature review was conducted to establish if evidence of research involving the identification of models for measuring critical thinking for individuals participating in discussion forums existed. As a result of a review of the literature a model developed by Perkins & Murphy (2006) to measure individual contributions was identified as having the potential to be applied to the computerised tool. The findings of the application of the COI model as it was used to measure evidence of critical thinking for individuals are detailed in Chapter 6.

The research associated with this phase of the study was conducted in the second semester of 2005 and involved a discussion forum activity undertaken by a second year undergraduate degree course at the same tertiary education institution used in the earlier research cycle. The class consisted of sixteen students, five females and eleven males, aged between 17 and 42, and of varying academic abilities who were involved in an internet and web development course. A content analysis approach was adopted to code the discussion forum transcript, with the transcript being coded by the computerised system and also by two human coders.

The findings related to the research were once again subjected to peer review. Locally the findings were reviewed at an online meeting of the ALTRC symposium and internationally at the 2007 Cognition and Exploratory Learning in the Digital Age (CELDA) Conference in the Algarve, Portugal (Corich et al., 2007a). While both groups acknowledged the apparent success of the automated tool to identify evidence of critical thinking questions were raised about the validity of the Garrison et al. (2000) COI model when being used to attempt to measure evidence of critical thinking for individual participants.

### 3.4.5 Phase 4: Model validation

The final phase in the action research cycle aimed to investigate and evaluate alternate methods for identifying critical thinking. Having identified the Perkins & Murphy (2006) model, the individual transcripts from a discussion forum were coded by a human coder using the model. The dictionary for the computerised system was modified to cater for the new model and the computerised results were compared to those produced by the human coder. To further validate the computerised tool and the effectiveness of the Perkins & Murphy (2006) model, the results using the computerised system were compared with the results obtained using a commercially available critical thinking assessment.

The discussion forum transcripts used in this study were obtained from a discussion forum used in a third year undergraduate web development course which took place in the first semester of 2007. The transcripts were obtained with ethical approval of the institute and involved sixteen students, aged between 18 and 36, and of varying academic abilities. Content analysis methods were used to code the transcript and coding was performed by the computerised system and one human coder.

The findings were presented internationally at the 2009 Cognition and Exploratory Learning in the Digital Age (CELDA) Conference in Rome, Italy (Corich et al., 2009). Attendees once again acknowledged the apparent success of the automated tool to identify evidence of critical thinking and suggested that the tool was worthy of further development to enable it to be used more widely.

Chapter 7 details the research activity conducted during Phase 4 of the action research cycle.

## 3.5 Issues of reliability and validity

The concepts of reliability and validity have their roots in the positivist or scientific paradigms of quantitative research where the world is seen as being made up of

observable or measurable facts (Glesne & Peshkin, 1992). In a positivist environment researchers employ experimental methods and quantitative measures to test hypothetical generalisations, emphasising measurement and analysis of causal relationships between variables (Denzin & Lincoln, 1998). Positivists look at information in the form of phenomena that can be quantified and summarised, analysing the resulting numerical data using mathematical processes and expressing results in statistical terms (Charles, 1995).

Qualitative research adopts a naturalistic or interpretive paradigm approach, seeking to understand phenomena in a context specific or "real world" setting using methods like interviews and observations. Qualitative researchers often produce findings not arrived at by statistical means or quantification, seeking instead illumination, understanding and extrapolation to similar situations (Hoepfl, 1997). Unlike quantitative research where researchers attempt to dissociate themselves as much as possible from the research process, qualitative researchers embrace their involvement and role within the research (Patton, 2001).

Whatever the research environment, researchers need to test and demonstrate that their studies are credible. In a quantitative research environment the credibility of the research depends on the instrument construction, while in qualitative research environment the "researcher is the tool" (Patton, 2001, p. 14). When a quantitative researcher talks about research validity and reliability they are seen to be referring to research that is credible, while in qualitative research the credibility relies on the ability and effort of the researcher (Golafshani, 2003). In quantitative research, reliability and validity are treated separately while in qualitative research the terms are not separated, and researchers prefer to talk about terms such as credibility, transferability, and trustworthiness.

### 3.5.1 Reliability in quantitative research

Reliability is a central concept in measurement and refers to whether a measure is consistent over time, is an accurate representation of the population under study and whether the results of a study can be reproduced using a similar methodology (Golafshani, 2003). A key element of establishing that a measurement is reliable is the idea that a result can be duplicated over time and across multiple similar measures. In the quantitative research setting reliability is most often measured using correlation coefficients, which measure the degree of relationship between variables.

Within the quantitative content analysis component of the current study, reliability is ensured by the use of what is often referred to as the use of test/retest reliability measures and the use of interrator reliability measures. When models are tested the operations performed by manual coders are duplicated and coefficients of interrator reliability are calculated. When a model is used to test the automated tool, the results are compared with those obtained by manual coders. To ensure the reliability of results obtained from the model and the human coders, alternate and more traditional models have been used for measurement and the results compared.

### 3.5.2 Validity in quantitative research

In addition to being reliable, results must be valid. Validity refers to whether a measure is trustworthy or genuine. It suggests that a valid measure is one that measures what it claims to measure. As with reliability, validity is measured by the use of correlation coefficients. Among the various approaches used to validate measurement instruments, three approaches are generally discussed, namely content validity, criterion validity and construct validity.

- Content validity ensures that the measures include an adequate and representative set of items that taps the concept (Cavana, Delahaye, & Sekaran, 2001). The common ways used to achieve content validity

involve conducting a literature review, using qualitative research and seeking advice and judgement from a panel of experts. All three methods of ensuring content validity have been used in this study.

- In criterion validity, an indicator is compared with another of the same construct in which the researcher has confidence. In the current study criterion validity has been addressed by comparing different tools and models when attempting to measure evidence of critical thinking.
- Construct validity testifies how well the results obtained from the use of a measure fit the theories around which the test is designed (Cavana et al., 2001). Researchers refer to convergent reliability which is established when scores obtained by two different instruments to measure the same concept are highly correlated. For this study, the last phase of this action research framework compares the results obtained using human coders, the automated tool and the Cambridge Thinking Skills Assessment to validate the automated tool.

### 3.5.3 Validity and reliability in qualitative research

While there is general agreement among researchers when discussing reliability and validity in the quantitative environment, the same cannot be said for the qualitative environment. Patton (2001) suggested that validity and reliability are factors that should concern any qualitative researcher when designing a study, analysing the results and judging the quality of the findings. Others argue that rather than trying to establish reliability and validity in the qualitative setting one should discuss issues of credibility, neutrality, confirmability, and transparency (Golafshani, 2003). Seale (1999) states that establishing trustworthiness is essential when discussing reliability and validity in the qualitative paradigm. Despite the disagreement on the value of attempting to prove reliability and validity in qualitative research, there does appear to be general agreement that triangulation is an accepted way of dealing with the issues of reliability and validity.

Triangulation is a test for improving the validity and reliability of research, Patton (2001) states that "triangulation strengthens a study by combining methods. This can mean several kinds of method or data, including the use of both quantitative and qualitative approaches" (p. 247). Steps taken by researchers when adopting triangulation techniques involve engaging in multiple methods of data collection, such as, interviews, observations and recordings. Others suggest involving several investigators or peer researchers at different times in a study. This study has adopted the involvement of peer reviewers at the end of each phase of the action research cycle and the keeping of an accurate account of comments made by the peer reviewers.

### 3.6 Conclusion

This chapter has examined the theoretical aspects of various research methodologies, with particular focus on the features and challenges of action research and content analysis. The action research framework under which the study was conducted was introduced along with the introduction of an artefact used as a tool to attempt to answer the research questions identified in chapter 1. The chapter concluded with an attempt to address the issues of reliability and validity as they relate to the study.

The next chapter provides a more detailed account of the second phase of the action research cycle and introduces the automated analysis tool and the quantitative content analysis models used to test the efficacy of the tool.



# CHAPTER 4: INITIAL DEVELOPMENT AND TESTING OF THE AUTOMATED TOOL

---

The real problem is not whether machines think but whether men do.

(Skinner, 1969, p.159)

The previous chapter described the research approach adopted for this study to explore the central questions related to automating the identification of the evidence of critical thinking among discussion forum participants. An action research approach provided a framework to support the development of a computerised tool built to help test the viability of using an automated tool. The previous chapter identified the action research phases undertaken during the study to allow the tool to be evaluated and refined.

This chapter, identified as phase 2 of the action research cycle, looks at the first stage of the development of the tool. It describes the conceptual design and explains how the tool was designed as an aid to reduce the workload associated with the manual content analysis process. The critical thinking content analysis models adopted to test the tool are described and the results of using the tool to evaluate evidence of critical thinking are listed and discussed. This chapter also describes how those results were shared and how feedback was sought to identify how the tool could be improved. The chapter concludes with a summary of the findings and an indication of the changes to be implemented in the second iteration of the tool.

## 4.1 Background

The literature review indicated that content analysis was one of the most widely used tools when trying to identify evidence of critical thinking among participants in discussions forums (Rourke et al., 2001; Henri,1992; Newman et al., 1995; Bullen, 1997; Gunawardena et al., 1997; Fahy et al., 2001; Garrison et al., 2000; Clulow et al., 2003; Perkins & Murphy, 2006). Since content analysis is seen as labour intensive and time consuming (Palmquist et al., 1997; Krippendorff, 2004), researchers have looked towards computers as a way of speeding up the categorisation process. The growth in availability of digitised texts and the development and adoption of computerised tools like NVivo and General Inquirer has led to a growing acceptance of the use of computers as a tool to assist with the task of analysing texts and classifying textual units into specified categories. Researchers like McKlin et al. (2002), Brook Wu and Chen (2005), and Dönmez et al. (2005) saw the potential of developing custom built computing applications to automate the content analysis process; they reported varying degrees of success.

This study attempts to build on the efforts of the small group of researchers who have developed customised tools to identify evidence of critical thinking among discussion forum participants. One of the principal aims of the study is to answer the question:

*" Is it possible to automate the identification of critical thinking ability exhibited by a group as they participate in a discussion forum?"*

An essential component of the study is the development of an automated tool, which is used to assist with the coding of transcripts and to assist with the identification of critical thinking indicators. The tool is referred to as the Automated Content Analysis Tool (ACAT).

## 4.2 The ACAT System

Rourke et al. (2001) suggest that content analysis involves four steps:

1. identify the representative sample and decide the unit of analysis;
2. identify an existing model (or create a new model) which will allow the transcript to be categorised and train coders how to use the model;
3. code the transcript components into the categories defined by the model;
4. compare the results for different coders to establish reliability.

When the initial version of the ACAT system was designed, the aim was to produce an automated system that reduced the labour intensive tasks associated with each of the four steps. The system was designed to import a discussion forum transcript (in text file format) and parse the transcript according to the prescribed unit of analysis.

The unit of analysis can be syntactical, such as the word, preposition, sentence or paragraph, which are delineated by syntactical criteria. It may also be a "thematic" unit which identifies a single thought or idea. As Rourke et al. (2001) point out, the thematic approach is difficult to implement, requiring in-depth processing to identify the unit, which in turn increases the opportunity for subjective ratings and is likely to result in low reliability scores.

A syntactical approach was therefore adopted, when designing the ACAT system. The syntactical approach is relatively easy to implement using advanced string handling features, and the syntactical approach avoids the issues associated with thematic coding identified by Rourke et al. (2001). The initial version of the ACAT system has the capability to parse a transcript on the basis of individual words, sentences and paragraphs.

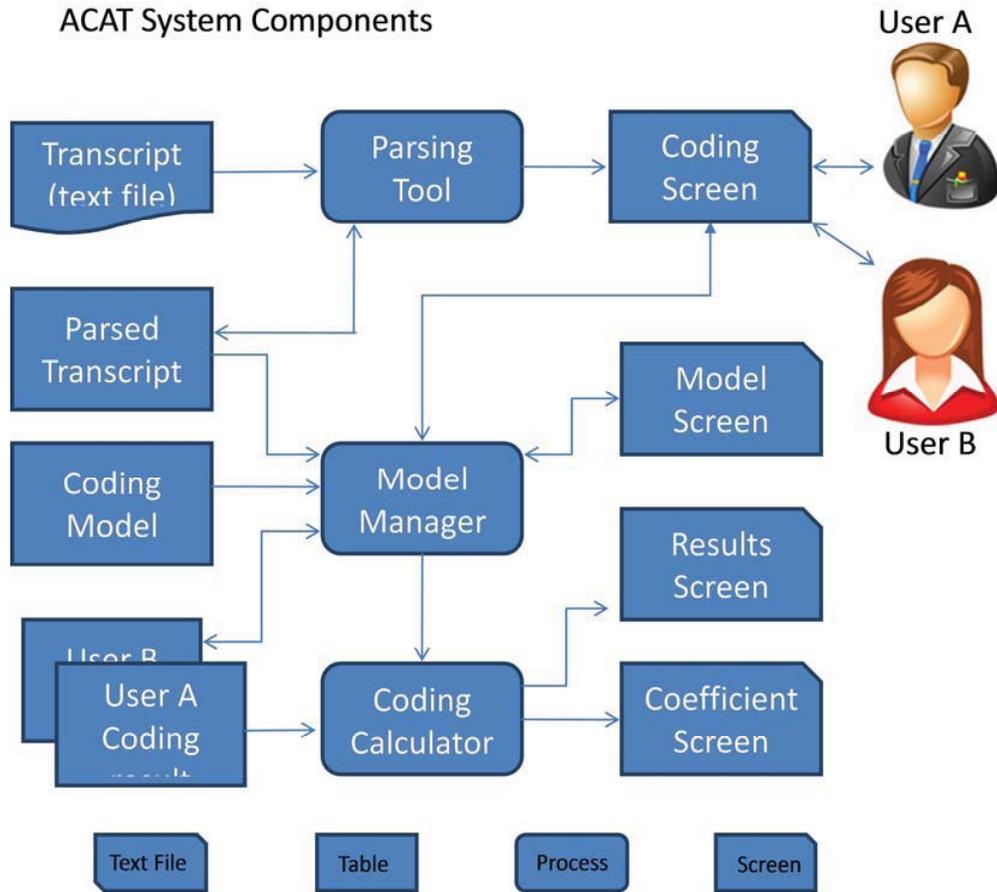
Based on the suggestion by Rourke & Anderson (2004), that researchers should use coding schemes that have already been developed and used in previous research,

rather than develop a new coding scheme, the ACAT system was designed to allow a variety of existing coding models to be adopted. As Rourke & Anderson suggest, such an approach encourages replicability and provides a means of further validating the chosen instrument.

The initial version of the ACAT system allowed the researcher to import a discussion forum transcript, then automatically parse the transcript to produce units for analysis presenting one unit at a time on screen. The system allowed the human coders to select the most appropriate category for coding, save the result and move on to the next unit. When the coder had finished coding the transcript, the results were saved in a table. The system was designed to store results for multiple coders. The system also provided a feature that automatically calculated the Holsti coefficient of reliability and the Cohen's kappa inter-rater reliability coefficient when the coding results of pairs of individual coders were compared.

When designing the initial system, the intention was to use it as the basis for another iteration of the system that would also automatically code the transcripts using a chosen model. The results of the manual coding exercise would be used to form the basis of a dictionary that would be used for the automatic coding process.

Figure 4.1 shows the components of the initial ACAT system components and indicates the table structures utilised for storing the parsed transcript, the coding model structure and the coding results for each of the individual coders. The diagram also lists the main processes and the user interface screens.



**Figure 4.1: ACAT System Components**

Coding models are added using the model screen (Figure 4.2) and stored in the coding model table. The transcript, in the form of a text file, is loaded into the system and passed to the parsing tool where the transcript is divided into analysis units and stored in the parsed transcript table. When a user wishes to code a transcript, he/she opens the coding screen (Figure 4.3), and the system presents analysis units one after the other. The user codes the transcript by allocating each of the units to the most appropriate model category. When the coding process is completed the coding results are saved in a coding result table. Summaries of the results for each coding exercise can be viewed using the results screen (Figure 4.4). Coding result tables are created

for each individual coder, allowing correlation coefficients to be calculated by comparing the codes allocated by pairs of coders. The results of the correlation calculations are presented on the coefficient screen (Figure 4.5).

**Edit Taxonomy Model**

Taxonomy Model Name	Garrison et al.
Taxonomy Levels	4
Taxonomy Code	GARRIS
<b>Taxonomy Level Details</b>	
Level 1	Triggering
Level 2	Explanation
Level 3	Integration
Level 4	Resolution

**Figure 4.2: ACAT Taxonomy Model Editing Screen**

The ACAT system coding screen allows users to create and edit coding models. To create a model the user must enter the model name and the number of levels associated with the model. When users enter the number of levels the ACAT system prompts them for descriptions of each of the levels.

ACAT - QCA Model Coding	
Data Source	transcript.txt
Analysis Type	QC Model Analysis
Taxonomy Type	Garrison et al.
Unit Text	
<div style="border: 1px solid black; padding: 5px;">           What is meant by the term quality of service         </div>	
Number of Levels	4
Select Model Level	
Level 1: Triggering	<input type="radio"/>
Level 2: Explanation	<input type="radio"/>
Level 3: Integration	<input type="radio"/>
Level 4: Resolution	<input type="radio"/>
Unclassified	<input type="radio"/>
Action	
<input type="button" value="Save Unit"/> <input type="button" value="Next Unit"/> <input type="button" value="ACAT Menu"/>	
- Automated Content Analysis Tool	

Text unit not yet saved....

**Figure 4.3: ACAT Transcript Coding Screen**

The ACAT system coding screen allows users to code units of analysis against the selected model. The system presents the text associated with the unit of analysis and automatically presents the user with the number of levels that correspond to the selected model. Users can allocate a unit to one of the levels or they may elect to place the unit in the unclassified category.

ACAT - Analysis Report				
Data Source	transcript.txt			
Analysis Type	QC Model Analysis			
Taxonomy Type	Garrison et al.			
Coder	S.Corich		D.Strode	
Number of Levels	4		4	
Level 1: Triggering	73	15.08%	70	14.6%
Level 2: Exploration	124	25.6%	120	24.8%
Level 3: Integration	209	43.2%	218	45.0%
Level 4: Resolution	58	12.0%	46	10.1%
Unclassified	20	4.13%	27	5.6%
Correlation Coefficients				
Holsti's Coefficient	0.82			
Cohen's Kappa	0.76			
Return to ACAT Menu				
Automated Content Analysis Tool				

**Figure 4.4: ACAT Results Screen**

The ACAT results screen presents a summary of the coding process for each coder. The system calculates the total number of analysis units coded; the number of units assigned by the coder to each category; the percentages of units in each category.

ACAT - Analysis Report				
Data Source	transcript.txt			
Analysis Type	QC Model Analysis			
Taxonomy Type	Garrison et al.			
Coder	S.Corich		D.Strode	
Number of Levels	4		4	
Level 1: Triggering	73	15.08%	70	14.6%
Level 2: Exploration	124	25.6%	120	24.8%
Level 3: Integration	209	43.2%	218	45.0%
Level 4: Resolution	58	12.0%	46	10.1%
Unclassified	20	4.13%	27	5.6%
Correlation Coefficients				
Holsti's Coefficient	0.82			
Cohen's Kappa	0.76			
Return to ACAT Menu				
Automated Content Analysis Tool				

#### **Figure 4.5: ACAT Coefficient Screen**

The ACAT system compares the coding decisions between any two pairs of coders and automatically calculates both Holsti's CR and Cohen's kappa. The coefficient screen presents a summary of the coding allocations for each coder and displays Holsti's CR and Cohen's kappa.

An open source web-based platform was chosen as the development environment for the ACAT system. The open source platform utilised a combination of a PHP application supported by a MySQL database. PHP is a recursive acronym for Hypertext Preprocessor, it is a popular general-purpose server-side scripting language which can be embedded into HTML to create a wide variety of web-based applications. MySQL is an open source database software based on the SQL vocabulary which can be employed in combination with most server-side languages, but which is most commonly employed with PHP. The platform was chosen as it allows the application to be shared on the web and if required it can be integrated into one of the freely available open source Learning Management Systems (LMS).

Having developed the initial version of the ACAT system the researcher reviewed the literature to identify which of the existing critical thinking measurement models should be applied to the coding component of the system. The literature review presented in chapter 2 identified a number of researchers who had used content analysis to develop models to measure evidence of critical thinking in discussion forums. Two of the most popular models were those developed by Henri (1992) and Garrison et al. (2000). Since both of these models had been duplicated or incorporated into models developed by other researchers, they were chosen for the first application of the ACAT system.

The model developed by Henri (1992) has five categories: elementary clarification, in-depth clarification, inference, judgment, and strategies. The paper in which Henri described using the model lacked details on how to allocate units to each of the five categories. A subsequent study conducted by Hara et al. (2000), using the same

coding categories, provided more detailed information and listed a number of indicators for each category to assist with the coding procedure. Table 4.1 lists the coding categories and the associated indicators for the model described by Hara et al. (2000).

**Table 4.1**

**Coding categories and indicators for Hara et al. (2000)**

<b>Category</b>	<b>Indicators</b>
Elementary clarification	Identifying relevant elements Reformulating the problem Asking a relevant question Identifying previously stated hypotheses Simply describing the subject matter
In-depth clarification	Defining the terms Identifying assumptions Establishing referential criteria Seeking out specialised information Summarising
Inferencing	Drawing conclusions Making generalisations Formulating a proposition which proceeds from previous statements
Judgment	Judging the relevance of solutions Making value judgments Judging inferences
Application of strategies	Making decisions, statements, appreciations, evaluations and criticisms Sizing up

Source: Hara et al., 2000, p. 10.

The COI model has four categories: triggering, exploration, integration, and resolution. Table 4.2 lists the four categories of the COI model and the associated behavioural indicators.

**Table 4.2**

**Coding categories and indicators for the COI model**

<b>Category</b>	<b>Indicators</b>
Triggering	Recognizing the problem Sense of puzzlement
Exploration	Divergence within online community Divergence within single message Information exchange Suggestions for consideration Brainstorming Leaps to conclusions
Integration	Convergence among group members Convergence within a single message Connecting ideas, synthesis Creating solutions
Resolution	Vicarious application to real world Testing solutions Defending solutions

Source: Garrison et al., 2001, p. 15 - 16.

The Henri (1992) model with its five categories, and the COI model with its four categories were added to the ACAT system and stored in the model tables.

### 4.3 Application of the Initial ACAT System

The initial version of the ACAT system was designed to reduce the time spent parsing a transcript, coding analysis units against a model and validating the results by calculating reliability coefficients between pairs of coders. The system was used to support two human coders, who coded the discussion forum transcript resulting from a discussion forum that was an assessed component of a first year undergraduate degree programme.

#### 4.3.1 Methodology

As identified in the previous chapter, the method used to assess evidence of critical thinking during this phase of the research was content analysis. Content analysis is described by White and Marsh (2006) as a highly flexible research method which has been widely used with varying research goals and objectives. Having established the method for this phase of the study, the next step was to decide on the categories into which segments of messages should be coded. Rather than design a new model, the decision was made to use two existing models. The first was the model developed by Henri (1992), which has five categories, the second was the model developed by Garrison et al. (2000), which has four categories. Having identified the models for coding, a decision on the unit of analysis for coding was required. Since one of the aims of the study is to automate the process of content analysis, a coding unit that could be reliably identified using a computer was required. An approach based on syntax was adopted for the system as it was viewed as being easier to implement than a thematic approach which would have been difficult to implement using computer technology. Following in the footsteps of Fahy et al. (2000) and McKlin et al. (2002) the sentence was chosen as the unit of analysis.

The research to test the efficacy of the initial ACAT system was conducted in the second semester of 2003 and involved a discussion forum activity undertaken by a first year undergraduate degree course at a tertiary education institution in New Zealand. All the students were enrolled in a computing systems degree and as such

were familiar with using information technology. The course was an introductory data communications and networking class. It was delivered using a blended learning environment, combining traditional face-to-face activities with web publishing, on-line review and discussion forum activities. On-line activities, which included publishing the results of a research project, evaluating the work of peers and participation in a discussion forum formed a significant component of the course. The use of the discussion forum was seen as a way to encourage participation as well as to provide a tool to promote discussion about a topic that was a key component of the course curriculum. As a way of encouraging participation, the discussion forum activity was included as an assessed activity with a course weighting of 15 percent. Previous offerings of the course had covered the same topic, the future of data communications, in a normal classroom setting, using face-to-face discussion over a two hour period. Using the discussion forum approach, students were allowed three weeks to participate in on-line discussion.

The software used to support the discussion forum was an integral part of the Blackboard learning management system. All the students had previously used Blackboard to retrieve course materials and to participate in on-line tests in their earlier courses; however none of the students had participated in discussion forums during their previous academic study.

The class consisted of fifteen students, three females and twelve males, aged between 18 and 38, and of varying academic abilities. Students were given the topic for the discussion early in the course and instructions were provided to the students as to what was expected in the discussion forum. The instructions were given as a guide to encourage higher level critical thinking. The student postings were monitored by an instructor who provided encouragement, added pedagogical comments and provided reinforcement and expert advice.

### 4.3.2 Findings and discussion

During the three weeks that the discussion forum was operational a total of 104 posts were made, 30 of which were made by the course instructor. Once the instructor postings were removed, the remaining 74 posts generated 484 sentences for coding.

Participation in the forum varied with almost 35% of postings being made by the three female class members. The six class members over the age of twenty-five accounted for 63% of postings. Of the under twenty five age group, one class member took no part in the discussion forum activities and the Blackboard software indicated no activity in the discussion forum area, while another who made no postings, obviously read postings as Blackboard indicated significant activity in the forum area. Two under twenty-five year olds made only a single posting; however monitoring software indicated significant activity for both.

Before coding the entire transcript, the two volunteers coders (one being the researcher and the other a lecturer who had previously taught the data communications and networking course but was no longer associated with the delivery of the course) looked at the first 150 sentences from the transcript and agreed on how the sentences should be coded against each of the two models. Having agreed on the coding procedures, both coders used the ACAT system to code the complete transcript firstly using the Hara et al. (2000) model and then using the COI model developed by Garrison et al. (2000). Once the entire transcript had been coded for both models by both coders, the ACAT system was used to establish the levels of agreement that existed, and results for Holsti's CR and Cohen's kappa were produced.

Table 4.3 and table 4.4 present the summary information on how the 484 sentences were classified for Hara et al. and COI model. respectively. Table 4.5 indicates the coefficients of reliability obtained using the ACAT system.

**Table 4.3****Number of postings using the Hara et al. model**

Category	Coder A		Coder B	
	Number of Sentences	Percent of Total Sentences	Number of Sentences	Percent of Total Sentences
Elementary clarification	63	13.02%	58	11.98%
In-depth clarification	121	25.00%	108	22.31%
Inferencing	145	29.96%	152	31.40%
Judgment	63	13.02%	71	14.67%
Application of strategies	53	10.95%	49	10.12%
Not categorised	39	8.06%	46	9.50%
Total number of postings	484	100.00%	484	100.00%

To establish the attitudes of the coders towards the text coding process the two coders discussed the coding process. The discussion was recorded and a transcript produced at a later date. The coders both indicated that the Hara et al. model was harder to code having five categories compared to the COI model with four. For both coders the majority of units were categorised as "inferencing", suggesting that participants were drawing conclusions, generalising and formulating propositions from previous statements. By far the biggest proportion of units were categorised in the three lower categories of the model (more than 65% for both coders), suggesting that most activity was lower end of the critical thinking disposition scale. Compared to the

findings presented by Hara et al. (2000), the students participating in this study displayed significantly lower levels of critical thinking.

The correlation between the coders using the Hara et al. (2000) model was 0.75 for the Holsti CR and 0.69 for the Cohen's kappa. While these appear acceptable when compared with previous studies, it should be noted that almost one third of the units (150 out of 484) were allocated to the same categories by both coders following the inter-coder discussion prior to the completion of the transcript coding exercise.

**Table 4.4**

**Number of postings using COI model**

Category	Coder A		Coder B	
	Number of Sentences	Percent of Total Sentences	Number of Sentences	Percent of Total Sentences
1. Triggering	73	15.08%	70	14.46%
2. Exploration	124	25.62%	120	24.79%
3. Integration	209	43.18%	218	45.04%
4. Resolution	58	11.98%	49	10.12%
Not categorised	20	4.13%	27	5.58%
Total number of postings	484	100.00%	484	100.00%

The discussion between coders indicated that the COI model was the easier of the two models to code. The coefficients of agreement were higher using the COI model and there were a lower number of uncategorised sentences. The findings using the COI model indicated small proportions of both "triggering" (approximately 15%) and "resolution" (approximately 11%) categorisations. The majority of the units were classified as belonging in the "exploration" and "integration" categories, indicating

that participants were exchanging information and building on each other's ideas. When compared with the findings published by Fahy et al. (2001), Meyer (2004) and Garrison et al. (2000) the participants in this study once again displayed lower levels of critical thinking.

The correlation between the coders using the COI model was 0.82 using Holsti's CR and 0.76 using Cohen's kappa. These compare favourably with previous studies using the Garrison et al. model.

**Table 4.5**  
**ACAT calculations for reliability coefficients**

<b>Model</b>	<b>Holsti's CR</b>	<b>Cohen's kappa</b>
Hara et al.	0.75	0.69
Garrison et al.	0.82	0.76

It was interesting to note that both models indicated similar levels of evidence of knowledge construction and low levels of synthesis and real world application. Both models reflected the preference of students to build and expand on ideas suggested by others.

When comparing the results of this study with previous studies, it should be noted that this is one of the few documented studies involving the measurement of critical thinking among discussion forum participants who were undergraduate students. The studies conducted by Hara et al. (2000), Fahy et al. (2001), Meyer (2004) and Garrison et al. (2000) all involved participants who were involved in post-graduate studies. The lower levels of critical thinking evident from the findings of this study compared with earlier studies is interesting and there is no evidence within the literature to suggest that this is an unexpected result. The finding does provide an

opportunity for a wider study to establish the extent to which critical thinking skills are developed as a student progresses through an undergraduate degree.

When the researcher and second coder discussed the results of the coding exercise, the second coder suggested that the quality of the first posting, which in this case was not of a particularly high calibre, set the tone for the remainder of the discussion forum activity. Once again there is an opportunity for further research, to establish the impact of initial postings on the quality of subsequent postings. Similarly opportunities exists for further studies to look at the impact that instructor postings may have on the quality of critical thinking exhibited among discussion forum participants.

When questioned about the experience of using discussion forums as a way of discussing topics which are relevant to the course prescription, thirteen of the fifteen students stated that they believed that it was a worthwhile exercise. Several indicated that the process was very time consuming but also said that the three week period gave them time to think about the topic and conduct research to assist with the postings. Students also stated that they found the discussion forum to be “addictive”, creating a desire to continually check to see if their postings had induced a response from the instructor or their fellow students. All those who participated in the exercise suggested that they had increased their knowledge as a result of the exercise and would be happy to participate in a similar exercise in the future.

The findings of this study indicate that both models appear to provide a useful tool for measuring the quality of student participation within an online discussion forum. Both models gave similar measures of the level of critical thinking and knowledge construction. Since no consideration was given to the identity of the person making the posting, neither of the models was able to measure the contributions of individual participants.

While attempting to evaluate the relative merits of each model was not really an aim of this part of the research, the fact that both models indicated similar patterns of critical thinking would suggest that both have their places in the field of discourse analysis. When asked about the coding exercise activity, both coders agreed that the process was time consuming and both indicated a preference for the COI model. The resulting coefficients of agreement between coders would suggest that even though there was a level of agreement, the classification of messages is open to individual interpretation. The researcher and second coder agreed that if the models were to be applied to a larger population with a significantly larger number of postings, then some form of automatic coding system would need to be considered. Such a tool would need to be efficient, reliable, valid and practical.

While the use of content analysis as a mechanism for measuring quality was shown to have merits, neither of the models tested gave any indication of how the critical thinking had been applied to the subject domain that the forum was discussing. This combined with the inability to identify evidence of an individual's critical thinking would suggest that the models and the tool in their present form could not be used as a tool to grade student performance in an assessed discussion forum relating to a particular subject domain area.

When asked about the use of discussion forums as a tool to complement the traditional face-to-face teaching environment, students indicated that for those who are familiar with technology, the use of discussion forums could enhance the learning process. The evidence of this research would also suggest that the discussion forum mechanism better suits relatively mature learners who have a desire to learn and take responsibility for knowledge construction. Discussion forums may also disadvantage students who have poor written communication skills and students who may be participating in a forum where the language used is not their first language.

#### 4.4 Peer review

The action research approach described in the methodology chapter indicated that peer review at the end of each action research cycle was an essential component of the study. The aim of the peer review process was to reflect on the findings of any action research cycle and establish the direction to be taken for subsequent research. For this study, the peer review process involved both national and international review. The avenue for national review was through meetings of the Advanced Learning Technologies Research Centre (ALTRC) at Massey University in Palmerston North, New Zealand, the avenue for international review was through presentations at a number of international conferences.

ALTRC was established in 2004 as a research centre with the aim of advancing research on the innovative paradigms, architectures and implementations of learning systems for individualised and adaptive learning. The centre was established within the School of Information Systems at Massey University in Palmerston North and involved a number of full-time Massey University academic staff, as well as a number of doctoral and master's students who were involved in educational technology related research. Students were expected to meet on a regular basis to present their research and receive feedback from staff and peers.

The initial ACAT system was demonstrated to a meeting of the ALTRC early in 2004 and the results for the research detailed in this chapter was presented for critical review. The meeting was attended by four Massey University staff members, eight PhD students and six masters students. The researcher explained to attendees that he was recording the presentation and that a written transcript of the meeting would be produced. The researcher also explained the important role that peer review played in the research methodology and that the peer review process was being used to inform the direction of the next phase of the research. Those attending the meeting all agreed that the research related to automating the identification of critical thinking among discussion forum participants was an interesting concept which was worthy of

investigation and a number of questions were asked relating to the future direction of the research. The questions included:

1. How will the automation of the posting allocation process be achieved?
2. What methods or algorithms will be used to automate the classification process?
3. Will a reference dictionary approach be adopted, and if it is how will it be built?
4. Will the system results be compared with those of human coders?
5. How will the system be validated?
6. How will the system deal with postings that are of a social nature and not indicative of critical thinking?
7. How will the automated system handle poor spelling and syntactical errors?

The findings of the first phase of the ACAT system research were also presented as a paper at an international conference, the Third Pan Commonwealth Forum on Open Learning, which was held in Dunedin, New Zealand (Corich et al., 2004a). The conference was attended by open learning representatives from throughout the Commonwealth. The paper presentation was attended by approximately 35 conference participants and before the presentation commenced the researcher explained that the paper which was being presented formed part of a PhD research project aimed at attempting to verify if it was possible to automate the identification of critical thinking among participants in discussion forums. The researcher also explained that he was using the conference presentation as part of a peer review process, to help inform the future direction of the study. The attendees all expressed

an interest in the concept of automating the identification of critical thinking, and suggested that should a system be built that it would most likely have applications that extend beyond those associated with discussion forum participation. There was a lively debate on the way that the research project might proceed and a number of questions were raised. The questions asked at the conference were directed at the plans for future development and were similar in nature to those asked at the ALTRC symposium.

The findings detailed in this chapter were also submitted as a paper to the *International Journal of Instructional Technology and Distance Learning*, which was published in December 2004 (Corich et al., 2004b). The review panel stated that the research covered an important aspect of educational research and that the system as presented merited further research. The only questions raised by the panel related to the proposed future developments planned for the ACAT system.

The feedback received from the ALTRC symposium, the Third Pan Commonwealth Forum on Open Learning and the paper submission to the *International Journal of Instructional Technology and Distance Learning* provided the basis for the next phase of the ACAT system development described in the following chapter.

## 4.5 Conclusion

This chapter described the development and testing of the initial phase of the ACAT system. It provided details on how two human coders used the semi-automatic system to code against two models designed to measure critical thinking in discussion forums and discussed the results. The chapter concluded by describing the peer review process used to inform the development of the next phase action research cycle.

The next chapter describes how the feedback received during the peer review process was used to assist with the redesign the ACAT system. The chapter also reports on how the system was used to code a transcript automatically using the COI model and

how the results obtained by using the tool compared to those obtained by a human coder.

# CHAPTER 5: USING THE ACAT SYSTEM TO MEASURE GROUP PARTICIPATION

---

Discovery consists in seeing what everyone else has seen and thinking what no-one else has thought.

Albert Szent-Gyorgyi (1893 - 1986)

The previous chapter described the planning, action and evaluation steps of the first phase of the action research cycle outlined in chapter 3. It explained the conceptual design of the initial ACAT system, a semi-automated computerised tool built to help answer the research question relating to the ability to automate the identification of critical thinking. The chapter also discussed the results obtained using the tool and the peer review process undertaken to identify how the tool could be refined.

This chapter, identified as phase 2 of the action research cycle, considers the questions that were raised as part of the peer review process and describes how the tool was refined in response to those questions. The chapter explains the changes that were made to the tool and discusses the results obtained using the tool to automatically code a discussion forum transcript. The chapter concludes with a summary of the findings of a second round of peer review as well as an indication of the questions that need to be addressed with the third iteration of the tool.

## 5.1 Background

The previous chapter described how the ACAT system was used to assist two human coders, as they attempted to categorise the postings obtained from the transcript of a discussion forum used to support the delivery of a first year data communications course. The ACAT system parsed the transcript, presenting sentences to the coders so

that they could be manually categorised using a model developed by Henri (1992) and refined by Hara et al. (2000). The system also allowed coders to categorise the same sentences using a second model, known as the cognitive component of the COI model.

The results of the categorisation process were shared at a doctoral symposium and an international conference, where a number of questions were raised which were used as the catalyst for redevelopment of the ACAT system. Several of the questions related to the design of the process that would be used to automatically code the discussion forum transcript. The questions, relating to the design of the automated coding process, that need to be addressed include:

1. How will the automation of the posting allocation process be achieved?
2. What text classification algorithms will be used?
3. How will the reference dictionary be built?
4. How will the automated system handle poor spelling and syntactical errors?

The answers to these questions lies in the experiences of researchers who have previously attempted to use computers to analyse and categorise digital texts. The automated content analysis of digital texts has not been limited to discussion forums and includes essay grading, media analysis, search engine design and automatic test marking.

## 5.2 Automatic Content Analysis

The literature review identified a number of researchers who have published their experiences of attempting to automate the content analysis process. McKlin et al. (2002); Brook Wu and Chen (2005); and Dönmez et al. (2005) all saw the potential of using computers to automate the content analysis process. Krippendorff (2004) discusses how computers have been used to aid the process of content analysis, he refers to the use of computer-aided text analysis (CATA) software and describes four different approaches:

- Accounts of character strings
- Text Searches
- Computational content analyses
- Interactive-hermeneutic approaches.

Software developed to provide accounts of character strings partition a body of text into convenient textural units and list, sort and cross tabulate them. The process relies on syntactical clues rather than meaning. Krippendorff (2004) suggests that the usefulness of such an approach relies on the analyst's ability to read and make sense of the character strings that are tabulated and counted. He also points out that software designed to provide accounts of character strings is adversely affected by poor spelling and grammatical errors. He explains that this type of software often requires use of "stop word" lists and "word stemming" procedures to improve the quality of the results that are produced.

Text searches are accomplished through the use of search engines. Text searching involves the scanning of large amounts of textual information to retrieve texts that can be expected to contain information that matches a specified search string. Users formulate queries in accordance with the search engine's rules and the search engine scans all the texts to which it has access and identifies and displays the units of texts that match the query.

Krippendorff (2004) states that computational content analysis software provides features that go beyond the syntactically based approach of accounts of character strings and text searches. Computational content analysis software processes texts according to theories of meaning that are presumed to exist within the context of the text being analysed. Such software systems often achieve their high level of sophistication by insisting that users perform significant pre-editing of the raw text. Within this category of text processing software, Krippendorff (2004) identifies four different approaches: coding/dictionary approaches, statistical association approaches, semantic network approaches and memetic approaches.

The developers of coding or dictionary-based CATA software adopt a taxonomy approach, categorising texts according to a predefined model. This is the approach adopted by some of the better known text analysis tools: General Inquirer, Wordstat and TextQuest.

Statistical association approaches rely on an assumption that important words in a text are identifiable by their relative frequency, their proximity to each other, and how they relate to each other. The statistical association approach is popular amongst those involved in the field of artificial intelligence and has been used by researchers attempting to grade essays and analyse interview transcripts.

When adopting a semantic network approach, researchers look at how concepts contained in a text relate to each other. Semantic network CATA systems often produce output in the form of a network diagram showing how various concepts identified within a document or text relate. Like the statistical association approach, semantic network modelling is popular among those involved in artificial intelligence research.

When adopting a memetic approach, researchers place less emphasis on the meanings within a text and more emphasis on how a specific text relates to previously

published texts. Turnitin, an Internet based software system designed to identify evidence of plagiarism is probably the most well-known example of a CATA system built using a memetic approach to content analysis.

Krippendorff (2004) referred to the final category of CATA software as interactive-hermeneutic software, better suited to qualitative research. He uses the word interactive to identify systems where the categories of analysis and the choices of analytical construct are not fixed, and he uses the word hermeneutic to recognise that the analysis is directed by a researcher's growing understanding of the body of texts being analysed. Krippendorff (2004) suggests that interactive-hermeneutic software should provide features that allow text to be entered into the system, displayed and manipulated with ease. The software should also allow researchers to code text into categories manually and to perform both hierarchical and non-hierarchical categorisation automatically.

### 5.2.1 The ACAT System: Coding Categorisation Process

A coding/dictionary approach was adopted for development of the automatic coding categorisation process for ACAT system. The coding/dictionary approach was chosen as it provided a relatively straightforward way of incorporating the existing critical thinking measurement models that were described in the previous chapter into the system. When compared with the statistical association, semantic network and memetic approaches, the coding/dictionary approach provided a solution that was less sophisticated and as a result was within the software building capabilities available to the researcher.

Having made a decision on how to automate the posting allocation process, an answer to the question concerning which text categorisation algorithm should be used to allocate individual coding units to a particular category needed to be addressed.

### 5.3 Text Categorisation Algorithms

Text categorisation refers to the way that text is assigned to the different categories in text documents. A number of statistical methods have been applied aimed at solving the text categorisation problem, including regression models (Gelman & Hill, 2007; Harrell, 2001), nearest neighbour classifiers (Cover & Hart, 1967; Liao & Vemuri, 2002), Naive Bayesian probabilistic classifiers (Pedersen, 2000; Ramoni & Sebastiani, 2001), decision trees (Janikow, 1998; Burham & de Wolf, 2002), support vector machines (SVM) (Vapnik, 1995; Joachims, 1998) and artificial neural networks (Chen, 2001; Ripley, 2008).

Regression analysis and nearest neighbour classifier are similar analysis techniques based on instance-based learning or memory based learning. With instance-based learning, categorisation of text is achieved by comparing the attributes of a new text unit with those seen in training (Harrell, 2001). The unit being classified is coded into the category with the closest fit. Both these analysis techniques are said to require large training samples before they become accurate, and they are reported to be less accurate than Bayesian probabilistic classifiers.

A Naive Bayesian classifier is a simple probabilistic classifier based on applying Baye's theorem that is commonly used in text categorisation (Khan, Baharudin, Lee, & Khan, 2010). With Naive Baye's the basic idea is to use joint probabilities of words and categories to estimate the probability of a coding unit falling within a specific category (Pedersen, 2000). Naive Baye's classifiers can be trained very efficiently and require very small amounts of training data to operate effectively (Khan et al., 2010). Naive Baye's classifiers have proven to perform well in a number of real world classification exercises (Rish, Hellerstein, & Thathachar, 2001; Domingos & Pazzani, 1997; Kim, Rim, Yook, & Lim, 2002; Isa, Hong-lee, Kallimani, & RajKumar, 2008). Because of its simplicity, Naive Baye's has become one of the most popular machine learning methods and has been used to classify email, web contents and spam (Khan et al., 2010).

The decision tree is an approach that is often applied to automatic text classification situations based on training data. The decision tree constructs well-defined true/false queries in the form of a tree structure (Khan et al., 2010). In such a structure, leaves represent the categories and branches represent features that lead to categories. Advantages of the decision tree approach are that it is easy to understand and relatively easy to implement using software. The most significant disadvantages of the decision tree approach are that it can handle only one category at a time and decision tree algorithms can be unstable (Pazzani & Billsus, 1997).

SVM uses discriminative classification methods and requires pairs of training sets. SVM training algorithms build models that predict whether a text string falls into one category or the other (Khan et al., 2010). Khan et al. (2010) suggest that while SVM classification is very effective, the training and categorisation algorithms are relatively complex and building the training data sets can be very time consuming.

An artificial neural network (ANN) is a computational model inspired by the structure and functional aspects of biological neural networks, consisting of an interconnected group of artificial neurons which processes information using a connectionist approach to computation (Khan et al., 2010). ANNs are able to handle documents with high-dimensional features and documents with noisy and contradictory data. The significant disadvantages of the ANN approach are that they are difficult to understand and require high computing processing power to implement.

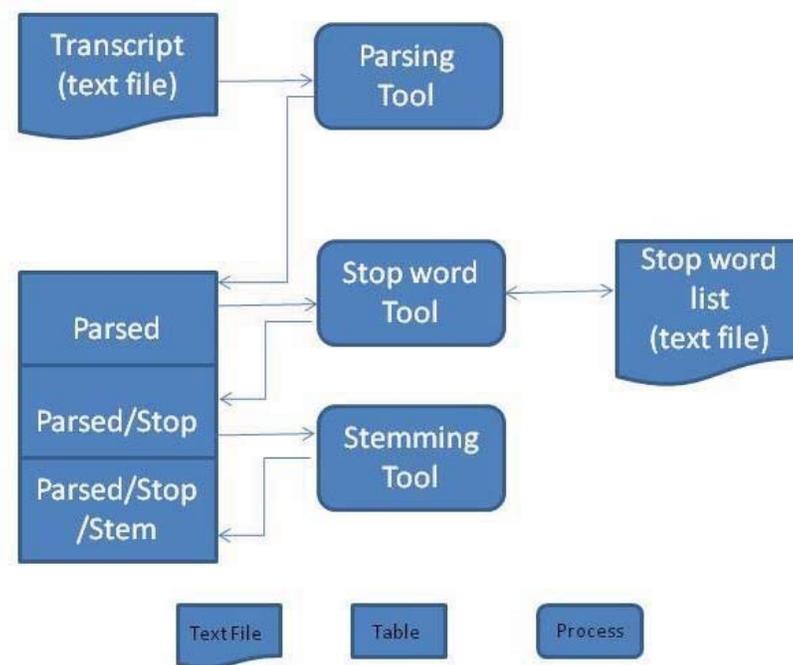
### 5.3.1 The ACAT System: Text Categorisation Algorithm

The Naive Bayes text categorisation approach was adopted for the ACAT system. This approach was chosen because of its simplicity and its popularity as a tool to classify web-based documents. When compared with nearest neighbour classifiers, decision trees, support vector machines and artificial neural networks, the Naive

Baye's classification approach provides a solution that works with small training data, is relatively simple to implement and once again is within the software building capabilities available to the researcher.

#### 5.4 The ACAT Automatic Coding Process Module

The ACAT system automatic coding process module attempts to automate the human coding process described in the previous chapter and produces a report indicating the total number of analysis units coded, the number of units assigned to each category and the percentages of units in each category. When designing the ACAT system automatic coding process module, some of the features found in traditional text categorisation systems were adopted. Khan et al. (2010) refer to document pre-processing techniques; these include tokenisation of texts, stopword removal and word stemming. Figure 4.1 illustrates the pre- processing techniques that are common to most text categorisation systems.

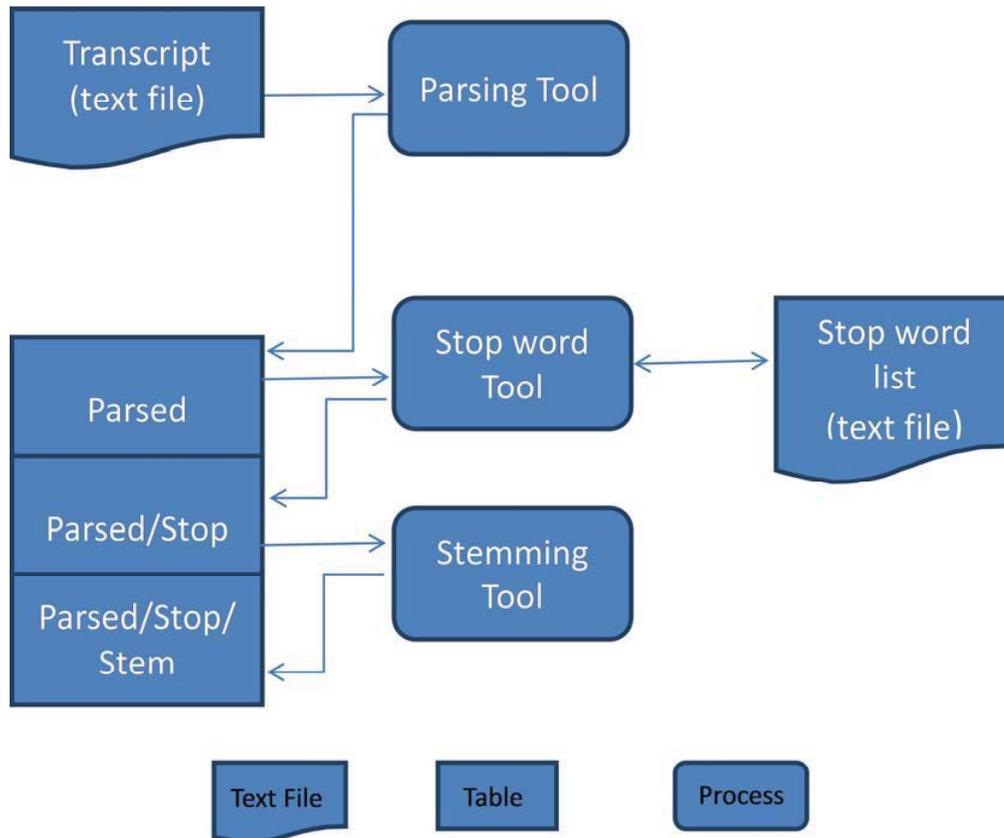


**Figure 5.1: Documentation Classification Process**

The documentation classification process described by Khan et al. (2010) starts with tokenising of the text; this involves treating the document as a string and partitioning it into tokens. This equates to the ACAT parsing process described in the previous chapter. The next two steps are designed to reduce the impact of spelling and grammatical errors and improve the speed of indexing when comparing documents. The first of these steps involves removing what are referred to as "stopwords". Stopword removal is accepted as a necessary part of information classification and retrieval systems (Riloff, 1995). Stopwords include frequently occurring words like "the", "a" and "and"; they are usually removed from documents and training texts to improve the text categorisation process. The second step involves the application of a stemming algorithm, an algorithm for reducing inflected or derived words to their stem, base or root form. Like stopword removal, stemming is an accepted practice in the world of computer assisted text classification and is said to improve the speed of indexing and reduce the impact of spelling and grammatical errors (Fuller & Zobel, 1998).

Both stopword removal and stemming algorithm features were incorporated into the ACAT system. The ACAT stopword removal system takes coding units (sentences) from the Parsed Transcript table, examines each word within the unit and compares it with a predefined list of stop words. Words that match the list are removed from the coding unit and the unit is stored in a separate field in the Parsed table. The literature suggests that there is no universally adopted English stop word list. The list selected for the ACAT system was downloaded from the Université de Neuchâtel web site and contains 571 words. Porter's stemming algorithm was chosen as the stemming algorithm. Porter's algorithm was developed in 1980 by Martin Porter of Cambridge University and is reported to be the most widely used stemming algorithm (Peng, Ahmed, Li, & Lu, 2007). The ACAT stemming algorithm system takes the coding units with the stop words removed from the Parsed table, the stemming algorithm is applied to each word and the stemmed coding unit is saved in another additional field within the Parsed table. Separate fields were created to store the parsed transcript, the

parsed/stop word transcript and the parsed/stop/stem transcript so that the coding results for each could be compared. Figure 5.2 shows how the stopword and stemming processes have been incorporated within the ACAT system.



**Figure 5.2: ACAT System Stop Word & Stemming Processes**

Since a coding/dictionary approach was adopted for development of the automatic coding categorisation process for the ACAT system, another table was created to store the training dictionary. The training dictionary table stores the training data, which is used by the text categorisation algorithm to examine each of the coding units and allocate the unit to the model category with the highest fit probability. The training dictionary has fields for storing raw text, text with stop words removed, text

with stop and stemmed words, and a number indicating the category or level of the coding scheme. Table 5.1 shows the table structure of the dictionary table.

**Table 5.1**

**Table structure of the ACAT training dictionary table**

<b>Field Name</b>	<b>Data Type</b>
Unit ID	Medium Integer
Unit_raw_text	Blob
Unit_stop_text	Blob
Unit_stop_stem_text	Blob
Category_nos	Tiny Integer

The population of the ACAT training dictionary requires human intervention. Human coders have to enter phrases which correspond to different levels of the coding scheme being adopted. To enable the training dictionary to be populated a training module was developed. The interface to the training module is similar to the interface to the ACAT transcript coding module described in the previous chapter. The interface allows the user to enter text into a text box and save the entry (figure 5.3).

ACAT - QCA Model Training	
Data Source	transcript.txt
Analysis Type	QC Model Analysis
Taxonomy Type	Garrison et al.
Original Unit Text	
<div style="border: 1px solid black; padding: 5px;">           What is meant by the term quality of service         </div>	
Action	
<input type="button" value="Save Unit"/> <input type="button" value="ACAT Menu"/>	
ACAT - Automated Content Analysis Tool	

Text unit not yet saved....

**Figure 5.3: ACAT System Training Text Entry Screen**

Once the text is stored the text is re-presented with the stop words removed and with the words in their stemmed form (figure 5.4). The coder then allocates the phrase to the appropriate model category. Units from previously manually coded transcripts can be imported directly into the coding dictionary.

ACAT - QCA Model Training	
Data Source	transcript.txt
Analysis Type	QC Model Analysis
Taxonomy Type	Garrison et al.
Original Unit Text	
What is meant by the term quality of service	
Stop Words Removed Unit Text	
What meant term quality service	
Stop/Stemmed Unit Text	
What meant term qualiti servic	
Number of Levels	4
Select Level	
Level 1	<input type="radio"/>
Level 2	<input type="radio"/>
Level 3	<input type="radio"/>
Level 4	<input type="radio"/>
Unclassified	<input type="radio"/>
Action	
<input type="button" value="Save Unit"/> <input type="button" value="Next Unit"/> <input type="button" value="ACAT Menu"/>	
ACAT - Automated Content Analysis Tool	

**Figure 5.4: ACAT System Training Text Allocation Screen**

Figure 5.5 illustrates how the various components of the ACAT system interact. The heart of the system is the automatic coding process which takes processed coding units one at a time from the Parsed table and compares them with all the records in the training/dictionary table, calculating the probability that the coding unit belongs in a specified category. The automatic coding process relies on a text categorisation algorithm which calculates the probability of a category match, based on the Naive Baye's classifier. When all the coding units have been processed the results are presented using the ACAT system results screen. Once the automatic coding process is completed the ACAT system allows the automatically produced results to be compared with the results produced by human coders and automatically calculates correlations between the two. Similar to the initial ACAT system, both Holsti's CR and Cohen's kappa are calculated.

## ACAT Automated System Conceptual Model

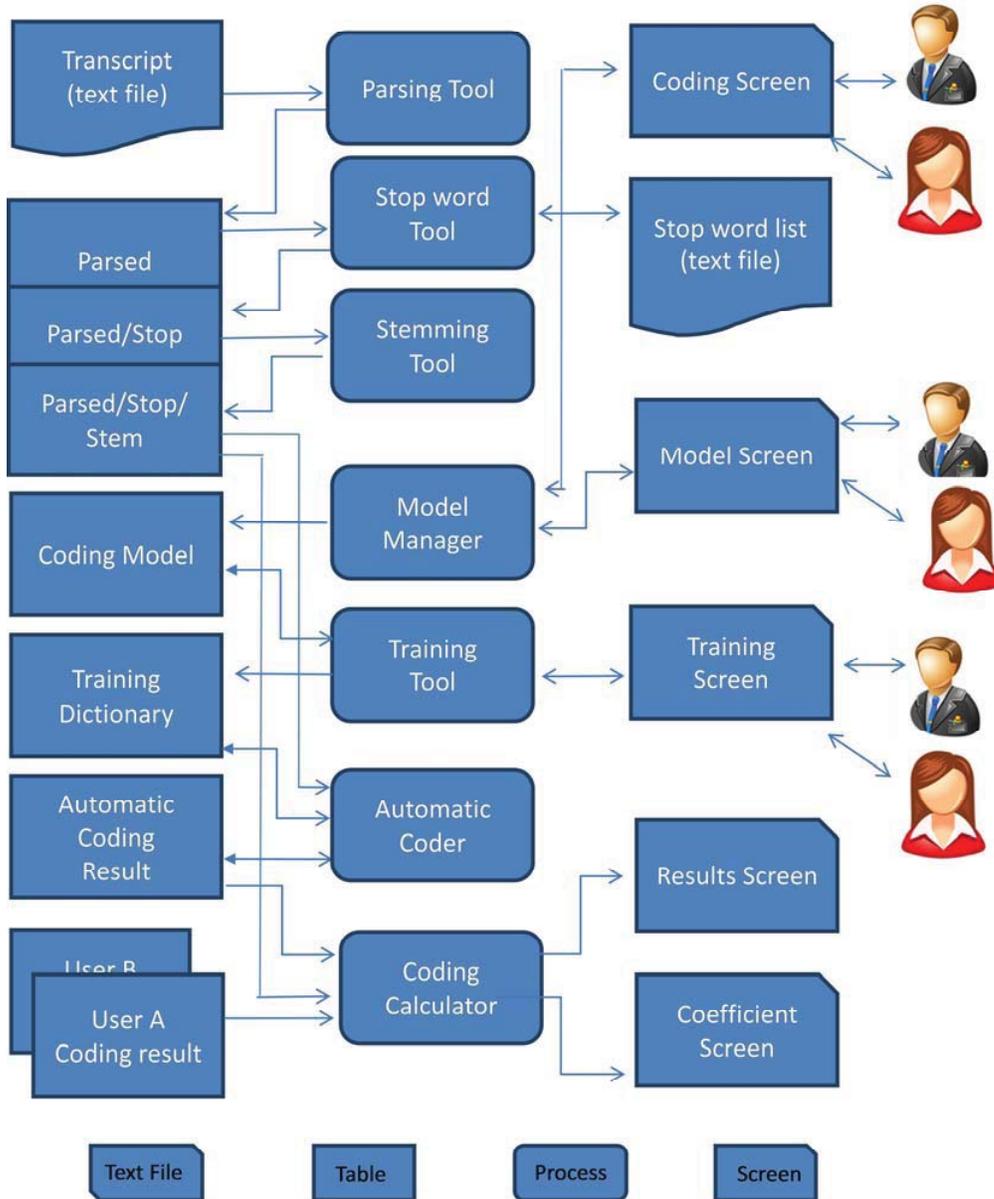


Figure 5.5: ACAT Automated System Conceptual Model

## 5.5 Application of the Automated ACAT System

The second iteration of the ACAT system provided a number of additional features designed to automate the coding process. The automated ACAT system parsed the discussion forum transcript, removed stopwords and reduced words in the coding units to their stems. The system also allowed users to build a training dictionary which enabled the system to categorise automatically the coding units extracted from an imported transcript. In addition, the system allowed the automatically produced results to be validated against those produced by human coders. The system was used to code the transcript described in the previous chapter and the results were compared with those obtained by the human coders.

### 5.5.1 Methodology

The methodology adopted for this phase of the study combines action research with content analysis. As described in the methodology chapter, each cycle of the action research includes a peer review process which acts as the catalyst for the redevelopment of the ACAT system and shapes the direction of the ongoing research. Unlike the previous chapter where the content analysis was conducted by human coders, the content analysis conducted in this phase of the research is fully automated.

The transcript resulting from the assessed discussion described in the previous chapter was used as the basis for this phase of the study and the results of the content analysis obtained using the automated ACAT system are compared with the results of the two human coders.

Having designed and built the training dictionary and automatic coding components of the ACAT system, the training dictionary needed to be populated. Time and coder resource limitations dictated that the training dictionary was built for one critical thinking measurement model only. Since the COI model produced higher correlation coefficients and was thought to be simpler to code than the Hara et al. (2000) model a decision was made to build a training dictionary for the Garrison et al. model.

The two coders who conducted the coding described in the previous chapter set about populating the training dictionary. Using the training module of the ACAT system the two coders took turns at adding statements to each of the four categories of the model. The initial content was based on the statements made by students and coded in the first iteration of the action research described in the previous chapter. The coders also referred to the examination scripts obtained from the previous offerings of the data communications' course pertaining to the same topic discussed in the discussion forum. Over a period of two months in the second half of 2004, the two coders added statements to all four categories of the dictionary. At regular intervals the coding units from the transcript were compared with the training dictionary using the ACAT automatic coding module and the correlation coefficients between the manual coders and those produced by the ACAT system were compared.

### 5.5.2 Findings and discussion

The previous chapter described the demographics of the participants and the results obtained by the human coders using both the COI model and the Hara et al. model. This chapter described the results obtained using the ACAT system to classify automatically the same discussion forum transcript used in the previous chapter utilising the COI model.

As previously stated the training dictionary was populated over a period of two months, and the ACAT system was used to categorise automatically the units obtained from the discussion forum transcript. Table 5.2 shows the results obtained using the ACAT system at various times during the two month period. The table indicates the number of coding units stored in the training dictionary and the number and percentage of units classified by the system into each of the four categories of the COI coding model at four different occasions.

**Table 5.2****Posting categorisation for different training dictionary populations**

Dictionary Units	1,000		2,500		3,500		7,000	
	Number of Units	Percent						
Triggering	117	24.17%	104	21.49%	90	18.60%	78	16.12%
Exploration	130	26.86%	134	27.69%	139	28.72%	128	26.45%
Integration	126	26.03%	158	32.64%	186	38.43%	218	45.04%
Resolution	111	22.93%	88	18.18%	69	14.26%	60	12.40%
Total	484	100.00%	484	100.00%	484	100.00%	484	100.00%

When the dictionary size was approximately two times that of the sample size, the ACAT system results were close to those that would have been obtained using a classifier that randomly classified units. As the size of the dictionary increased, the performance of the ACAT classifier improved. The results obtained using the ACAT system suggest that as the size of the dictionary increased the ability of the system to categorise discussion forum posting units improved. This supports the findings of earlier studies using the Naive Baye's classifier that suggested that there is a relationship between the dictionary size and the effectiveness of the classifier (Khan et al., 2010; Nigam, Mccallum, Thrun, & Mitchell, 2000).

Table 5.3 shows the results obtained using the ACAT system when coding the raw text unit, the text unit with the stop word removed and the text unit with stop word removal and word stemming implemented. The table also compares the results obtained using the ACAT system with the results obtained by the two human coders.

It should be noted that the ACAT automated system did not leave any items uncategorised, whereas the human coders were unable to categorise almost 5% of the total. The algorithm which categorised the coding units, categorised on the basis of the greatest probability of fit in a specific category. An analysis of the categorised

postings identified that none of the 484 postings that were categorised had a matching probability that was less than the probability of a random occurrence in an individual category. The analysis of the posting categorisation process also identified a problem with the classification algorithm, when it categorised units with equal match probabilities. When the algorithm identified a sentence having identical probabilities in more than one category the algorithm automatically placed the sentence in the first of the categories in the processing loop.

**Table 5.3**

**ACAT (raw text, stopword, stemmed) and manual posting categorisations**

<b>Category</b>	<b>Automatic Coding (Raw Text)</b>	<b>Automatic Coding (Stopwords removed)</b>	<b>Automatic Coding (Stemmed text)</b>	<b>Manual Coder A</b>	<b>Manual Coder B</b>
Triggering	78 (16.1%)	80 (16.5%)	75 (15.5%)	73 (15%)	70 (14.5%)
Exploration	128 (26.5%)	129 (26.7%)	119 (24.6%)	124 (25.6%)	120 (24.8%)
Integration	218 (45%)	217 (44.8%)	225 (46.5%)	209 (43.2%)	218 (45%)
Resolution	60 (12.4%)	58 (12%)	65 (13.4%)	58 (12%)	49 (10.1%)
Not categorised	0	0	0	20 (4.1%)	27 (5.6%)
Total number of units	484 (100%)	484 (100%)	484 (100%)	484 (100%)	484 (100%)

While the percentages of sentences were similar across all categories for manual coding, raw text automatic coding, stopwords removed text automatic coding and

stemmed text automatic coding, the values for the coefficients of reliability did not indicate such a high level of coding correlation.

Table 5.4 shows the correlation (Holsti's CR and Cohen's kappa) between the results produced by the ACAT system and the results produced by the two human coders. The correlations are shown for the raw text comparison, the stopword removal comparison and the stopword and stemmed text comparison.

**Table 5.4**  
**Correlation between the ACAT system and the human coders**

	ACAT and Coder A		ACAT and Coder B	
	Holsti's CR	Cohen's kappa	Holsti's CR	Cohen's kappa
Raw text	0.64	0.61	0.63	0.61
Stopwords removed	0.65	0.63	0.64	0.62
Stopped and Stemmed	0.71	0.68	0.69	0.65

The correlation coefficients between the ACAT system and the manual coders improved slightly when the comparison was made using units with the stopwords removed. The correlation coefficients improved further when the comparison was made using units with both stopwords removed and words in their stemmed format. This result supports the claim made by Khan et al. (2010), that text classification can be improved by incorporating stopword removal and word stemming algorithms.

The correlation coefficients between the ACAT system using stopword removal and word stemming is lower than that calculated between the human coders described in the previous chapter. Since no evidence can be found of earlier research where both

Holsti's CR and Cohen's kappa have been calculated, it is difficult to make comparisons. There are however reports available where content analysis techniques have been used and Hoslti's Coefficient has been reported. The results for the ACAT system compare favourably with both the manual coding exercise described by Garrison et al. (2000) and the research conducted by McKlin et al. (2002) using a neural network automated system.

While the automated text classification process may not be quite as accurate as the manual coding process, the results were encouraging. The ACAT system was able to reduce the time spent categorising a transcript and was able to reduce the risk of coder subjectivity.

## 5.6 Peer review

Having built and tested the training dictionary and automatic text classification components of the ACAT system, the research findings were once again presented for national and international review. The second version of the ACAT system was demonstrated at an ALTRC PhD research symposium towards the end of 2005 and a paper describing the second phase of the action research cycle was presented at the Fifth International Conference on Networked Learning held at Lancaster University in England in April 2006 (Corich et al., 2006a).

The ALTRC meeting was attended by four Massey University staff members, a visiting international research professor, nine PhD students and six masters' students. Attendees were once again reminded that the presentation was being recorded and that a written transcript would be produced which would be used to inform the direction of future research. There was once again general agreement among those attending the meeting that the ACAT system appeared worthy of further investigation and once again advice was given concerning possible directions for future research and a number of questions were raised.

The questions raised at the ALTRC symposium included:

1. Will the system work with larger samples?
2. Will the system work for discussion forums that are discussing a different topic?
3. Is the success of the system subject domain related?
4. How can you be sure that the postings are not of a social nature?
5. Could the system work for individuals to identify an individual's critical thinking ability?

The web site for the International Conference on Networked Learning describes the Networked Learning Conference as an international, research-based conference, which since its inception in 1998, has developed a strong following by international researchers. It claims to be well supported by practitioners, managers and learning technologists interested in contributing to and hearing about research in the area of networked learning (Networked Learning, 2010). The fifth conference was held at Lancaster University, a research centre which is well known for its Centre for Studies in Advanced Learning Technology (CSALT), a centre that has a reputation for research in the field of technology enhanced learning (TEL) applied to adult education and training . The conference was attended by researchers from around the world, with the majority from the United Kingdom and Europe.

The conference paper presentation was attended by approximately 50 conference participants. Before the presentation commenced, the researcher explained that the paper which was being presented formed part of a PhD research project, to verify if it was possible to automate the identification of critical thinking among participants in discussion forums and that the presentation was being recorded. The researcher also

explained that he was using the conference presentation as part of a peer review process, to help inform the future direction of the study. The attendees expressed a great deal of interest in the ACAT system and suggested that the system definitely merited further research and development. A number of questions were raised, directed at the future plans for the development and testing of the ACAT system. Similar to the ALTRC meeting, a conference participant inquired if the system had the potential to identify critical thinking for an individual discussion forum participant. Perhaps the most challenging questions came from a member of the CSALT research team who asked:

1. How can you be sure that the model you have adopted actually measures critical thinking?
2. Has the critical thinking model been empirically validated?

The findings detailed in this chapter were also submitted as a paper to the *Journal of e-Learning and Knowledge Society*, which was published in March 2006 (Corich et al., 2006b). The review panel agreed that the area of critical thinking was an important aspect of e-learning and encouraged further research into the area of automating critical thinking measurement.

The feedback received from the ALTRC symposium, the Fifth International Conference on Networked Learning and the paper submission to the *Journal of e-Learning and Knowledge Society* helped direct the direction of the next phase of the ACAT system development described in the next chapter.

## 5.7 Conclusion

This chapter described the second phase of the action research cycle which included the development and testing of the ACAT dictionary training and automatic coding modules of the ACAT system. It described the rationale for the selection of the

coding classification system and the text classification algorithm, both of which are key components of the ACAT system. The chapter also described how the automatic coding module of ACAT system was used to automatically code a discussion forum transcript and how the results that the system generated were compared with the results of two human coders coding the same transcript. The chapter concluded by describing the peer review process used to inform the development of the next phase action research cycle.

The next chapter describes how the feedback received during the peer review process was used to inform the approach adopted for the next phase of the action research cycle. The chapter also explains how the ACAT system was used to code the transcripts of individual participants automatically in an attempt to establish if the system could be used to identify evidence of the critical thinking ability of individuals.



# CHAPTER 6: USING THE ACAT SYSTEM TO MEASURE INDIVIDUAL PARTICIPATION

---

Traditional thinking is all about "what is". Future thinking will also need to be about what can be.

Edward De Bono (1933 - )

The previous chapter described the planning, action and evaluation steps of the second phase of the action research cycle outlined in chapter 3. It explained the rationale for the selection of the coding classification system and the text classification algorithm, both of which are key components of the ACAT system. The chapter also described how the automatic coding module of ACAT system was used to code a discussion forum transcript automatically and how the results that the system generated were compared to the results of two human coders coding the same transcript. The chapter concluded by describing the peer review process used to inform the development of the next phase action research cycle.

This chapter addresses the issues that were identified in the previous cycle and attempts to answer some of the questions that were raised as part of the ongoing peer review process. The chapter explains how the ACAT system was once again used to code a discussion forum transcript automatically and how the system was then used to evaluate the critical thinking skills of individual discussion forum participants. The chapter also discusses the results obtained using the tool and compares them to those obtained using human coders. The chapter concludes with a summary of the findings of a third round of peer review and provides an indication of the questions that need to be addressed with the fourth iteration of the tool.

## 6.1 Background

The previous chapter described how the ACAT system was used to code the transcript of a discussion forum automatically. The results obtained using the tool were encouraging. When compared with the results produced by human coders, the coefficients of reliability between the ACAT system automated classifications and the classifications obtained using human coders were similar to those obtained by earlier researchers using only manual coding methods (Fahy, 2001; Garrison et al., 2000; Meyer 2004; Hara et al., 2000).

The peer review process which included presentations at the ALTRC research symposium and an international conference in England raised a number of questions which provided the catalyst for the research described in this chapter. The questions which have been addressed include:

1. Will the system work for discussion forums that are discussing a different topic?
2. Is the success of the system subject domain related?
3. How can you be sure that the postings are not of a social nature?
4. Could the system work for individuals to identify an individual's critical thinking ability?
5. How can you be sure that the model you have adopted actually measures critical thinking?
6. Has the critical thinking model been empirically validated?

In an attempt to address the first two questions, the ACAT system was used to classify the discussion forum transcript resulting from an online discussion forum of a group of students participating in a internet and web development course. While the subject area was still computing, the topic of discussion differed from that used in the study described in the previous chapter. The third question was addressed by having the human coders review the coding units before the automated analysis process to remove postings that were of a social or flippant nature and not related to the topic being discussed. Attempting to address the fourth question, which was raised at both the ALTRC symposium and the international conference, led to an investigation to see if the ACAT system could be used to identify evidence of critical thinking for individual discussion forum participants. The answer to the fifth question which relates to having confidence that the model being used actually measures evidence of critical thinking is addressed in the next chapter. The next section in this study attempts to address the final question, regarding the empirical validation of the model.

## 6.2 Validation of the Garrison et al. (2000) Model

Zelkowitz (2006) suggests that empirical validation is required for a model or hypothesis to gain acceptance in the research community. There are a number of ways that validation may be achieved, including; experimental design, case and field studies, duplication of results, peer review, adversarial review and conference presentations. Arbaugh (2008) suggests that while there are several models for measuring critical thinking in discussion forums, the Community of Inquiry model developed by Garrison, Anderson and Archer is by far the most popular. Google Scholar shows that the COI model has been cited in other works at least 894 times as of September 2010, making it by far the most cited article from the Internet and Higher Education journal. The model has been adopted and used on numerous occasions by researchers attempting to find evidence of social presence, teaching presence and critical thinking (Anagnostopoulos, Basmadjian, & McCrory, 2005; Arnold & Ducate, 2006; Meyer, 2004; Shea, Li, & Pickett, 2006; Garrison &

Cleveland-Innes, 2005; Heckman & Annabi, 2005; Oriogun, Ravenscroft, & Cook, 2005; Stodel, Thompson, & MacDonald, 2006; Arbaugh & Hwang, 2006; Richardson & Swan, 2003; Wise, Chang, Duffy, & del Valle, 2004; Arbaugh, 2008). In the majority of these studies researchers report similar findings, adding credibility to the model and providing evidence to support its validity.

The research conducted as part of this study and described in chapter 4, further validates the COI model. The results obtained from using the Garrison model were similar to those obtained using the model developed by Henri (1992) and adapted by Hara et al. (2000).

### 6.3 Second Application of the ACAT Automatic Classification System

In an attempt to address the questions relating to the ability of the ACAT system to identify evidence of critical thinking in an area other than data communications (i.e. a different subject domain) the ACAT automatic text classification system was applied to a discussion forum transcript obtained from a discussion forum activity associated with an internet and web development course. The transcript was parsed using the ACAT system and coded using the automatic text classification feature of the ACAT system. The ACAT system was also used to assist two human coders as they manually parsed the same transcript.

#### 6.3.1 Methodology

As with the previous phase, the methodology adopted for this phase of the action research uses content analysis. The review process at the end of the second phase of the action research cycle was the catalyst which helped inform the direction of this cycle of action research activity. Once again content analysis is used to classify the coding units into the categories recommended by the Garrison et al. (2000) cognitive process component of the COI model. To enable comparisons to be made with the

results obtained in the previous phase of the action research cycle, the sentence was used as the coding unit.

There are four different activities associated with this phase of the study. The first activity involves the manual categorisation of the coding units using the COI model, the second involves the automatic coding of the same transcript using the ACAT text classification module with the COI model. The third activity involves using the COI model to code the postings made by individual participants, the fourth activity involves participation in the peer review process.

The research associated with this phase of the study was conducted in the second semester of 2005 and involved a discussion forum activity undertaken by a second year undergraduate degree course at the tertiary education institution described in chapter 4. All the students were enrolled in a computing systems degree and as such were familiar with using information technology. The course was on internet and web development and was delivered using a blended learning environment, which combined traditional face-to-face activities with web publishing, on-line review and discussion forum activities. Participation in the discussion forum activity was a mandatory component and the activity was assessed with 20% of the assessment grade for the course being awarded for the quality and relevance of the student postings. The topic for discussion was "Web Scripting Tools". Students were given three weeks to participate in the discussion forum activity. Similar discussion forum activities had been used for the course in the two previous offerings in semester 2 of 2003 and semester 2 of 2004.

The software used to support the discussion forum was an integral part of the open source Moodle learning management system (LMS). All participants had previously used Moodle to retrieve course materials and to participate in on-line tests in their earlier courses. They had also participated in discussion forum activities using the Blackboard LMS.

The class consisted of sixteen students, five females and eleven males, aged between 17 and 42, of varying academic abilities. Students were given the topic for the discussion halfway through the course and instructions were given as a guide to encourage higher level critical thinking. By the time the discussion forum activity commenced, three of the original students had withdrawn from the course leaving only twelve participants (4 female and 8 male).

### 6.3.2 Findings and discussion

During the three week's period when students were expected to post to the forum, an instructor monitored postings on a daily basis. The instructor provided encouragement, added pedagogical comments and provided reinforcement and expert advice. As with the previous exercise, the proportion of postings made by female students (44% of the total postings) was higher than those made by their male counterparts. The seven students over the age of 25 accounted for 69% of the postings, however this time all of the participants actively participated in the discussion forum activity.

#### 6.3.2.1 Manual Coding for Group Participation

In total 115 posts were made, 35 of which were made by the course instructor. Once the instructor postings were removed, the remaining 80 posts were imported into the ACAT system and the parsing process generated 462 sentences for coding. Using the ACAT system to manually review and categorise the coding units, the two human coders agreed to remove 152 of the sentences which were viewed as being social in nature or not contributing to the discussion topic, leaving 310 sentences for analysis.

Both coders then used the ACAT system to manually code the sentences using the Garrison et al. (2000), COI model, categorising sentences into the triggering, exploration, integration or resolution categories. The coefficient of agreement between the two coders was 80% using Holsti's CR and 76% using Cohen's kappa. Table 6.1 shows the summary information on how the 310 sentences were classified.

**Table 6.1**  
**Number of postings using the COI model**

Category	Coder A		Coder B		ACAT System	
	Number of coding units	Percent of total coding units	Number of coding units	Percent of total coding units	Number of coding units	Percent of total coding units
1. Triggering	60	19.35%	58	18.71%	82	26.45%
2. Exploration	84	27.10%	86	27.74%	125	40.32%
3. Integration	111	35.81%	106	34.19%	78	25.16%
4. Resolution	55	17.74%	60	19.35%	25	8.06%
Total number of coding units	310	100.00%	310	100.00%	310	100.00%

The coefficients of agreement were very similar to those generated in the coding exercise described in chapter 4. Considering that in the coding exercise described in chapter 4 the coders had met and agreed on how to code more than one third of the total units, the result is better than expected. It should be noted that this time none of the units was placed in the uncategorised category, this is a consequence of the removal of the units that were considered not contributing to the discussion topic prior to the coding exercise.

Compared with the previous exercise the proportion of "triggering" (approximately 19% compared with 15%) was higher and the same could be said for the proportion of "resolution" (approximately 18% compared with 11%) categorisations. As with the last exercise, the majority of units were coded as belonging to the "exploration" and "integration" categories. Having higher proportions in both the lower and higher levels of critical thinking would suggest a wider range of critical thinking abilities among the participants, however unless a way can be found to identify individual levels of critical thinking there is no way of verifying this statement.

### 6.3.2.2 Automatic Coding for Group Participation

Having completed the manual coding process, the 310 coding units resulting from the removal of the instructor postings and the removal of the postings that were considered to be social in nature or not contributing to the discussion topic, were automatically coded by the ACAT system. The ACAT system used the same training dictionary that was used in the previous coding exercise, a dictionary that was built to categorise postings for the data communications' discussion.

The results obtained using the ACAT system to categorise the postings resulting from the "Web Hosting Environments" discussion topic are shown in table 6.1. The coefficients of reliability comparing the automatically coded results with the results obtained by the two human coders are shown in table 6.2.

The results produced by the ACAT system, using the training dictionary populated in the previous discussion forum study, suggest that the change in discussion topic has had a negative impact on the ability of the system to effectively categorise the postings.

**Table 6.2**  
**Correlation between the ACAT system and the human coders**

ACAT and Coder A		ACAT and Coder B	
Holsti's CR	Cohen's kappa	Holsti's CR	Cohen's kappa
0.43	0.41	0.42	0.40

The correlation coefficients which are approximately 0.4 are significantly lower than the previous exercise using the ACAT automated coding system. According to Capozzoli et al. (1999), correlations below 0.4 represent poor agreement beyond chance, while values greater than 0.75 represent excellent agreement. In this case the agreement level is only slightly better than the expectation for a chance result. The result suggests that the training dictionary has a bias that favours the topic that it was trained to categorise. Since there is little evidence of similar automated coding activities within the existing literature, it is difficult to judge how this finding compares with previous research. The finding does suggest that there is an opportunity for further research into the production of training dictionaries that are not subject specific.

Having obtained such a disappointing result, the coders re-evaluated the contents of the training dictionary and noticed that almost half the units in the dictionary referred to items which have a data communications' focus, the majority of which were imported following human coding exercises.

The study described in the previous chapter provided evidence that reducing the dictionary size impacts on the ability of the text classification system to classify text automatically, so rather than strip out all the references that have a data communications' focus a decision was made to increase the number of generic coding units and add units to the dictionary that have an internet and web development focus. This was achieved by importing transcripts from discussion forum transcripts obtained from two previous offerings of the course and manually coding the transcript using the training module of the ACAT system so that they can be added to the training dictionary. The human coders also populated the training dictionary by manually entering and coding units directly to the dictionary.

Over a two-months period the two human coders doubled the size of the training dictionary. The new training dictionary was then applied to the ACAT system to automatically code the transcript used with the 2005 semester 2 offering of the course. The results of the automated coding exercise and the two manual coders are shown in table 6.3.

**Table 6.3**  
**Posting categorisation for automatic and manual coding**

<b>Category</b>	<b>Automatic Coding</b>		<b>Manual Coder A</b>		<b>Manual Coder B</b>	
Triggering	51	16.5%	60	19.35%	58	18.71%
Exploration	80	25.8%	84	27.10%	86	27.74%
Integration	126	40.6%	111	35.81%	106	34.19%
Resolution	53	17.1%	55	17.74%	60	19.35%
Total number of units	310	100%	310	100.00%	310	100.00%

The ACAT automatically coded results placed a greater proportion of coding units in the “integration” category than the manual coders, but produced similar results to the human coders in the other three categories.

Table 6.4 shows the correlation (Holsti’s CR and Cohen's kappa) between the results produced by the ACAT system and the results produced by the two human coders. The average correlation between the human coders and the ACAT system is 0.65, a correlation that Capozzoli et al. (1999) describe as a strong agreement.

**Table 6.4**  
**Correlation between the ACAT system and the human coders**

ACAT and Coder A		ACAT and Coder B	
Holsti's CR	Cohen's kappa	Holsti's CR	Cohen's kappa
0.65	0.63	0.67	0.65

The results of the automatic coding process in this activity of the action research cycle suggests that the effectiveness of the text categorisation system is dependant on the size and subject focus of the training dictionary. The findings suggest that a training dictionary populated for a specific subject area is unlikely to be effective at coding transcripts from a different subject area. The findings also suggest an avenue for further research investigating the possibility of creating a generic training dictionary that is effective at coding transcripts from multiple subject areas.

### 6.3.2.3 Automatic Coding of Individual Participation

The third activity associated with this phase of the study attempts to answer the question relating to the ability of the ACAT system to measure the critical thinking abilities of individual discussion forum participants. The postings for each of the 12

individual participants were exported from the Moodle LMS and were loaded one at a time into the ACAT system.

Both coders used the training module of the ACAT system to manually code the coding units for each individual. When both the coders had completed the exercise, they met and reviewed all of the postings, discussing the postings that had been classified differently and reclassifying them into mutually agreed categories. When the manual coding process had been completed, the ACAT system was used to code the coding units for the individual participants automatically.

Table 6.5 represents the individual participant's critical thinking classifications, as agreed by the human coders and as coded by the ACAT system.

The correlations for individual postings of the ACAT system and the human coders was 0.69 for Holsti's CR and 0.67 for Cohen's kappa, suggesting that the ACAT system has the potential to provide useful information about the critical thinking activities of individual participants within a discussion forum.

Table 6.6 represents the percentages for individual participants critical thinking classifications against each category of the COI model, as agreed by the human coders and as coded by the ACAT system.

**Table 6.5**  
**Individual participant classification (number of sentences)**

	Sentences	Triggering (Human Coder)	Triggering (ACAT System)	Exploration (Human Coder)	Exploration (ACAT System)	Integration (Human Coder)	Integration (ACAT System)	Resolution (Human Coder)	Resolution (ACAT System)
1	24	4	5	6	6	9	8	5	5
2	22	4	4	5	6	9	8	4	5
3	28	5	6	7	7	11	9	5	6
4	30	4	5	8	8	12	10	6	7
5	26	5	5	7	7	10	9	5	5
6	34	5	7	10	10	14	11	5	6
7	28	5	5	7	8	11	10	5	5
8	26	4	5	6	7	10	10	3	4
9	26	4	6	7	7	11	9	4	4
10	28	4	5	8	8	11	10	5	5
11	16	3	4	4	4	6	5	3	3
12	24	4	5	6	7	10	9	4	3
Average	26	4.25	5.12	6.75	7.08	10.33	9.00	4.50	4.833
Standard Deviation	4.43	0.62	0.83	1.54	1.44	1.92	1.54	0.90	1.93

**Table 6.6**  
**Individual participant classification (as a percentage)**

	Sentences	Triggering (Human Coder)	Triggering (ACAT System)	Exploration (Human Coder)	Exploration (ACAT System)	Integration (Human Coder)	Integration (ACAT System)	Resolution (Human Coder)	Resolution (ACAT System)
1	24	16.7%	20.8%	25.0%	25.0%	37.5%	33.3%	20.8%	20.8%
2	22	18.2%	18.2%	22.7%	27.3%	40.9%	36.4%	18.2%	22.7%
3	28	17.9%	21.4%	25.0%	25.0%	39.3%	32.1%	17.9%	21.4%
4	30	13.3%	16.7%	26.7%	26.7%	40.0%	33.3%	20.0%	23.3%
5	26	19.2%	19.2%	26.9%	26.9%	38.5%	34.6%	19.2%	19.2%
6	34	14.7%	20.6%	29.4%	29.4%	41.2%	32.4%	14.7%	17.6%
7	28	17.9%	17.9%	25.0%	28.6%	39.3%	35.7%	17.9%	17.9%
8	26	16.7%	19.2%	25.0%	26.9%	41.7%	38.5%	12.5%	15.4%
9	26	15.4%	23.1%	26.9%	26.9%	42.3%	34.6%	15.4%	15.4%
10	28	14.3%	17.9%	28.6%	28.6%	39.3%	35.7%	17.9%	17.9%
11	16	18.8%	25.0%	25.0%	25.0%	37.5%	31.3%	18.8%	18.8%
12	24	16.7%	20.8%	25.0%	29.2%	41.7%	37.5%	16.7%	12.5%
Average	26	15.5%	18.7%	25.9%	27.1%	39.9%	34.6%	17.5%	18.6%
Standard Deviation	4.43	1.90%	2.39%	1.80%	1.58%	1.60%	2.22%	2.30%	3.20%

An unexpected result for both the human coded results and the ACAT system results was that all participants appear to operate at similar levels of critical thinking with both sets of results appearing to show only minor differences in the levels of critical thinking between individuals. A Chi square test was conducted across each of the categories for the results obtained by the human coder. When comparing the observed result in each category with the average result for each category (which would be the expected result assuming no differences existed) the Chi square test indicated insignificant differences between observed and expected results for all the categories. Conducting a similar test across each of the categories for the automatically coded results also indicated insignificant differences between observed and expected results.

When the lecturer responsible for teaching the class was asked if all students exhibited similar levels of critical thinking in face-to-face discussions, he suggested that he would have expected the critical thinking abilities of individuals within the class to have varied. He explained that he would have expected participants to display significant differences in levels of critical thinking abilities. The estimation of critical thinking abilities of participants were based on observations over time, no formal means of measurement had been applied.

The similarities in critical thinking activities between individuals could be explained in part by the fact that students were given a marking rubric which indicated how postings would be graded and that it appeared that most of them followed the instructions. Another factor that could help account for the apparent similarities in critical thinking activities, could be related to the design of the actual COI model itself. The COI model assumes

that learning occurs within the Community through the interaction of three core elements; cognitive presence, social presence, and teaching presence. (Garrison et al., 2000 p. 88).

This study has considered only the cognitive presence aspects of the model, ignoring the teaching presence and the social presence. Concentrating on the cognitive presence aspects only of the model could impact on the results achieved by both the human coders and the ACAT system. It should also be noted that the COI model has been specifically designed to focus on the "critical thinking within a group dynamic as reflected by the perspective of a community of inquiry" (Garrison et al. 2000, p. 11). When the aim of a study is to examine evidence of critical thinking of a group as a whole, then concentrating on the group dynamic is appropriate. The same approach might not be as relevant when the focus is on the individual member of the community (Perkins & Murphy, 2006).

It is interesting to note that a study conducted by Perkins and Murphy in 2006, using a model specifically designed to measure evidence of critical thinking by individual discussion forum participants was able to distinguish differences in levels of critical thinking. The model was based on models presented by earlier researchers (Norris & Ennis, 1989; Henri, 1992; Clulow et al., 2003; Garrison et al., 2000; Newman et al., 1995; Bullen, 1997) and was designed with a focus on individual participant contribution. Using their model, Perkins and Murphy (2006) were able to identify different levels of critical thinking, which they felt were a reflection of the critical thinking abilities of the individual participants.

The unexpected result produced using the COI model, where individuals appeared to exhibit similar levels of critical thinking, suggest that while the model appears able to identify evidence of critical thinking of a group of participants, it is not an appropriate tool for measuring individual contribution. The promising results achieved by Perkins and Murphy (2006), indicate that their model might provide a more fitting tool worthy of further research.

## 6.4 Peer review

Having used the ACAT system to code automatically a transcript obtained from a discussion forum discussing a different subject area from that described in the last chapter, then using the system to attempt to measure the critical thinking of individuals, the research findings were again presented for national and international review. The results described in this chapter were discussed at a number of online meetings of a group of Doctoral and Masters students towards the end of 2006. A paper describing the third phase of the action research cycle was presented at the IADIS International Conference on Cognition and Exploratory Learning in Digital Age held in Portugal in July 2007 (Corich et al., 2007a).

The online symposia were arranged to provide an avenue for people who had been part of the ALTRC community to continue to discuss their research after the ALTRC ceased to meet face-to-face. Like the previous face-to-face ALTRC meetings, the online meetings involved a number of post graduates students who were studying within New Zealand and overseas. The purpose of the online meetings was to provide an opportunity for researchers to share their findings and seek advice and direction. The online meeting participants provided encouragement for the further development of the ACAT system, in particular the possibility of using it to inform individual student's feedback on their level of critical thinking when participating in a discussion forum. Participants once again gave advice on how the ACAT tool could be used and a number of questions were raised. The questions included:

1. Will the system work with larger samples?
2. Will the system work with individual students?
3. How could you validate the results for individual participants?
4. What are the alternate ways of measuring the critical thinking of individuals?

5. Could the system be used to inform students about their ability to think critically?
6. Could the system be integrated into a LMS so that it could provide constructive feedback to enable participants to improve their critical thinking?

The aim of the IADIS International Conference on Cognition and Exploratory Learning in the Digital Age was to address the main issues concerned with evolving learning processes and supporting pedagogies and applications in the digital age. The conference web-site suggests that the conference sought to address issues such as just-in-time learning, constructivism, student-centred learning and collaborative approaches that are being supported by technological advancements such as simulations, virtual reality and multi-agents' systems. The paper describing the development and testing of the ACAT system, which attempts to automate the process of identifying evidence of critical thinking, was double-blind refereed and the reviewers commented that the paper provided an interesting insight in the area of combining cognitive psychology with computing.

The conference paper presentation was attended by more than 70 conference participants from all around the world. Before the presentation commenced the researcher explained that the paper which was being presented formed part of a PhD research project aimed at attempting to verify if it was possible to automate the identification of critical thinking among participants in a discussion forums. The researcher also explained that he was using the conference presentation as part of a peer review process, to help inform the future direction of the study. The presentation attendees expressed interest in the ACAT system suggesting that the area of critical thinking was an essential element of the current educational system and that the ACAT system merited further research and development. The questions asked were similar to those already raised during the online ALTRC discussion, with most interest being shown in the ability of the system to measure evidence of critical

thinking for individual participants and the need to validate by alternate means any individual participant measurements.

The findings detailed in this chapter were also submitted as a book chapter which appeared in an edited volume which was published in 2007 (Corich et al., 2007b). Like the previous journal article the reviewers encouraged further research into the area of automating critical thinking measurement stating that it was an exciting development worthy of further research.

The feedback received from the online post-graduate discussion, the IADIS International Conference on Cognition and Exploratory Learning in Digital Age and the book chapter submission helped inform the next phase of the use of the ACAT system described in the next chapter.

## 6.5 Conclusion

This chapter described the third phase of the action research cycle which included the testing of the ACAT system using the transcript from a subject area that was different from that used in the earlier study. The chapter discussed how the ACAT system was used to attempt to measure evidence of critical thinking for individual participants. It described the unexpected result, which indicated little variance between the individuals.

As with the previous chapter, this chapter concluded by describing the peer review process used to inform the development of the next phase action research cycle.

The next chapter describes how the model developed by Perkins and Murphy (2006) was used with the ACAT system in an attempt to measure the critical thinking of the individual discussion forum participants. The chapter also explains how the results obtained using the Perkins & Murphy model were validated using a traditional critical thinking measurement tool.



# CHAPTER 7: VALIDATING THE ACAT SYSTEM USING THE PERKINS & MURPHY MODEL

---

The world we have created is a product of our thinking; it cannot be changed without changing our thinking.

Albert Einstein.

The previous chapter described the planning, action and evaluation steps of the third phase of the action research cycle outlined in chapter 3. It explained how the ACAT system was used to analyse the transcript obtained from an internet and web development course and how the system was used with the same transcript to analyse the postings of individual participants. The chapter concluded that while the Garrison et al. COI Model was able to identify levels of critical thinking for a group of discussion forum participants, it appeared to have difficulty identifying differences in the levels of critical thinking for individuals. An alternate model specifically designed to work with individuals was proposed as having the potential for ongoing investigation. The chapter concluded by describing the peer review process undertaken to inform the direction of the next phase of the action research cycle.

This chapter, identified as the last phase (phase 4) of the action research cycle, considers the questions that were raised as part of the ongoing peer review process. It describes how the ACAT tool was used with the Perkins and Murphy model to identify critical thinking activities for individual participants. The chapter explains how the results obtained using the ACAT system were validated by comparing them with the results obtained using a more traditional critical thinking measurement tool. The chapter concludes with a summary of the findings and suggests areas for further research.

## 7.1 Background

The previous chapter indicated that the COI model, which appeared to be able to identify evidence of critical thinking of discussion forum participants when considering the group as a whole, was not such an effective tool when it came to measuring evidence of critical thinking for individual participants. The model developed by Perkins and Murphy, designed to measure the critical thinking abilities of individual participants, appears to provide an alternative that might be more effective. The same model has also been used in subsequent studies by Jacob & Sam (2008, 2010) and by Murphy (2009).

The peer review process, which included online discussion with a number of Doctorate and Masterate level students and a presentation at the IADIS International Conference on Cognition and Exploratory Learning in Digital Age in Portugal, raised a number of questions, which combined with the findings of study described in the previous chapter provide the catalyst for the research described in this chapter. The questions which were raised and have been considered, include:

1. Will the system work with individual students?
2. How could you validate the results for individual participants?
3. What are the alternate ways of measuring the critical thinking of individuals?
4. Could the system be used to inform students about their ability to think critically?

The two remaining questions:

1. Will the system work with larger samples?
2. Could the system be integrated into a LMS so that it could provide constructive feedback to enable participants to improve their critical thinking?

will be discussed in the concluding chapter of this thesis.

## 7.2 The Perkins and Murphy (2006) Model

The Perkins and Murphy (2006) model is described as being specifically designed to measure evidence of critical thinking for each of the individual participants contributing towards a discussion forum. Like the COI model the model is based upon four critical thinking indicators. When developing their model, Perkins and Murphy examined a number of existing models all of which had been designed to measure evidence of critical thinking of the discussion forum participants as a whole. The existing models that influenced the design of the Perkins and Murphy model included models developed by Norris & Ennis (1989), Henri (1992), Clulow et al. (2003), Garrison et al. (2000), Newman et al. (1995) and Bullen (1997).

The first category of the Perkins and Murphy model, clarification, was common to all models except the Garrison et al. COI model. The second category, assessment, was found only in the Bullen (1997) model. The third category, inference and the fourth category, strategies were common to all but the COI model. Perkins and Murphy describe clarification as including everything involved in proposing, describing and defining an issue. They describe assessment as covering various types of judgment, including the using of evidence to support or refute a judgment. Inference is said to cover the thinking skills, which include; induction, deduction and generalizing. The strategies category is said to include practical proposals for dealing with the issues under discussion. Having identified the four categories, Perkins and Murphy

identified a number of contextually based indicators which could be used when allocating transcript components into one of the four categories.

Since the Perkins and Murphy model was the only available model designed specifically to measure evidence of an individual's critical thinking it was adopted for the study as a mechanism for identifying the critical thinking skills of the individual discussion forum participants.

Having identified a model that could be used with the ACAT system to evaluate evidence of critical thinking of individual discussion forum participants, it was necessary to address how the results produced could be validated. Comparing the results produced by the ACAT system with those produced by a human coder would validate the ability of the ACAT system to reproduce the human coders results; however such a comparison would not validate the ability of the Perkins and Murphy (2006) model to actually measure critical thinking. To address this validation issue, alternate ways of measuring critical thinking for individuals were investigated.

The literature review identified three commonly used assessment techniques that have been adopted to measure the ability of students to think critically (Paul & Elder, 2001; Facione, 2007). The three main approaches that were identified were:

- commercially available standard tests;
- researcher/instructor designed assessment tools built around a particular course or research project;
- encouraging students to assess their own thinking.

Since this study has concentrated on tools which belong to the second category and no evidence can be found of any other tools which have been specifically designed to measure evidence of individual participation of discussion forum participants a

decision was made to adopt a tool from the first category to validate the Perkins and Murphy (2006) model.

The literature review identified a number of commercially available standard tests, and suggested that one of the most widely used tests is the Cambridge Thinking Skills Assessment (TSA) (Fisher, 2005; Forster, 2004). A practice version of the Cambridge TSA is available online, and since it is widely used and freely available it was chosen as the preferred method of validating the results obtained from using the Perkins and Murphy (2006) model.

### 7.3 The Cambridge TSA

According to the Cambridge Assessment website, the Cambridge TSA was developed in 2001 as an entrance test for certain programmes offered at Cambridge University. The Cambridge TSA is now licensed around the world and is used as an admission exam by many colleges and universities. The Cambridge TSA consists of 50 multiple choice questions and is used to measure critical thinking and problems solving skills. The problem solving skills include numerical and spatial reasoning and the critical thinking skills include understanding arguments and reasoning using everyday language. The TSA uses a multiple-choice format and allocates 1 mark per question. Scores are calibrated onto a scale (running from approximately 0-100), with scores being reported to one decimal place. The scale is not a percentage based on raw scores, but a statistical estimate of a candidate's ability. The average TSA score of an applicant to the University of Cambridge is typically in the high 50s, with only around 10% of applicants scoring over 70 (Cambridge Assessment, 2008).

The origins of the Cambridge TSA can be traced back to the early 1990s and the NEMO project, a project named after Nemo, a student of Socrates who was said to have prompted Socrates to propound his theory of knowledge as recollection. The aim of the NEMO project was to develop a test that could identify the thinking skills that are crucial to success in Higher Education. The project was a collaborative

exercise involving Universities, Polytechnics and Colleges of Higher Education within the United Kingdom and across Europe. The developments arising from the NEMO project led to the creation of the Cambridge TSA as a mechanism for testing critical thinking (Chapman, 2005).

Since introducing the Cambridge TSA in 2001, the University of Cambridge has conducted a number of studies to show that the TSA can be a successful predictor of critical thinking abilities and its use has expanded throughout the university (Cambridge Assessment, 2008; Fisher, 2005; Forster, 2004). Currently 27 of the 29 colleges at Cambridge University now require applicants to take the TSA for a number of courses, including Natural Sciences, Engineering, Computer Science and Economics.

#### 7.4 Using the ACAT system with the Perkins and Murphy Model

A model dictionary was prepared for the Perkins and Murphy (2006) model and added to the ACAT system. The dictionary was built using the phrases from the dictionary used with the Garrison et al. model and described in chapter 6. The phrases were re-coded in accordance with the indicators outlined in Table 7.2. The initial coding was done by both the coders identified in the previous chapters. Having coded the first 400 units, the two coders examined each of the units and discussed the coding decisions. Particular attention was paid to the units which had been coded differently; the coding decision for each of these was discussed and the units recoded to a mutually agreed category. Having decided on an agreed coding approach, each of the coders took turns coding the remaining units. Shortly after completing the dictionary coding process one of the coders took up a position overseas and was no longer available to assist with the final manual coding process described later on in this chapter. This meant that the only remaining coder was the researcher who is the author of this thesis. It was decided, that since it would take a significant amount of time to find and train a new coder to assist with the final coding exercise, that the coding would be conducted by the researcher.

The model was then applied to the transcript of each individual participant and the ACAT system produced a report for each participant showing the number of sentences coded against each category of the Perkins and Murphy (2006) model.

To test the validity of the ACAT system, a human coder was employed to code the transcripts of each individual participant against each category of the Perkins and Murphy (2006) model. Correlation coefficients were calculated to compare the results from the two systems.

To further test the validity the results of the ACAT system using the Perkins and Murphy (2006) model, the individual participants were asked to sit the practice version of the Cambridge TSA. The results obtained from the Cambridge TSA were compared with those obtained using the ACAT system.

#### 7.4.1 Methodology

The methodology for this phase of the study once again included content analysis within the phase of an action research cycle. The findings of the previous chapter combined with the questions raised as part of the peer review process drove the research described in this chapter. The content analysis involves both human coder and automated system, this time classifying coding units into categories developed by Perkins and Murphy (2006). Validation of the automated system is provided by comparing the results of the ACAT system coding with that of a human coder coding the same transcript. Validation of the critical thinking model is undertaken by comparing the results of the automated coding with those obtained using the Cambridge TSA.

The discussion forum transcripts used in this study were obtained from a discussion forum used in a third year undergraduate web development course which took place in the first semester of 2007. The transcripts were obtained with ethical approval of

the institute and involved sixteen students, aged between 18 and 36, and of varying academic abilities. All participants were informed of the research activity and given a research information sheet (copy provided at Appendix C) and were asked to sign a consent form indicating agreement to participate in the study (copy provided at Appendix D). The course was delivered using a blended learning environment, combining traditional face-to-face activities with web publishing, on-line review and discussion forum activities. The discussion forum activity was assessed and given a 15% weighting for the final assessment allocation. Students were asked to discuss “Web Hosting Environments”; they were informed that they would be expected to demonstrate aspects of critical thinking in their posts. Prior to the commencement of the discussion forum activity, students were given a marking rubric that indicated the type of activities that would be recognized as contributing to the four different levels of critical thinking (clarification, assessment, inference and strategies).

The software used to support the discussion forum was an integral part of the Moodle learning management system, which allows the discussion forum transcripts to be exported and individual participant's transcripts to be identified. The transcript exported from Moodle had to be manually parsed to separate the individual contributions and to create individual text files for each participant. The individual text files were then imported into the ACAT system and results were obtained for each individual participant. All students had previously used the Moodle learning management system and most students had participated in an assessed discussion forum earlier in their studies.

#### 7.4.2 Findings and discussion

During the three week period when students were expected to post to the forum, an instructor monitored postings on a daily basis. The instructor provided encouragement, added pedagogical comments and provided reinforcement and expert advice. Of the 16 students, only 12 participated in the discussion and then went on to participate in the critical thinking test. One hundred and forty-two student posts were

made which generated 436 sentences for coding. The resulting transcripts were reviewed by a human coder who removed 148 of the sentences which were assessed as being social in nature or not contributing to the discussion topic, leaving 288 sentences for manual and automatic analysis.

The coding units for each individual participant were imported one at a time into the ACAT system; the human coder then used the ACAT training module to manually code the sentences using the Perkins and Murphy (2006) model, where coding units are categorised as belonging to one of the clarification, assessment, inference and strategies categories. The coding units for each individual were then automatically coded by the ACAT system, and individual reports for each of the participants were produced.

Following the completion of the discussion forum exercise the 12 students who had participated in the discussion forum were asked to complete the Cambridge TSA practice test.

Table 7.1 shows the results of the analysis of the 288 coding units obtained from the discussion forum transcripts after the social and non-contributing sentences had been removed. The table displays the number of coding units allocated to each category of the Perkins and Murphy (2006) model for each of the 12 individual participants, as agreed by both the human coder and as calculated by the ACAT system.

**Table 7.1**  
**Individual participant classification using the Perkins and Murphy (2006) model (number of sentences)**

Participant	Sentences	Clarification		Assessment		Inference		Strategies	
		Human Coder	ACAT System	Human Coder	ACAT System	Human Coder	ACAT System	Human Coder	ACAT System
1	18	5	5	5	6	4	3	4	4
2	26	8	8	6	7	7	6	5	5
3	19	5	5	5	5	7	7	2	2
4	21	5	6	6	5	6	5	4	5
5	24	7	7	7	6	7	8	3	3
6	28	9	8	11	11	4	6	4	3
7	30	8	9	9	8	6	7	7	6
8	19	6	6	5	4	6	7	2	2
9	27	9	9	10	9	5	5	3	4
10	25	7	7	7	6	6	7	5	5
11	31	10	8	10	10	6	7	5	6
12	20	6	6	6	6	6	7	2	1

When the individual transcripts were combined and treated like a group, the coefficients of reliability between the manually coded results and the automatically coded results were calculated. The coefficients of reliability comparing the automatically coded results with the results obtained by the human coders were 0.67 using Holsti's CR and 0.65 using Cohen's kappa.

The correlation coefficients compare favourably with those of the previous two cycles using the ACAT system suggesting that the ACAT system was able to produce results similar to those produced by the human coder.

Table 7.2 represents the number of coding units allocated to each category of the Perkins and Murphy (2006) model for each of the 12 individual participants, as agreed by the human coder and as calculated by the ACAT system shown as a percentage.

**Table 7.2**  
**Individual participant classification using the Perkins and Murphy (2006) model (percentages)**

Participant	Sentences	Clarification		Assessment		Inference		Strategies	
		Human Coder	ACAT System	Human Coder	ACAT System	Human Coder	ACAT System	Human Coder	ACAT System
1	18	28	28	28	33	22	17	22	22
2	26	31	31	23	27	27	23	19	19
3	19	26	26	26	26	37	37	11	11
4	21	24	29	29	24	29	24	19	24
5	24	29	29	29	25	29	33	13	13
6	28	32	29	39	39	14	21	14	11
7	30	27	30	30	27	20	23	23	20
8	19	32	32	26	21	32	37	11	11
9	27	33	33	37	33	19	19	11	15
10	25	28	28	28	24	24	28	20	20
11	31	32	26	32	32	19	23	16	19
12	20	35	30	30	30	30	35	5	5

Unlike the study described in the previous chapter, which used the COI model, this study indicates differences in levels of critical thinking between different individual participants. A Chi square test was conducted across each of the categories for the results obtained by the human coder. When comparing the observed result in each category with the average result for each category (which would be the expected assuming no differences existed) the Chi square test indicated insignificant differences between observed and expected results in the clarification and the inference categories, but significant differences in the assessment and strategies categories. Conducting a similar test across each of the categories for the automatically coded results once again showed insignificant differences between observed and expected results for the clarification and the inference categories, but significant differences in the assessment and strategies categories. This tends to support the findings made by Perkins and Murphy (2006) that the model they had developed to specifically measure evidence of critical thinking for individuals provided a better indicator of an individual's critical thinking than the COI model.

Similar to the results obtained by Perkins and Murphy (2006) the group as a whole tended to engage more in clarification, assessment and inference, and less in strategy. Looking at the individual results there appears to be no indication of a relationship between the number of postings and evidence of higher or lower cognitive activity.

The 12 participants who completed the discussion forum posting exercise were asked to sit the practice version of the Cambridge TSA. The participants were asked to complete reminded that participation was optional and they were asked if they would like to know the results of the assessment. All 12 participants indicated that they would like to know their scores and they were then given 90 minutes to complete the assessment exercise. When the participants had completed the assessment, the assessment was marked using the marking schedule obtained from the Cambridge Assessment website and the resulting scales were recorded for each of the individuals.

Table 7.3 shows the results for both the Cambridge Thinking Skills Assessment demonstration test and the cumulative ACAT scores for the same individuals that participated in the discussion forum activities.

**Table 7.3**  
**Cambridge Thinking Skills Assessment demonstration test and cumulative ACAT score results**

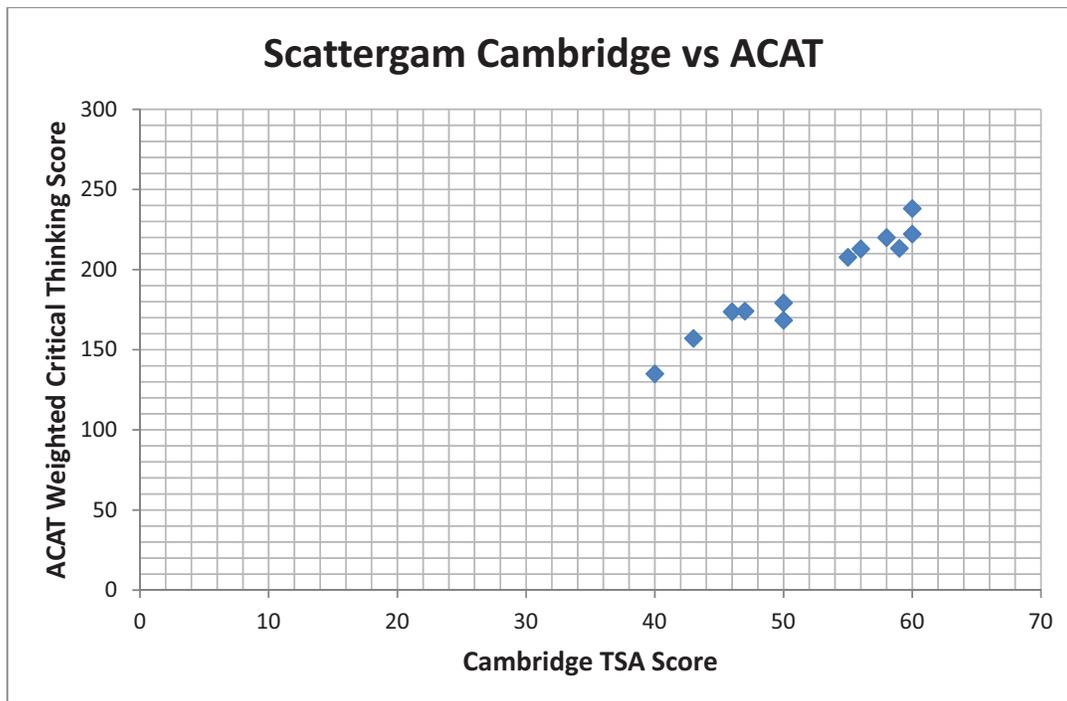
Participant	Cambridge Skills Assessment Score	Cumulative ACAT Score
1	60	222.2
2	55	207.7
3	46	173.7
4	60	238.1
5	50	179.2
6	43	157.1
7	59	213.3
8	50	168.4
9	47	174.1
10	58	220
11	56	212.9
12	40	135
Mean	52	191.4

The scores from the Cambridge Thinking Skills Assessment demonstration test suggest that the participants in the study differed from each other in their critical thinking skills abilities. When compared with the scores reported at the Cambridge Assessment website, where the average of a Cambridge applicant is in the high 50s,

the average score was 52, suggesting that the participants demonstrated slightly lower critical thinking and problem solving skills than their Cambridge counterparts.

The results obtained from the manual coding and automated coding of transcripts also indicated different levels of critical thinking. Since the results from the manual coding and the ACAT coding do not indicate an overall critical thinking score which can be compared with the Cambridge Thinking Skills Assessment results, an algorithm was used to calculate a score that recognises higher levels of critical thinking for the ACAT system produced results. The algorithm applies weightings to each of the four categories (lower weightings for the lower level categories and higher weightings for the higher level categories), multiplies the percentages in each category by the weightings and totals the result to arrive at a cumulative score for each participant.

A comparison of the results of the Cambridge TSA scale and the ACAT weighted average score, suggests a correlation between the two. To test for evidence of correlation between the two sets of data a scattergram chart was produced (see Figure 7.1). The chart reinforces the suggestion that there is a strong correlation between the two sets of results.



**Figure 7.1: Scattergram of Cambridge Thinking Skills Assessment score and the ACAT system weighted score**

Having established that a relationship existed between the Cambridge Thinking Skills Assessment score and the ACAT system weighted score a statistical test was conducted to determine the strength of the correlation. Two of the most commonly used measures of correlation between two sets of data are the Pearson's product moment correlation coefficient and the Spearman's rank correlation coefficient. Since the two data items are measures of physical quantities that can be ranked and the two variables are not jointly normally distributed, the Spearman's rank correlation coefficient was selected as the most appropriate measure of correlation or association for the study.

The Spearman correlation coefficient was 0.961 ( $p < 0.01$ ) suggesting a strong correlation between the scores for the Cambridge Thinking Skills Assessment and the ACAT system.

The purpose of this study was to evaluate whether an automated content analysis tool could be used to identify levels of critical thinking by individual participants in a discussion forum and to validate the results against a traditional measure of critical thinking. Since the sample size was small, generalisations cannot be made on the basis of the findings. The results however do suggest that the use of automated tools for identifying levels of critical thinking in discussion forums is worthy of further research. The automated tool was able to produce results that are similar to those obtained by human coders and also comparable with results obtained using more traditional critical thinking measurement tools.

The strong correlation between the results obtained using the ACAT system with the Perkins and Murphy (2006) model and the results obtained using the Cambridge Thinking Skills Assessment (Spearman coefficient = 0.961,  $p < 0.01$ ) suggest that the ACAT system provides an indication of critical thinking that is similar to that produced by the Cambridge Thinking Skills Assessment.

The study adds credibility to the Perkins and Murphy (2006) model. The model proposed by Perkins and Murphy, tested with a human coder and the ACAT system does appear to be able to differentiate different levels of critical thinking for individuals participating in discussion forums. The close correlation between the results obtained using the Perkins and Murphy model and the results obtained using the Cambridge Thinking Skills Assessment provide strong evidence to validate the model.

The validity of the results obtained from analysing the discussion forum transcripts could be questioned, since students participating in the study were given a rubric indicating how levels of critical thinking can be measured and they were informed

that their discussion forum participation was being tested for evidence of critical thinking. Knowing how critical dispositions towards critical thinking were being measured could have influenced the way that students contributed towards the forum discussion. However, the comparisons with the results obtained using the Cambridge Thinking Skills Assessment would suggest that the impact of informing the participants had little effect. The strong correlation between the two systems would suggest that the provision of the marking rubric had minimal impact on the level of critical thinking identified.

#### 7.4 Peer Review

Having used the ACAT system to code automatically a transcript obtained from a discussion forum using the Perkins and Murphy (2006) model having compared the results with those obtained by a human coder and with those obtained for the same group of participants using the Cambridge TSA, the research findings were presented for international review. A paper describing this fourth phase of the action research cycle was submitted for the IADIS International Conference on Cognition and Exploratory Learning in Digital Age held in Rome, Italy in November 2009 (Corich et al., 2009). The paper describing the use of the ACAT system to code a discussion forum transcript using the Perkins and Murphy model and validation of the model using the results obtained from the Cambridge TSA was double-blind refereed. The reviewers stated that the paper provided an interesting and timely update on the previously presented work.

The conference paper was accepted and the presentation was attended by more than 50 conference participants who once again expressed interest in the ACAT system and suggested that the area of critical thinking was a vital component of today's educational setting. Very few questions were asked, but those that were asked related to the future plans for the ACAT system and possibility of applying the tool to areas other than discussion forum transcripts.

The feedback received from the IADIS International Conference on Cognition and Exploratory Learning in Digital Age will be discussed in the final chapter.

## 7.5 Conclusion

This chapter described the fourth and final phase of the action research cycle which included the testing of the ACAT system using a model designed by Perkins and Murphy (2006) and specifically designed to measure evidence of critical thinking for individual discussion forum participants. The chapter also described how the students that participated in the discussion forum completed the Cambridge TSA and how the results were used to validate the findings of the Perkins and Murphy study. The chapter concluded by suggesting that the ACAT system appeared to have been successful in measuring evidence of critical thinking for individuals and that the tool has the potential for further development.

The next chapter is the final chapter in this thesis and it summarises the action research project and discusses the significance of the findings of the project. The final chapter also reviews the original research questions and explains to what extent they have been addressed. The chapter closes by discussing the limitations associated with the study and looks at the potential for further research.



# CHAPTER 8: SUMMARY, CONCLUSIONS, LIMITATIONS AND FUTURE DIRECTIONS

---

By wisdom a house is built, and through understanding it is established; through knowledge its rooms are filled with rare and beautiful treasures.

Proverbs 24:3-4 (NIV).

The previous chapter described the last phase (phase 4) of the action research cycle. It considered the issues related to the apparent inability of the COI model to measure evidence of critical thinking for individual discussion forum participants and attempted to address the questions that were raised as part of the ongoing peer review process. The chapter explained how the coding dictionary was adapted so that it could be used with the Perkins and Murphy's (2006) model and how the ACAT tool was used to identify critical thinking activities for individual participants. The chapter also described how the students participating in the discussion forum activity completed the Cambridge TSA and how the results obtained were used to test the validity of the Perkins and Murphy (2006) model. The chapter concluded with a summary of the findings and a description of the final peer review process.

This chapter summarises the activities that were undertaken as part of this study and attempts to address the research questions that were identified in chapter 1. The chapter starts with an overall summary of the research and then discusses the limitations associated with the study. This concluding chapter also looks at the potential for further research in the area of automating the measurement of critical thinking with discussion forums. The chapter concludes with a personal reflection on the research journey that has been a significant part of my life for the last seven years.

## 8.1 Background

The previous chapter provided evidence that the model developed by Perkins and Murphy (2006) to measure individual engagement in critical thinking in online discussion forums has the potential as a tool to aid educationalists. The successful adoption of the model within the ACAT system and its subsequent use to analyse the transcripts of individual participants suggest that the combination of the model and the automated tool have the potential to reduce the time spent undertaking manual content analysis. The use of the Cambridge TSA to measure evidence of critical thinking abilities and the comparison of the results with those obtained using the ACAT system incorporating the Perkins and Murphy (2006) model suggest that both the automated tool and the model could help identify evidence of critical thinking. The previous chapter described the last cycle of a four phase action research framework, which was adopted to attempt to address the research questions that were the basis of this study.

Chapters 4, 5, 6 and 7, described the activities undertaken within the action research framework, identified in chapter 3. The principal aim of this research project was to answer the following two questions:

- 1. Is it possible to automate the identification of critical thinking ability exhibited by a group as they participate in a discussion forum?*
- 2. Is it possible to automate the identification of an individual participant's critical thinking ability in a discussion forum?*

This chapter is the conclusion of the thesis, it provides answers to those wishing to address the questions raised in the research and provides advice for those who may be interested in conducting further research in the area of automating the identification of evidence of critical thinking in online discussion forums.

## 8.2 Summary of the Research Activities

The study adopted a qualitative action research framework as its methodology; a research framework that is popular amongst researchers concerned with activities that involve the development and testing of computerised information systems (Baskerville & Wood-Harper, 1996; Checkland, 1991). The action research framework had four cycles of activities, each of which involved peer review, planning an activity, undertaking the activity and evaluating the activity. Within each cycle, a content analysis approach was used to classify the text document of the discussion forum transcripts against one or more of the existing models developed to identify evidence of critical thinking.

The study commenced with a literature review that attempted to place the research in context by presenting an overview of critical thinking and the use of discussion forums in the educational setting. The literature review traced the roots of critical thinking back to the early Greek philosophers and found a lack of consensus on how critical thinking is currently defined in the literature. There is however general consensus on the cognitive skills and affective dispositions of critical thinkers, and researchers looking for evidence of critical thinking have tended to concentrate on identifying these skills and dispositions. The literature review established that considerable energy has been directed towards identifying evidence of critical thinking and several researchers have attempted to identify evidence of critical thinking among discussion forum participants. A number of models were identified as having been developed by researchers to classify the text obtained from discussion forum transcripts into different categories of critical thinking attributes. The most widely referenced model was the COI model developed by Garrison et al. (2000). The literature review also identified evidence of researchers using computerised tools to analyse discussion forum transcripts; however, the amount of research in this area was very limited.

Following the literature review and identification of the appropriate research methodology the first cycle of the research framework was undertaken. Since the literature review identified that the COI model was the most widely referenced and used model for measuring evidence of critical thinking in discussion forums, the ACAT system was developed to incorporate the model. Chapter 4 described how the ACAT system was developed to help two human coders code a discussion forum transcript which was an assessed component of an undergraduate data communications course.

The system automatically parsed the transcript, identifying and displaying sentences which the human coders could then categorise. The results of the categorisation process were saved allowing the system to calculate the coefficients of reliability between the two coders' results. The coefficients of reliability (Holsti's and Cohen's) were comparable with earlier studies and the results obtained indicated lower overall levels of critical thinking among participants. Having demonstrated that the ACAT system could assist coders with coding using the COI model, the exercise was replicated, this time using a model developed by Henri (1992) and refined by Hara et al. (2000). Comparing the results of the two models, both indicated similar levels of critical thinking and the coefficients of reliability were again similar to those achieved by earlier researchers.

Having completed the initial study using two existing models, the findings were shared with both national and international audiences and the feedback helped inform the second cycle of the research framework. The second cycle of the action research saw the development and testing of the ACAT system. The system was developed by adopting a computational content analysis approach that utilised a coding dictionary. Text classification of the coding units against the coding dictionary was achieved by adopting the Naive Bayesian classifier. Both the coding dictionary and the Naive Bayesian classifier approaches were chosen because they are widely accepted as text classification tools and their lack of sophistication made the building of the

computational tool a task within the software building capabilities of the researcher. A coding/training dictionary was populated by two human coders and the system was used to categorise the text described in the first cycle of activity. The model adopted for the categorisation process was the COI model. The study was able to demonstrate that the accuracy of the system was improved as the dictionary size increased and when compared with the results of the human coders, the ACAT system was able to achieve coefficients of reliability comparable with those achieved by pairs of human coders.

On completion of the first trial of the fully automated ACAT system, the findings were shared with both national and international audiences and once again the feedback helped inform the next cycle of the research framework. The third action research cycle sought to identify whether the ACAT system using the COI model could identify evidence of critical thinking for individual discussion forum participants and whether the coding dictionary approach could be applied across a different subject area.

Using the dictionary from the first automated test, the ACAT system was used to categorise the transcript obtained from a group of students enrolled in a web development course. The correlation between the results obtained by human coders compared with the results obtained from the ACAT system was slightly higher than 0.4 for both correlation coefficients. An examination of the coding dictionary revealed that almost half of the units within the dictionary were subject oriented and related to the data communications' field. As a result the two human coders spent time adding to the coding dictionary, using units obtained from previous discussion forum transcripts related to the web development subject area. Using the expanded dictionary the ACAT system was able to achieve correlations close to 0.66, a result that would be classified as acceptable for human coder comparison.

Once the ACAT system had been used to code the transcripts for the web development course successfully, the human coders both coded the coding units for

each of the individual participants and on completion met to agree a combined coding classification. By providing separate transcripts for each individual the ACAT system was used to code the individual transcripts. The correlation coefficients for the combined results for the human coders and the ACAT system were close to 0.68. When the categorisation figures were compared, both the human coded results and the automatically produced results indicated very little difference in critical thinking abilities between the individual participants. Conducting a Chi square analysis over both data sets also indicated insignificant variance between the critical thinking abilities of the individuals. When asked if the result was expected, the course lecturer suggested that he would have expected some noticeable variation in critical thinking abilities.

On completion of the third cycle of the action research framework, the findings were once again shared with both national and international audiences and once again the feedback helped inform the next cycle of the research framework. The final action research cycle sought to establish whether a model developed specifically to measure individual engagement in critical thinking could identify differences in critical thinking between individuals. The model adopted for this phase of the study was developed by Perkins and Murphy (2006). The first task was to recode the coding dictionary for the new model. This was done by the two human coders. Once the coding dictionary had been populated, one of the coders then went on to code the transcripts for each participant individually, using the Perkins and Murphy model. The same individual transcripts were applied to the ACAT system and the system produced individual categorisation reports for each participant. The correlation between the human coder and the ACAT system averaged 0.66 and this time a difference in the individual classifications was apparent.

Having no way of testing the validity of the results obtained using the Perkins and Murphy model, other than looking at the correlation between human and machine coders, a decision was made to ask the discussion forum participants to undertake a Cambridge TSA so that a comparison of results could be made. The Cambridge TSA

results were compared with those obtained from the ACAT system and applying a simple weighted average algorithm. Using a Spearman's Rank Correlation test a correlation coefficient of 0.961 was obtained, suggesting a strong correlation between the results obtained using the ACAT system with the Perkins and Murphy model.

### 8.3 Conclusions

The aim of the research presented in this thesis was to attempt to address the apparent gap in the literature relating to the use of technology to automate the process of identifying evidence of critical thinking among discussion forum participants. The two research questions that were identified as needing to be addressed were:

*Is it possible to automate the identification of critical thinking ability exhibited by a group as they participate in a discussion forum?*

*Is it possible to automate the identification of an individual participant's critical thinking ability in a discussion forum?*

The literature indicated that researchers have been actively involved in measuring critical thinking skills since the early 1970s . A number of different measurement techniques have been used, ranging from commercially designed and publicly available tests to teacher and researcher developed rubrics and tests. The validity of the commercial tests are widely accepted, although evidence of debate does exist concerning exactly which component of critical thinking is being measured. A number of researchers have reported success in using teacher and researcher designed tools, although there is insufficient evidence to suggest that the results obtained have been empirically validated.

There are numerous examples of researchers who have developed models to measure evidence of the critical thinking abilities of discussion forum participants; the majority of them involve using QCA (Henri, 1992; Norris & Ennis, 1989; Newman et

al., 1995; Bullen, 1997; Gunawardena et al., 1997; Clulow et al., 2003; Garrison et al., 2000; and Fahy et al., 2001; Perkins & Murphy, 2006). Most of the models have been tested by the developers and in most cases validated using multiple coders. Unfortunately there is little evidence to suggest that many of the models have subsequently been adopted by others and re-tested to provide further validation of their critical thinking measurement abilities. There is also little evidence to suggest that researchers have attempted to compare the results of the different models to see if they come to the same conclusions about the critical thinking abilities of discussion forum participants.

The response to the first research question;

*Is it possible to automate the identification of critical thinking ability exhibited by a group as they participate in a discussion forum?*

appears to have a positive response.

The literature review identified a study conducted by McKlin (2004), who developed an automated system that used artificial neural network (ANN) software to measure evidence of critical thinking of discussion forum participants. When comparing the coding results achieved by his automated system with those achieved by human coders he was able to achieve correlation coefficients of 0.7 and he concluded that his automated system could classify text with near human accuracy.

The results described in this study using the ACAT system would also suggest that within the limitations of the study, the system was able to produce results that were comparable with those achieved by human coders. When comparing the results achieved using the ACAT system with those achieved by human coders, the resulting correlation coefficients ranged from 0.61 to 0.71. The results were achieved when attempting to measure evidence of critical thinking of the participants when considered as a group. The correlation coefficients in this study are similar to those

reported by McKlin (2004), suggesting once again an automated system had been able to classify text with near human accuracy.

The results of this study also suggest that the success of the automated process is dependent on a number of factors, including the size and quality of the coding dictionary. It should also be noted that the text classification process was not completely automated, as postings that were considered as social in nature or not contributing to the discussion topic were removed by human coders before the ACAT system was applied.

The response to the second research question;

*Is it possible to automate the identification of an individual participant's critical thinking ability in a discussion forum?*

also appears to have a positive response.

While the literature review did not reveal any evidence of researchers attempting to use an automated system to measure individual engagement in critical thinking in discussion forums, there was evidence of researchers manually coding transcripts to measure individual engagement. Perkins and Murphy (2006) developed the model that was used in this study. Subsequent studies using the Perkins and Murphy model, such as the ones conducted by Jacob and Sam (2008, 2010), also relied on the use of manual coders.

A coding dictionary based on the Perkins and Murphy (2006) model was built for the ACAT system and the system was able to achieve correlation coefficients that averaged 0.66, which is accepted as being an acceptable level for human coder comparisons. Perhaps the most compelling evidence, for the success of the automation of the identification of critical thinking ability for individuals participating in a discussion forum, comes from the comparison of the results of the

ACAT system using the Perkins and Murphy (2006) model with the results obtained from applying the Cambridge TSA, where using a Spearman's Rank Correlation test a correlation coefficient of 0.961 was obtained.

#### 8.4 Limitations of the Study

The study described in this thesis was exploratory in nature and because of the limitations associated with the study, caution is needed while attempting to generalise the findings. While the answers to the first two research questions are based on evidence uncovered during the literature review process, the answers to the last two questions are based on the findings of this research, which, it should be noted, are based on small sample sizes from students enrolled in a single qualification in a single institution. The aim of the research was not to make findings that could be applied to the general population; rather, it was to investigate whether an automated tool has the potential to measure evidence of individual engagement in critical thinking in online discussion forums. The findings should encourage further research in the area of automated text classification and the identification of the evidence of critical thinking, and help provide evidence of validation for the findings of previous researchers.

The samples used in this study were chosen as an audience of convenience. Since the author did not have local access to a research team and he did not work in an institute that had a strong research culture, almost all the research activity, including software development, content analysis and data analysis was conducted by the author. The only additional support came from a colleague who assisted with transcript coding and helped build the coding dictionaries used in the ACAT system to support the automated classification process. The courses that were used to test the critical thinking analysis models and the ACAT system were courses that were readily accessible and delivered by colleagues who were happy to support the research goals. While the sample sizes (15, 12 and 12 students) were small, they are comparable with the sample sizes used by many of the earlier studies (Hara et al., 2000 used 20

students; Bullen, 1997 used 13 students; Fahy et al., 2000 used 13 students; Rourke et al. 2001 used 11 - 14 students; Garrison et al., 2000 used 11 students; Perkins & Murphy, 2006 used 8 students). The most significant reason for the small sample sizes is the nature of the content analysis process which Rourke et al. (2001, p. 12) describe as being “difficult, frustrating, and time-consuming.” Since manual coding was used to validate all the applications of the ACAT system, sample sizes and the resulting number of units available for coding could not be large.

The study was conducted at a single tertiary institute; the participating students were all involved in courses that belonged to a single programme. The programme was an undergraduate degree programme in Computer Systems, chosen because it was easily accessible to the researcher. While the decision to choose students who were enrolled in a computing programme, could be said to have provided a sample with a bias towards technology adoption, it did avoid problems associated with research subjects who may have been technophobic and required training in the use of discussion forums.

Since all of the software development associated with the study was undertaken by the author, a number of decisions relating to the design of the ACAT system were made on the basis of programming ability and the time available for development. The decision to adopt the sentence as the coding unit was made on the basis that it was relatively simple to build a system that divided a transcript based on grammatical syntax. DeWever et al. (2006) point out that the choice of unit of analysis is dependent on context and should be well considered. The choice of coding unit may affect coding decisions and make comparisons with earlier studies problematic. Similarly, the decision to adopt the Naive Bayesian classifier as the preferred classification algorithm for the ACAT system was influenced by the fact that it is relatively easy to implement with computer software. Having raised both of these issues it should be noted that earlier researchers have adopted the sentence as the coding unit (Fahy et al., 2000; Pena-Shaff & Nicholls 2004) and the Naive Bayesian

classifier is a well-proven and popular text classification algorithm (Khan et al., 2010; Pedersen, 2000; Ramoni & Sebastiani, 2001).

The study suggests that size of the coding dictionary does impact on the ability of a text classification system to classify text. This finding is supported by Khan et al. (2010) and Joachims (2009). The size of the dictionaries used with the ACAT system were very much dependent on the time availability of the two coders who built and coded the dictionaries used with the COI model and the Perkins and Murphy (2006) model.

The study was limited to analysing the discussion forum transcripts of two closely related subject areas. Even though the data communications and web-development areas have a lot in common, the study indicated that the differences are significant enough to impact on the effectiveness of the coding dictionary as a tool to classify text. Resource and time limitations associated with the study did not allow time to be spent investigating the potential of developing a generic dictionary that could possibly work over a large number of different subject areas.

The validity of the ACAT system using the Perkins and Murphy (2006) model was tested using the Cambridge TSA practice test. Part of the reason for selecting the Cambridge TSA was that the test was freely available and relatively simple to implement. Even though the test is widely accepted and widely used, it is designed to test both critical thinking and problem solving skills. The test has been validated by a number of studies (Fisher, 2005; Forster, 2004). For this study, the researcher was able to establish that the Cambridge TSA was an effective measure of critical thinking ability.

Having developed the ACAT system, it should be noted that the system is not completely automated. While the system has reduced many of the time-consuming activities associated with content analysis it still requires human intervention to operate effectively. The system relies on a coding dictionary, which requires human

intervention to populate. The dictionary population process is not only time-consuming, but it also needs human coders to make the categorisation decisions. Early in the study, a reviewer asked about the ability of the system to identify and remove postings that were social in nature or not relevant to the discussion topic. Rather than attempt to automate this aspect of the coding process, a decision was taken to use human coders to manually remove such postings prior to the automatic coding process.

## 8.5 Contributions

By verifying that an automated tool can be used to identify evidence of critical thinking among discussion forum participants, this study has added to the existing body of knowledge and built on a similar study conducted by McKlin et al. (2002). The use of the COI model and the Henri's (1992) model in the first cycle of the action research cycle provided evidence of successful human transcript coding, adding to the credibility of both models by validating their ability to produce similar measures of evidence of critical thinking.

The second action research cycle validated the COI model for a second time, and the automatic coding exercise using the ACAT text classification system provided additional evidence of the ability of the Naive Bayesian classifier to classify text by calculating the probability that a post matches a category within a coding dictionary.

The third action research cycle provided additional evidence of the relationship between the ability of a text classifier and the size of the coding dictionary, supporting the claims made by Khan et al. (2010) and Joachims (2009).

The fourth action research cycle helped validate the model developed by Perkins and Murphy (2006) and provided evidence of one of the first recorded attempts to use an automated tool to measure individual engagement in critical thinking in an online discussion forum. The use of the Cambridge TSA to validate both the automated tool

and the Perkins and Murphy model is also an activity that has not appeared in previously published research. The final research activity provides evidence that automated tools like the ACAT system have the potential to reduce the time and energy required to perform content analysis and may provide a useful mechanism to inform instructors and students about the critical thinking process.

## 8.6 Implications and Potential for Further Research and Development

The study provides compelling evidence of the potential for the use of computerised systems to identify evidence of critical thinking among discussion forum participants. The ACAT system was able to classify text and make coding decisions that were similar to those made by human coders. Even though the system was not tested with large samples or across a wide range of subject disciplines, findings indicate that the system merits further research and development.

Automated text classification systems, like the ACAT system have the potential to provide indicators of the evidence of critical thinking among discussion forum participants. Both instructors and students could benefit from using a system similar to that developed in this study. Instructors would have the potential to understand better the critical thinking abilities of their students; they would also be able to track the effect of any strategies that they may adopt to encourage the development of critical thinking skills. By being better informed about the relative critical thinking abilities of their students, instructors could focus their support on learners who exhibit low levels of critical thinking or they could provide extension activities for those who appear to be at the higher end of the scale.

If a system like the ACAT system were incorporated into a learning management system (LMS), it could provide feedback directly to students, indicating the level of critical thinking that they exhibit as they contribute to discussion forum activities. This would potentially allow students to self-assess their critical thinking abilities,

enhance their understanding of the role of critical thinking and allow them to develop their own improvement strategies. The literature review identified that teaching students to assess their own thinking is the most appropriate way to assess students' critical thinking (Reed, 1997; Elder & Paul, 1994). A system like the ACAT system, if incorporated within a LMS has the potential to allow students to assess their own levels of critical thinking routinely, and with the help of their instructors develop strategies to improve the quality of their thinking.

The testing of the ACAT system was limited to small samples and was only applied to courses that were part of an undergraduate computing degree. Future studies could be directed at seeing how the system copes with larger samples involving greater numbers of postings. Future studies could also examine how the system is able to identify evidence of critical thinking across a number of different subject areas. The problems identified with the subject specific nature of the coding dictionary could also be further investigated, and solutions could be investigated aimed at developing a generic coding dictionary that works successfully across a wide range of subjects.

This study has limited its investigation to discussion forums, where the use of automated systems to measure evidence of critical thinking appears to have merits. Today there are a number of places where students make electronic postings, including blogs and wikis. The potential exists to evaluate the use of a system like the ACAT system in a number of different environments.

The findings described in this study were limited to a single New Zealand tertiary institution. Similar studies could be conducted using the ACAT system at a variety of institutions within New Zealand and overseas. Studies across multiple sites would enable comparisons in critical thinking between institutes and across year levels.

In its current form, transcripts from individual participants must be imported into the ACAT system one at a time and the analysis results exported individually. If the system were to be used to identify evidence of individual engagement in critical

thinking among discussion forums, it will require modification to separate the individual participant transcripts automatically and provide individual reports. If the ACAT system was to be used as a mechanism to provide feedback about individual engagement in critical thinking to instructors and/or students, significant development work would be required. Since the system was developed using an open source web based approach, integration into an existing open source LMS like Moodle should be a relatively straightforward development task.

## 8.7 Concluding Remarks

The study described within this thesis explored the potential of using computer technology to identify evidence of individual engagement in critical thinking among discussion forums. An action research framework was chosen as the methodology to carry out development of the automated tool and test its effectiveness. Within each of the action research cycles, content analysis was chosen as the methodology to perform the text classification required to make the tool operate.

The study, which commenced with a literature review, showed that while there are several models that have been designed to measure evidence of critical thinking in discussion forums, there was little evidence of using computers to automate the process and no evidence of using computers to identify evidence of individual engagement in critical thinking.

As the study progressed, a number of the models designed to accommodate human coders were tried and validated, and finally a model designed by Perkins and Murphy (2006) was used with an automated system to measure evidence of how individuals engage in critical thinking.

Having completed the study, it is hoped that it will provide a catalyst for further research into the area of automating the measurement of evidence of critical thinking in discussion forums. It is also hoped that it may lead to the development of a

computerised tool that could be integrated into a LMS and provide a mechanism that will allow students to assess their own levels of critical thinking routinely, then with the help of their instructors students could develop strategies to improve the quality of their thinking.



# 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.
- Adler, M. J. (1982). *The paideia proposal: An educational manifesto*. New York: Macmillan.
- Akyol, Z., & Garrison, D. R. (2011). Understanding cognitive presence in an online and blended community of inquiry: Assessing outcomes and processes for deep approaches to learning. *British Journal of Educational Technology, 42*(2), 233-250.
- Anagnostopoulos, D., Basmadjian, K. G., & McCrory, R. S. (2005). The decentered teacher and the construction of social space in the virtual classroom. *Teachers College Record, 107*, 1699-1729.
- Anderson, T., & Kanuka, H. (1997). On-line forums: New platforms for professional development and group collaboration. (*ERIC Document Reproduction Service, ED 418 693*).
- Andrade, H.G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership, 57*(5). Retrieved from [http://www.smallschoolsproject.org/PDFS/coho103/using\\_rubrics.pdf](http://www.smallschoolsproject.org/PDFS/coho103/using_rubrics.pdf)
- Arbaugh, J. B. (2008). Does the Community of Inquiry framework predict outcomes in online MBA courses? *International Review of Research in Open and Distance Learning, 9* (2). Retrieved from <http://www.irrodl.org/index.php/irrodl/article/view/490/1048>
- Arbaugh, J. B., & Hwang, A. (2006). Does “teaching presence” exist in online MBA courses? *The Internet and Higher Education, 9*(1), 9-21.
- Archibald, D. (2010). Fostering the development of cognitive presence: Initial findings using the community of inquiry survey instrument. *Internet & Higher Education, 13*(1-2), 73-74.
- Arnold, N., & Ducate, L. (2006). Future foreign language teachers’ social and cognitive collaboration in an online environment. *Language Learning & Technology, 10*(1), 42-66.

- Bailin, S., & Siegel, H. (2003). Critical thinking. In N. Blake, P. Smeyers, R. Smith, & P. Standish (Eds.), *The Blackwell guide to the philosophy of education* (pp. 181–193). Oxford, UK: Blackwell.
- Barratt, J. (2009, August 10). *A Plea for more critical thinking in design, please*. Retrieved from <http://www.fastcompany.com/blog/john-barratt/design-day/plea-more-critical-thinking-design-please>
- Baskerville, R. (1997). Distinguishing action research from participative case studies. *Journal of Systems and Information Technology*, 1(1), 25-45.
- Baskerville, R.L., & Wood-Harper, A.T. (1998). Diversity in information systems action research methods. *European Journal of Information Systems*, 7(2), 90-107.
- Berelsen, B. (1952). *Content analysis in communications research*. New York: Free Press.
- Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W.H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain*. New York: David McKay Co. Inc.
- Bourner, T. (1996). The research process: four steps to success. In T. Greenfield (Ed.), *Research methods: guidance for postgraduates*, London: Arnold.
- Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of Research in Education*, 24, 61.
- Brazell, A. (2003). *The forum moderator's guide to life - parts 1 and 2*. Retrieved 9 June 2010 from <http://articles.sitepoint.com/article/guide-life-parts-1-2>
- Brook Wu, Y., & Chen, X. (2005) Assessing student learning with automated text processing techniques. *Journal of Asynchronous Learning Networks* , 9 (3) . Retrieved from [http://www.sloan-c.org/publications/jaln/v9n3/pdf/v9n3\\_wu.pdf](http://www.sloan-c.org/publications/jaln/v9n3/pdf/v9n3_wu.pdf)
- Breu, K., & Peppard, J. (2001). The Participatory paradigm for applied information systems research. *Proceedings of the 9th European Conference on Information Systems*, pp. 243-252. Slovenia, June 27-29. Retrieved from <http://is2.lse.ac.uk/asp/aspecis/20010085.pdf>

- Budd, R., Thorp, R., & Donohew, L. (1967). *Content analysis of communications*. London: The Collier-McMillan Limited.
- Bullen, M. (1997). *A case study of participation and critical thinking in a university-level course delivered by computer conferencing*. (Unpublished doctoral dissertation). University of British Columbia, Vancouver, Canada. Retrieved from <http://bullenmark.files.wordpress.com/2009/05/dissertation.pdf>
- Bullen, M. (1998). Participation and critical thinking in online university distance education. *Journal of Distance Education*, 13(2), 1-32.
- Burham, H., & de Wolf, R. (2002). Complexity measures and decision tree complexity: a survey. *Theoretical Computer Science*, 288(1), 21-43.
- Cambridge Assessment. (2008). *Cambridge Thinking Skills Assessment Home*. Retrieved from <http://www.admissionstests.cambridgeassessment.org.uk/adt/>
- Capozzoli, M., McSweeney, L., & Sinha, D. (1999). Beyond Kappa: A review of interrater agreement measures. *The Canadian Journal of Statistics*, 27(1), 3-23.
- Cavana, R. Y., Delahaye, B. L., & Sekaran, U. (2001). *Applied business research: Qualitative and quantitative methods*, Queensland, Australia: J. Wiley.
- Chapman, J. (2005). *The development of thinking skills assessment*. Report prepared for the University of Cambridge Local Examinations Syndicate. Cambridge, UK.
- Charles, C. M. (1995). *Introduction to educational research* (2<sup>nd</sup> ed.), San Diego: Longman.
- Checkland, P. (1981). *Systems thinking, systems practice*, Chichester, UK: J. Wiley.
- Checkland, P. (1991). From framework through experience to learning: the essential nature of action research. In H. E. Nissen, H.K. Klein., & R.A. Hirschheim (Eds.), *Information systems research: Contemporary approaches and emergent traditions* (pp. 397-403), Amsterdam: North-Holland.
- Chen, T. (2001). Global exponential stability of delayed Hopfield neural networks. *Neural Networks*, 14(8), 977-980.
- Christian, G. R., Blumenthal, C., & Patterson, M. (2000). The information explosion and the adult learner: Implication for reference librarians. *The Reference Librarian*, 33(69-70), 19-30.

- Clouse, S. T. (2003, November 2). *Tips and tricks for online chats, threaded discussions, and PowerPoint lectures*. Retrieved from <http://www.business.umt.edu/Faculty/clouse/Tips-and-Tricks>
- Clulow, V., Clulow, J., & Brace-Govan, J. (2003), Web-based education: experience-based research. In A. Aggarwal (Ed.), *Web-based education: Learning from experience*, (pp 49-70). Hershey, PA: Information Science Publishing.
- Cohen, W. (2004). *Minorthird: Methods for identifying names and ontological relations in text using heuristics for inducing regularities*. Retrieved from <http://sourceforge.net/apps/trac/minorthird/wiki>
- Cohen, J. (1968). Weighed kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213–220.
- Cohen, C., Manion, L., & Morrison, K. (2007). *Research methods In education* (6<sup>th</sup> ed.). London: RoutledgeFalmer.
- Corey, S. (1953). *Action research to improve school practice*. New York: Teachers College, Columbia University.
- Corich, S., Kinshuk., & Jefferey, L. (2007a). Changing focus from group to individual: Using an automated tool to measure evidence of critical thinking in discussion forums. *Proceedings of the IADIS International Conference on Cognition and Exploratory Learning in Digital Age*, December 7–9, Algarve, Portugal.
- Corich,S., Kinshuk., & Jeffrey,L.M. (2007b) The use of discussion forums in learning communities, pp 87 - 108. In R. Lupplicini, (Ed.), *Learning communities in online education.*, Charlotte, NC: Information Age Publishing.
- Corich, S., Kinshuk., & Jefferey, L. (2009). Automating the measurement of critical thinking for individuals participating in discussion forums. *Proceedings of the IADIS International Conference on Cognition and Exploratory Learning in Digital Age*, November 20–22, Rome, Italy.
- Corich, S. Kinshuk., & Hunt, L.M . (2004a). *Using discussion forums to support collaboration*, Proceedings of Third Pan Commonwealth Forum on Open Learning, Dunedin. July 4 – 8, 2004. Retrieved 23 August 2010 from [http://www.col.org/pcf3/Papers/PDFs/Corich\\_Stephen.pdf](http://www.col.org/pcf3/Papers/PDFs/Corich_Stephen.pdf)
- Corich, S., Kinshuk., & Hunt, L. (2004b). Assessing discussion forum participation: In search of quality. *International Journal of Instructional Technology & Distance Learning*, 1(12), 3-13. Retrieved from [http://www.itdl.org/journal/dec\\_04/article01.htm](http://www.itdl.org/journal/dec_04/article01.htm)

- Corich, S., Kinshuk., & Hunt, L. (2006a). Measuring critical thinking within discussion forums using a computerised content analysis tool. *Proceedings of the Fifth International Conference on Networked Learning 2006*, April 10-12, Lancaster, UK. Retrieved from <http://www.lancs.ac.uk/fss/organisations/netlc/past/nlc2006/abstracts/pdfs/P0720Corich.pdf>
- Corich, S., Kinshuk., & Hunt, L.M. (2006b). Computerised content analysis for measuring critical thinking within discussion forums. *Journal of e-Learning and Knowledge Society*, 2(1), 47 - 60. Retrieved from [http://www.je-lks.it/archive/01\\_06/03.html](http://www.je-lks.it/archive/01_06/03.html)
- Cormack, D. S. (1991). *The research process*. Oxford: Black Scientific.
- Cover, T., & Hart, P. (1967). Nearest neighbour pattern classification. *IEEE Information Theory Society*, 13(1), 21-27.
- Creswell, J. W. (1994). *Research design: Qualitative & quantitative approaches*. Thousand Oaks, CA: Sage.
- Critical Thinking Community. (n.d). *The critical thinking community foundation for critical thinking*. Retrieved from <http://www.criticalthinking.org/resources/assessment/index.cfm>
- Davies, B. (1993). *The thought of Thomas Aquinas*. Oxford: Oxford University Press.
- DeLeng, B.A., Dolmans,D.H.J.M., Jöbsis, B., Muijtjens, A.M.M., & van der Vleuten, C.P.M. (2009). Exploration of an e-learning model to foster critical thinking on basic science concepts during work placements. *Computers & Education*, 53(1), 1-13.
- Denzin, N. K., & Lincoln, Y. S. (1998). *The landscape of qualitative research: Theories and issues*. Thousand Oaks: Sage Publications.
- Denzin, N.K., & Lincoln, Y.S. (2000). Introduction: the discipline and practice of qualitative research. In N.K. Denzin & Y.S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 1 - 28). London: Publications.
- Dewey, J. (1909) *How we think Boston, MA*: D.C. Heath and Co.
- Dewey, J. (1993). *How we think: A restatement of the relation of reflective thinking to the educative process*. Boston: Houghton Mifflin Company.

- DeWever, B., Schellens, T., Valcke, M., & Van Keer, H. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers & Education*, 46, 6–28.
- Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning*, 29(2-3), 103-130.
- Dönmez, P., Rosé, C., Stegmann, K., Weinberger, A., & Fischer, F. (2005). Supporting CSCL with automatic corpus analysis technology. In T. Koschmann, D. Struthers & T. W. Chan (Eds.), *Computer Support for Collaborative Learning 2005: the Next 10 Years!..Proceedings of The 2005 Conference on Computer Support For Collaborative Learning*. (pp. 125-134). Taipei, Taiwan.
- Dovring, K. (1954). Quantitative semantics in 18th century Sweden. *Public Opinion Quarterly*, 18(4), 389-94.
- Doyle, C. S. (1994). *Information literacy in an information society: A Concept for the information age*. Retrieved from Database. (ED372756)
- Elder, L., & Paul, R. (1994). Critical thinking: Why we must transform our teaching. *Journal of Developmental Education*, 18(1), 34-35.
- Ennis, R.H. (1987). A taxonomy of critical thinking dispositions and abilities. In J.B. Baron & R.J. Sternberg (Eds.), *Teaching thinking skills: theory and practice* (pp. 9-26). New York: W.H. Freeman.
- Ennis, R.H. (1993). Critical thinking assessment. *Theory into Practice*, 32(3), 179-186.
- Ennis, R. H. (2004). Applying soundness standards to qualified reasoning. *Informal Logic*, 24(1), 23-39.
- Facione, P. A. (1990). *Critical thinking: A Statement of expert consensus for purposes of educational assessment and instructions. The Dephi Report*. Millbrae, CA: California Academic Press.
- Facione, P. A. (2007). *Critical thinking: what it is and why it counts*. Millbrae, CA: California Academic Press.
- Fahy, P.J. (2001). Addressing some common problems in transcript analysis. *The International Review of Research in Open and Distance Learning*, 1(2). Retrieved from <http://www.irrodl.org/index.php/irrodl/article/view/321/531>
- Fahy, P. J. (2002a). Epistolary and expository interaction patterns in a computer conference transcript. *Journal of Distance Education*, 17(1), 20–35.

- Fahy, P. J. (2002b). Use of linguistic qualifiers and intensifiers in a computer conference. *The American Journal of Distance Education*, 16(1), 5–22.
- Fahy, P. J., & Ally, M. (2005). Student learning style and asynchronous computer-mediated conferencing. *American Journal of Distance Education*, 19(1), 5–22.
- Fahy, P.J., Crawford, G., & Ally, M. (2001). Patterns of interaction in a computer conference transcript. *International Review of Research in Open and Distance Learning*, 2(1). Retrieved from <http://www.irrodl.org/index.php/irrodl/article/view/36/73>
- Fahy, P. J., Crawford, G., Ally, M., Cookson, P., Keller, V., & Prosser, F. (2000). The development and testing of a tool for analysis of computer mediated conferencing transcripts. *Alberta Journal of Educational Research*, 46(1), 85-88.
- Fisher, A. (2001). *Critical thinking: An introduction*. Cambridge, England: Cambridge University Press.
- Fisher, A. (2005). *Thinking skills' and admission to higher education*. Report prepared for the University of Cambridge Local Examinations Syndicate.
- Forster, M. (2004). *Final analysis of the 2003 TSA tests*. Report prepared for the University of Cambridge Local Examinations Syndicate.
- Friedman, M. (1999). *Reconsidering logical positivism*. Cambridge, U.K.: Cambridge University Press.
- Frey, B. A., Sass, S. S., & Alman, S. W. (2006). Mapping MLIS asynchronous discussions. *International Journal of Instructional Technology and Distance Learning*, 3(1), 3-17.
- Fujita, N., & Teplovs, C. (2009). Automating analysis of collaborative discourse: Identifying idea clusters. In C. O'Malley, D. Suthers, P. Reimann & A. Dimitracopoulou (Eds.), *Computer supported collaborative learning practices* (pp.162-164). Rhodes, Greece: International Society of the Learning Sciences, Inc.
- Fuller, M., & Zobel, J.(1998, August 21). *Conflation-based comparison of stemming algorithms*. Paper presented at the Third Australian Document Computing Symposium, Sydney, Australia.

- Garrison, D. R. (1991). Critical thinking and adult education: a conceptual model for developing critical thinking in adult learners. *International Journal of Lifelong Education*, 10, 287–303.
- Garrison, D. R. (2003). Cognitive presence for effective asynchronous online learning: The role of reflective inquiry, self-direction and metacognition. In J. Bourne & J. C. Moore (Eds.), *Elements of quality online education: Practice and direction* (pp.47-58). Volume 4 in the Sloan C Series, Needham, MA: The Sloan Consortium.
- Garrison, D.R., Anderson, T., & Archer, W. (2000). Critical thinking in a text-based environment. Computer conferencing in higher education. *Internet in Higher Education*, 2(2), 87-105.
- Garrison, D. R., & Cleveland-Innes, M. (2005). Facilitating cognitive presence in online learning: Interaction is not enough. *American Journal of Distance Education*, 19(3), 133-148.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, England: Cambridge University Press.
- Giancarlo, C.A., Blohm, S. W., & Urdan, T. (2004). Assessing secondary students' disposition toward critical thinking: Development of the California measure of mental motivation. *Educational and Psychological Measurement*, 64, 347-364.
- Gilbert, P. K., & Dabbagh, N. (2004). How to structure online discussions for meaningful discourse: a case study. *British Journal of Educational Technology*, 36(1), 5-18.
- Glaser, E. M. (1941). *An experiment in the development of critical thinking*, New York: Teachers College, Columbia University.
- Glesne, C., & Peshkin, P. (1992). *Becoming qualitative researchers: An introduction*. New York: Longman.
- Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report*, 8(4), 597-607.
- Grundy, S. (1988). Three modes of action research. In S. Kemmis & R. McTaggart (Eds.). *The action research reader* (3rd ed.) Geelong, Australia : Deakin University Press.

- Guevarra, J. (2009). Discontent with content analysis of online transcripts. *ALT-J: Research in Learning Technology*, 17(2), 101-113.
- Gunawardena, C., Lowe, C., & Anderson, T. (1997). Analysis of a global on-line debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. *Journal of Educational Computing Research*, 17(4), 395-429.
- Hacket, L. (1992). *The age of enlightenment, the European dream of progress and enlightenment*. Retrieved from [http://history-world.org/age\\_of\\_enlightenment.htm](http://history-world.org/age_of_enlightenment.htm)
- Halpern, D. F. (1998). Teaching critical thinking for transfer across domains. *American Psychologist*, 53(4), 449-455.
- Halpern, D.F. (2003). *Thought and knowledge: An introduction to critical thinking*. (4th Ed.). Mahwah, NJ: Erlbaum.
- Hara, N., Bonk, C., & Angeli, C., (2000). Content analyses of on-line discussion in an applied educational psychology course. *Instructional Science*, 28(2), 115-152.
- Harmse, M., Pothas, A.-M., & de Wet, A. G. (2002). Action research on operations research for street children. *Systemic Practice and Action Research*, 15(1), 37-49.
- Harrell, F. E. (2001). *Regression modelling strategies: with applications to linear models, logistic regression, and survival analysis*. New York: Springer-Verlag.
- Hart, E., & Bond, M. (1994). *Action research for health and social care*. Buckingham, England: Open University Press.
- Heath, R. L. (2005). *Encyclopaedia of public relations volume 2* (2nd ed.), California: Sage Publications.
- Heckman, R., & Annabi, H. (2005). A content analytic comparison of learning processes in online and face-to-face case study discussions. *Journal of Computer-Mediated Communication*, 10(2). Retrieved from <http://jcmc.indiana.edu/vol10/issue2/heckman.html>
- Henri, F. (1992). Computer conferencing and content analysis. In A. R. Kaye (Ed.), *Collaborative learning through computer conferencing: The Najaden papers* (pp. 116-136). Berlin: Springer-Verlag.

- Heron, J. (1981). Philosophical basis for a new paradigm. In P. Reason & J. Rowan (Eds.), *Human inquiry: A sourcebook of new paradigm research* (p 19). Chichester, England: Wiley.
- Heron, J., & P. Reason (1997). A participatory inquiry paradigm. *Qualitative Inquiry*, 3(3), 274-294.
- Hiltz, S. R., & Turoff, M. (1993). *Network nation: Human communication via computer*. Boston, MA:MIT Press, revised edition, 1993. Originally published in 1978.
- Hoepfl, M. (1997) Choosing qualitative research: A primer for technology education researchers, *Journal of technology education*, 9(1), 47-63.
- Holsti, O. R. (1969). *Content analysis for the social sciences and humanities*. Reading, MA: Addison-Wesley.
- Holter, I.M., & Schwartz-Barcott,D. (1993). Action research: What is it? How has it been used and how can it be used in nursing? *Journal of Advanced Nursing*, 128, 298-304.
- Hughes, I., Ndonko, F., Ouedraogo, B., Ngum, J., & Popp, D. (2004). International education for action research: The Bamenda Model. *Action Research e-Reports*, 20. Retrieved from <http://sites.google.com/site/ourpinehill/Home/ianh/publications>.
- Isa, D., Hong-lee, L., Kallimani, V. P., & RajKumar, J. (2008). Text documents preprocessing with the Bayes' formula for classification using the support vector machine. *Traction of Knowledge and Data Engineering*, 20(9), 1264-1272.
- Insight Assessment. (n.d.). *Measuring critical thinking worldwide*. Retrieved from <http://www.insightassessment.com/>
- Internet Forum. (2010). Retrieved from Wikipedia: [http://en.wikipedia.org/wiki/Discussion\\_forum](http://en.wikipedia.org/wiki/Discussion_forum)
- Jacob, S. M., & Sam H. K. (2008). *Effect of online discussion forums to improve mathematics problem solving and critical thinking – A comparative study*. Proceedings of the 5th IEEE International Conference on Innovations in Information Technology (Innovations 2008), United Arab Emirates, ISBN 978-1-4244-3397-1, pp. 366-370.
- Jacob, S. M., & Sam, H. K. (2010). Perspectives on critical thinking through online discussion forums in engineering mathematics. In T. Sobh, M. Iskander, V.

- Kapila, M. Karim, & A. Mahmood (Eds.) *Technological developments in networking, education and automation*, pp.121-126. Netherlands: Springer.
- Janikow, C.Z.(1998). Fuzzy decision trees: issues and methods. *Systems, Man, and Cybernetics*, 28(1), 1-14.
- Jeong, A.C. (2003). The sequential analysis of group interaction and critical thinking in online threaded discussions. *The American Journal of Distance Education*, 17(1), 25-43.
- Joachims, T. (1998). *Text categorisation with support vector machines: Learning with many relevant features*. In Proceeding of 10th European Conference on Machine Learning, pp. 137 - 142.
- Joachims, T. (2009). Retrospective on transductive inference for text classification using support vector machines. *Proceedings of the 26th International Conference on Machine Learning (ICML 2009)*, Montreal, Canada, June 14-18.
- Johnson, B. R., & Christensen, L. B. (2004). *Educational research: Quantitative, qualitative, and mixed approaches*, (2nd ed.). New York: Pearson Education.
- Jonassen, D. H. (1994). Technology as cognitive tools: Learners as designers. *IT Forum Paper #1*. Retrieved from <http://it.coe.uga.edu/itforum/paper1/paper1.html>
- Kanuka, H., & Anderson, T. (1998). Online social interchange, discord, and knowledge construction. *Journal of Distance Education*, 13(1), 57-75.
- Kanuka, H., & Garrison, D. R. (2004). Cognitive presence in online learning. *Journal of Computing in Higher Education*, 15(2), 21-39.
- Kanuka, H., Rourke, L., & Laflamme, E. (2007). The influence of instructional methods on the quality of online discussion. *British Journal of Educational Technology*, 38(2), 260-271.
- Kassirer, J., & Campion, E. W. (1994). Peer review crude and understudied, but indispensable. *Journal of the American Medical Association*, 272(2), 96-97.
- Keeley, S. M., & Browne, M. N. (1986). How college seniors operationalize critical thinking behavior. *College Student Journal*, 20, 389-95.
- Kember, D. (2000). *Action learning and action research: Improving quality of teaching and learning*. London: Kogan Page.

- Khan, A., Baharudin, B., Lee, L.H., & Khan, K. (2010). A review of machine learning algorithms for text-documents classification. *Journal of Advances in Information Technology*, 1(1), 4-20.
- Kim, S., Rim, H., Yook, D., & Lim, H. (2002). Effective methods for improving naïve Bayes text classification. In I. Mitsuru & S. Abdul (Eds.), *PRICAI 2002: Trends in artificial intelligence: Lecture notes in computer science*. Berlin, Germany: Springer.
- King, A. (1990). Enhancing peer interaction and learning in the classroom through reciprocal questioning. *American Educational Research Journal*, 27(4), 664-687.
- Klemm, W., & Snell, J. (1996). Enriching computer-mediated group learning by coupling constructivism with collaborative learning. *Journal of Instructional Science and Technology*, 1(2). Retrieved from <http://www.ascilite.org.au/ajet/e-jist/docs/vol1no2/article1.htm>
- Knowles, M., Holton, E., & Swanson, R. (1998). *The adult learner: the definitive classic in adult education and human resource development* (5th ed.). Houston, TX: Gulf Publishing.
- Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy of educational objectives; the classification of educational goals. Handbook II: The affective domain*. New York: Longmans Green.
- Krigger, T. J. (2001). A virtual revolution: Trends in the explosion of distance education. *United States Distance Learning Association Journal*, 15(11), 30-50.
- Krippendorff, K. (2004). *Content analysis An introduction to its methodology* (2nd ed.), California: Sage Publications.
- Kuhn, T. S. (1962) *The structure of scientific revolutions*, (1st ed.). Chicago: University of Chicago Press.
- Lacey, A. R. (1999). *A dictionary of philosophy*.(3rd ed.). New York: Routledge.
- Lam, W. (2004) Encouraging online participation. *Journal of Information Systems Education*, 15(4), 345-348.
- Lamy, M. N., & Goodfellow, R. (1999). “Reflective conversation” in the virtual language classroom. *Language Learning & Technology*, 2(2), 43-61.
- Lather, P. (1986). Research as praxis. *Harvard Educational Review*, 56(3), 257-77.

- Lewin, K. (1946). Action research and minority problems. *Journal of Social Issues*, 2(4), 34-46.
- Lewins, A., & Silver, C. (2007). *Using software in qualitative research: A step-by-step guide*. California: Sage Publications Ltd.
- Liao, Y., & Rao Vemuri, V. (2002). Use of K-Nearest Neighbor classifier for intrusion detection. *Computers & Security*, 21(5), 439-448.
- Lincoln, Y.S., & Guba, E.G. (2000). Paradigmatic controversies, contradictions, and emerging confluences. In N.K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 163-188). Thousand Oaks, CA: Sage Publications.
- Lyman, P., & Varian, H. R. (2003). *How much information?* . Retrieved from <http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/>.
- Marra, R. M., Moore, J. L., & Klimczak, A. K. (2004). Content analysis of online discussion forums: a comparative analysis of protocols. *Educational Technology Research Development*, 52, 23-40.
- Mason, R. (1997). Moderating educational computer conferencing. *DEOSNEWS*, 1(19). Retrieved from [http://www.ed.psu.edu/acsde/deos/deosnews/deosnews1\\_19.asp](http://www.ed.psu.edu/acsde/deos/deosnews/deosnews1_19.asp).
- Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data*. Mahwah, NJ: Lawrence Erlbaum.
- Mayadas, F. (2001). Testimony to the Kerrey commission on web-based education. *Journal of Asynchronous Learning Networks* 5(1), 134-138.
- McCutcheon, G., & Jurg, B., (1990). Alternative perspectives on action research. *Theory into Practice*, 24(3), 144-151.
- McDonald, J. (1998). Interpersonal group dynamics and development in computer conferencing: The rest of the story. In *Wisconsin Distance Education Proceedings Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans*. Retrieved from <http://www.mrc.gc.ca/publications/publications.html>
- McKenzie, W., & Murphy, D. (2000). I hope this goes somewhere: evaluation of an online discussion group. *Australian Journal of Educational Technology*, 16, 139-257.

- McKernan, J. (1991). *Curriculum action research. A handbook of methods and resources for the reflective practitioner*. London: Kogan Page.
- McKlin, T.E. (2004). *Analyzing cognitive presence in online courses using an artificial neural network*. (Unpublished doctoral dissertation). Georgia State University, USA.
- McKlin, T. E., Harmon, S., Evans, W., & Jones, M. (2002). Cognitive presence in web-based learning: A content analysis of student's online discussion. *IT Forum*, 60. Retrieved from <http://it.coe.uga.edu/itforum/paper60/paper60.htm>
- McMahon, M. (1997, December). *Social constructivism and the world wide web – A paradigm for learning*. Paper presented at the Australian Society for Computers In Learning In Tertiary Education, Perth WA.
- Meyer, K.A. (2004). Evaluating online discussions: Four different frames of analysis. *Journal of Asynchronous Learning Networks*, 8(2), 101-114. Retrieved from [http://www.sloan-.org/publications/jaln/v8n2/v8n2\\_meyer.asp](http://www.sloan-.org/publications/jaln/v8n2/v8n2_meyer.asp)
- Moller, L. (1998) Designing communities of learners for asynchronous distance education. *Educational Technology Research and Development*, 46(4), 115-122.
- Morley, D. (1991). Resource analysis as action research. In P. F. Wilkinson & W. C. Found (Eds.), *Resource analysis research in developing countries*. Toronto: York University Press.
- Mumford, E., & Weir, M. (1979). *Computer systems work design: The ETHICS method*, London: Associated Business Press.
- Murphy, E. (2009). A framework for identifying and promoting metacognitive knowledge and control in online discussants. *Canadian Journal of Learning and Technology*, 34 2). Retrieved from <http://www.cjlt.ca/index.php/cjlt/article/view/491/222>
- Newman, G., Webb, B., & Cochrane, C. (1995). A content analysis method to measure critical thinking in face-to-face computer supported group learning. *Interpersonal Computing and Technology*, 3(2), 56-77.
- Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, California: Sage Publications.

- Nigam, K., McCallum, A. K. , Thrun, S., & Mitchell, T. (2000). *Text classification from labeled and unlabeled documents using machine learning*. Netherlands: Springer.
- Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: Big five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology*, 93(1), 116-130. doi: 10.1037/0022-3514.93.1.116
- Norris, S. P., & Ennis, R. H. (1989). *Evaluating critical thinking*. Pacific Grove, CA: Midwest Publications.
- O'Brien, R. (2001). Um exame da abordagem metodológica da pesquisa ação [An Overview of the Methodological Approach of Action Research]. In R. Richardson (Ed.), *Teoria e Prática da Pesquisa Ação [Theory and Practice of Action Research]*. João Pessoa, Brazil: Universidade Federal da Paraíba. (English version) Retrieved from <http://www.web.ca/~robrien/papers/arfinal.html>
- Oriogun, P. K., Ravenscroft, A., & Cook, J. (2005). Validating an approach to examining cognitive engagement in online groups. *American Journal of Distance Education*, 19(4), 197-214.
- Ozkan, B. C. (2004). Using NVivo to analyze qualitative classroom data on constructivist learning environment. *The Qualitative Report*, 9(4), 589-603.
- Palloff, R. M., & Pratt, K. (1999). *Building learning communities in cyberspace*. San Francisco: Jossey- Bass Inc.
- Palloff, R.M., & Pratt, K. (2002). What we know about electronic learning. In L. Foster, B. L. Bower, & L. W. Watson (Eds.), *ASHE reader distance education: Teaching and learning in higher education* (pp. 224-231). Boston, MA: Pearson Custom Publishing.
- Palmquist, M. E., Carley, K.M., & Dale, T.A. (1997). Two applications of automated text analysis: Analyzing literary and non-literary texts. In C. Roberts (Ed.), *Text analysis for the social sciences: Methods for drawing statistical inferences from texts and transcripts* (pp. 171-189). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Patton, M. Q. (2001). *Qualitative research and evaluation methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Paul, R. W. (1997). The Critical thinking movement: 1970-1997: Putting the 1997 conference into historical perspective. *The Critical Thinking Community*.

Retrieved from <http://www.criticalthinking.org/articles/documenting-history.cfm>.

- Paul, R.W., & Elder, L. (2001). *Critical thinking: Tools for taking charge of your learning and your life*. Upper Saddle River NJ: Prentice Hall.
- Paul, R. W., Elder, L., & Bartell, T. (1997). *California teacher preparation for instruction in critical thinking: Research findings and policy recommendations*. Sacramento, CA: California Commission of Teacher Credentialing.
- Pazzani, M., & Billsus, D. (1997). Learning and revising user profiles. The identification of interesting web sites. *Machine Learning*, 27(3), 313-331.
- Pedersen, T. (2000). A simple approach to building ensembles of Naive Bayesian classifiers for word sense disambiguation. *Proceedings of the 1st North American chapter of the Association for Computational Linguistics* (pp. 63 - 69). Seattle, Washington.
- Pena-Shaff, J. B., & Nicholls, C. (2004). Analyzing student interactions and meaning construction in computer bulletin board discussions. *Computers & Education*, 42, 243-265.
- Peng, F., Ahmed, N., Li, X., & Lu, Y. (2007). Context sensitive stemming for web search. In: *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, Amsterdam, Netherlands, pp. 639 - 646.
- Perkins, D. N. (1985). Postprimary education has little impact on informal reasoning. *Journal of Educational Psychology*, 77(5), 562-571.
- Perkins, D. N. (1989). Reasoning as it is and could be: An empirical perspective. In D. M. Topping, D. S. Cromwell, & V. N. Kobayaski (Eds.), *Thinking across cultures: Third international conference on thinking* (pp. 175-194). Hillsdale, NJ: Erlbaum.
- Perkins, C., & Murphy, E. (2006). Identifying and measuring individual engagement in critical thinking in online discussions: An exploratory case study. *Educational Technology & Society*, 9(1), 298-307.
- Peters, K.M. (2000). *Creative use of threaded discussion areas, Part 1*. *WebCT Online Teaching and Learning* (newsletter), Retrieved from <http://www.webct.com/OTL/ViewContent?contentID=898084>

- Ramoni, M., & Sebastiani, P. (2001). Robust Bayes classifiers. *Artificial Intelligence*, 125(1), 209-226.
- Reason, P. (1994). Three approaches to participative inquiry. In N.K. Denzin & Y.S. Lincoln (Eds.). *Handbook of qualitative research* (p. 324). London, England: Sage.
- Reason, P. (1998). A participatory world. *Resurgence*, 186, 43-44.
- Reed, J. H. (1997). *Effect of a model for critical thinking on student achievement in primary source document analysis and interpretation, argumentative reasoning, critical thinking dispositions, and history content in a community college history course*. (Unpublished doctoral dissertation), University of South Florida, Tampa, Florida. Retrieved 14 April 2010 from <http://www.criticalthinking.org/resources/JReed-Dissertation.pdf>
- Rheingold, H. (1994). A slice of life in my virtual community. In L.M. Harasin (Ed.), *Global networks Computers and international communication* (pp. 57-80). Cambridge, MA: MIT Press.
- Riel, M. (2007) *Understanding action research*, Center For Collaborative Action Research. Pepperdine University. Retrieved from <http://cadres.pepperdine.edu/ccar/define.html>.
- Riffe, D., Lacey, S., & Fico, F.(1998). *Analyzing media messages: Quantitative content analysis*. New Jersey: Lawrence Erlbaum Associates, Inc.
- Rigou, M. (2004, February). *On the development of adaptive web-based learning communities*. Proceedings of the IASTED International Conference Web-based Education, Innsbruck, Austria.
- Riloff, E. (1995). Little words can make a big difference for text classification. In the *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.130-136.
- Ripley, B. D. (2008). *Pattern recognition and neural networks*. Cambridge, England: Cambridge University Press.
- Rish, I., Hellerstein, J., & Thathachar, J. (2001). *An analysis of data characteristics that affect Naïve Baye's performance*. Hawthorne, NY: IBM T.J. Watson Research Center. Retrieved from <http://www.research.ibm.com/PM/icml01.pdf>
- Rourke, L., & Anderson, T. (2004). Validity in quantitative content analysis . *Educational Technology Research and Development*, 52(1), 5-18.

- Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (2001). Methodological issues in the content analysis of computer conference transcripts. *International Journal of Artificial Intelligence in Education*, 12(1), 8-22. Retrieved from [http://www.iaied.org/pub/951/file/951\\_paper.pdf](http://www.iaied.org/pub/951/file/951_paper.pdf)
- Rovai, A. P. (2002). Sense of community, perceived cognitive learning and persistence in asynchronous learning networks. *Internet and Higher Education*, 5(4), 319-332.
- Schafersman, S., D. (1991). *An introduction to critical thinking*. Retrieved from <http://www.freeinquiry.com/critical-thinking.html>
- Schellens, T., & Valcke, M. (2005). Collaborative learning in asynchronous discussion groups: What about the impact on cognitive processing? *Computers in Human Behavior*, 21, 957-975
- Schrire, S. (2006). Knowledge building in asynchronous discussion groups: Going beyond quantitative analysis. *Computers & Education*, 46(1), 49-70.
- Seale, C. (1999). Quality in qualitative research. *Qualitative Inquiry*, 5(4), 465-478.
- Sergiovanni, T. (1999). The story of community. In J. Retallick, B. Cocklin, & K. Coombe (Eds.), *Learning Communities in Education: Issues, Strategies and Contexts* (pp. 9-25). London: Routledge.
- Shea, P., & Bidjerano, T. (2009a). Cognitive presence and online learner engagement: a cluster analysis of the community of inquiry framework. *Journal of Computing in Higher Education*, 21, 199-217.
- Shea, P., & Bidjerano, T. (2009b). Community of inquiry as a theoretical framework to foster “epistemic engagement” and “cognitive presence” in online education. *Computers & Education*, 52(3), 543-553.
- Shea, P., Li, C. S., & Pickett, A. (2006). A study of teaching presence and student sense of learning community in fully online and web-enhanced college courses. *The Internet and Higher Education*, 9(3), 175-190.
- Sheffler, I. (1973). *Reason and teaching*. Indianapolis, IN: Hackett.
- Skinner, B.F. (1969). Contingencies of reinforcement in the design of a culture. *Behavioral Science*, 11(3), 159-166.
- Smith, M. K. (1997) 'Paulo Freire and informal education'. *The encyclopaedia of informal education*. Retrieved 14 April 2010 from <http://www.infed.org/thinkers/et-freir.htm>.

- Spatariu, A., Hartley, K., & Bendixen, L.D. (2004, Spring) Defining and measuring quality in on-line discussion. *Journal of Interactive Online Learning*. 2(4). Retrieved from <http://www.ncolr.org/jiol/issues/PDF/2.4.2.pdf>
- Squire, K., & Johnson, C. (2000). Supporting distributed communities of practice with interactive television. *Educational Technology Research and Development*, 48(1), 23–43.
- Stacey, E., & Gerbric, P. (2003. December 7 - 10). Investigating the impact of computer conferencing: Content analysis as a manageable research tool. In G.Crisp, D.Thiele, I.Scholten, S.Barker and J.Baron (Eds.), *Interact, Integrate, Impact: Proceedings of the 20th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education*. Adelaide, Australia.
- Stein, D.S., Wanstreet, C.E., Glazer, H.R., Engle, C.L., Harris, R.T., Johnston, S.M., Simons, M.R., & Trinko, L.A. (2007). Creating shared understanding through chats in a community of inquiry. *The Internet and Higher Education*, 10, 103 - 115.
- Sumner, W. G. (1906). *Folkways: a study of the sociological importance of usages, manners, customs, mores, and morals*. Boston: Ginn and Co.
- Susman, G. I., & Evered, R.D. (December, 1978). An assessment of the scientific merits of action research. *Administrative Science Quarterly* 23, 582-603.
- Swan, K., & Shea, P (2005). Social presence and the development of virtual learning communities. In S. Hiltz & R. Goldman (Eds.), *Learning together online: Research on asynchronous learning networks* (pp.239–260). Mahwah, NJ: Lawrence Erlbaum.
- Tandon, R. (1998). Participation, citizenship and democracy: reflections on 25 years' of PRIA. *International Journal of Adult Education*, 21(2-3), 5-18.
- The Critical Thinking Community. (n.d.). *International critical thinking test*. Retrieved from <http://www.criticalthinking.org/assessment//ICAT-info.cfm>
- Tishman, S., & Andrade ,A. (1999). Thinking dispositions: a review of current theories, practices, and issues. *Harvard University, Active Learning Practices for Schools*. Retrieved from <http://learnweb.harvard.edu/alps/thinking/docs/Dispositions.pdf>
- Tishman, S., Perkins, D., & Jay, E. (1995). *The thinking classroom*. Boston: Allyn and Bacon.

- Thomas, M. (1999). The impacts of technology on communication – Mapping the limits of online discussion forums. *The University of Adelaide Intranet Project*. Retrieved from [http://online.adelaide.edu.au/LearnIT.nsf/URLs/technology\\_and\\_communication](http://online.adelaide.edu.au/LearnIT.nsf/URLs/technology_and_communication)
- Trist, E. (1976). Engaging with large-scale systems. In: A. Clark (Ed.), *Experimenting with organizational life: The action research approach*, pp. 43-75. New York: Plenum.
- Underbakke, M., Borg, J. M., & Peterson, D. (1993). Researching and developing the knowledge base for teaching higher order thinking. *Theory into Practice*, 32 (3), 138-146.
- Vapnik, V. N. (1995). *The Nature of statistical learning*. New York: Springer.
- Vaughan, N., & Garrison, D. R. (2005). Creating cognitive presence in a blended faculty development community. *Internet and Higher Education*, 8, 1-12.
- Veerman, A., & Veldhuis-Diermanse, E. (2001). Collaborative learning through computer-mediated communication in academic education. In *Euro CSCL 2001* (pp. 625–632). Maastricht: McLuhan institute, University of Maastricht
- Washington State University. (n.d.). *Critical thinking project*. Retrieved from <http://wsuctproject.wsu.edu/>
- Weiss, R. E., & Morrison, G. R. (1998). Evaluation of a graduate seminar conducted by list-serv. *Proceedings of selected paper presentations at the convention of the Association for Educational Communications and Technology* (pp. 463–478). Ames: Iowa State University.
- White, M. D., & Marsh, E. E. (2006) Content analysis: A flexible methodology. *Library Trends*, 55(1), 22-45.
- Whittaker, S., Isaacs, E., & O'Day, V. (1997). Widening the Net. In *SIGCHI Bulletin*, 29(1). New York: ACM Press.
- Williams, L., & Lahman, M. (2009). Online discussion, student engagement, and critical thinking (2009). *APSA 2009 Toronto Meeting Paper*. Retrieved from SSRN: <http://ssrn.com/abstract=1448970>
- Wilson, A., & Rayson, P. (1993) The automatic content analysis of spoken discourse: A report on work in progress. In C. Souter & E. Atwell (Eds.), *Corpus-based computational linguistics* (pp. 215-226). Amsterdam: Rodopi.

- Winter, R. (1989a) *Action research and the nature of social inquiry: Professional innovation and educational Work*. Aldershot, England: Gower.
- Winter, R. (1989b). *Learning from experience: Principles and practice in action-research*. Philadelphia: The Falmer Press.
- Wise, A., Chang, J., Duffy, T., & del Valle, R. (2004). The effects of teacher social presence on student satisfaction, engagement, and learning. *Journal of Educational Computing Research*, 31, 247-271.
- Wood-Harper, A. T. (1989). *Comparison of Information Systems Definition Methodologies: An Action Research Multiview Perspective*. (Unpublished doctoral dissertation). University of East Anglia, Norwich, UK.
- Wood-Harper, A.T., Antill, L., & Avison, D. E. (1985) *Information systems definition: The multiview approach*, Oxford, England: Blackwell Scientific.
- Xenophon. (1852). Xenophon's memorabilia of Sokrates / tr. from the text of Raphael Kuhner ; with notes and prolegomena by George B. Wheeler. London: W. Allan.
- Yang, Y.C., & Chou, H. (2008). Beyond critical thinking skills: Investigating the relationship between critical thinking skills and dispositions through different online instructional strategies. *British Journal of Educational Technology*, 39(4), 666–684.
- Yeh, S. S. (2001). Tests worth teaching to: Constructing state-mandated tests that emphasize critical thinking. *Educational Researcher*, 30(9), 12-17.
- Zelkowitz, M.V. (2006). Techniques for empirical validation. In V. R. Basili, H. D. Rombach, K. Schneider, B. A. Kitchenham, D. Pfahl & R. W. Selby (Eds.), *Empirical software engineering issues: Lecture notes in computer science* (pp. 4 - 9). New York: Verlag.
- Zhu, E. (1997). Meaning negotiation, knowledge construction, and mentoring in a distance learning course. (ERIC Document Reproduction Service, ED 397 849).



# APPENDICIES

---



## Appendix A: Ethics Documentation: Massey University



**Massey University**

OFFICE OF THE ASSISTANT  
TO THE VICE-CHANCELLOR  
(Research Ethics)  
Private Bag 11 222  
Palmerston North 4442  
New Zealand  
T 64 6 350 5573/350 5575  
F 64 6 350 5622  
humanethics@massey.ac.nz  
animalethics@massey.ac.nz  
gtc@massey.ac.nz  
www.massey.ac.nz

1 August 2008

Stephen Corich  
c/o Eastern Institute of Technology  
Private Bag 1201  
Taradale  
NAPIER

Dear Stephen

**Re: Measurement of Critical Thinking Ability Using an Automated Discussion Forum Evaluation Tool and the Cambridge University Critical Thinking Assessment**

Thank you for your Low Risk Notification which was received on 31 July 2008.

Your project has been recorded on the Low Risk Database which is reported in the Annual Report of the Massey University Human Ethics Committees.

The low risk notification for this project is valid for a maximum of three years.

Please notify me if situations subsequently occur which cause you to reconsider your initial ethical analysis that it is safe to proceed without approval by one of the University's Human Ethics Committees.

Please note that travel undertaken by students must be approved by the supervisor and the relevant Pro Vice-Chancellor and be in accordance with the Policy and Procedures for Course-Related Student Travel Overseas. In addition, the supervisor must advise the University's Insurance Officer.

**A reminder to include the following statement on all public documents:**

*"This project has been evaluated by peer review and judged to be low risk. Consequently, it has not been reviewed by one of the University's Human Ethics Committees. The researcher(s) named above are responsible for the ethical conduct of this research.*

*If you have any concerns about the conduct of this research that you wish to raise with someone other than the researcher(s), please contact Professor Sylvia Rumball, Assistant to the Vice-Chancellor (Research Ethics), telephone 06 350 5249, e-mail humanethics@massey.ac.nz".*

Please note that if a sponsoring organisation, funding authority or a journal in which you wish to publish requires evidence of committee approval (with an approval number), you will have to provide a full application to one of the University's Human Ethics Committees. You should also note that such an approval can only be provided prior to the commencement of the research.

Yours sincerely

Sylvia V Rumball (Professor)  
Chair, Human Ethics Chairs' Committee and  
Assistant to the Vice-Chancellor (Research Ethics)

cc Assoc Prof Kinshuk  
Department of Information Systems  
PN311

Prof Claire Massey, HoD  
Department of Management  
PN214

---

Massey University Human Ethics Committee  
Accredited by the Health Research Council





## Appendix B: Ethics Documentation: EIT Hawke's Bay



Ref: 18/08

29 August 2008

Steve Corich  
C/- Faculty of Business & Computing  
EIT Hawke's Bay

Dear Steve

Your research project, "Measurement of critical thinking ability using an automated discussion forum evaluation tool and the Cambridge University Critical Thinking Assessment" has been reviewed by the EIT Research Approvals Committee.

I am pleased to advise that the Committee has approved your project for a period of two years, ending in August 2010.

You are reminded that if the proposal changes in any significant way, then you must inform the Research Committee.

We wish you well for the project.

Yours sincerely



**Jeanette Fifield**  
Secretary, Research Approvals Committee

**EASTERN INSTITUTE OF TECHNOLOGY**  
MAIN CAMPUS Gloucester Street, Private Bag 1201, Taradale 4141, New Zealand. Telephone 06 974 8000, Facsimile 06 974 8910  
Web [www.eit.ac.nz](http://www.eit.ac.nz) Email [info@eit.ac.nz](mailto:info@eit.ac.nz)  
HASTINGS CENTRE Cnr Lyndon & Railway Roads, PO Box 1477, Hastings 4156. Telephone 06 878 4738, Facsimile 06 878 2965  
CENTRAL HAWKE'S BAY CENTRE 53 Russell Street, PO Box 230, Waipukurau 4242. Telephone 06 858 7009, Facsimile 06 858 7018  
FLAXMERE CENTRE Flaxmere Village, Swansea Road, Flaxmere 4120. Telephone 06 974 8943  
WAIROA CENTRE Cnr Paul & Queen Streets, Wairoa 4108. Telephone 06 838 7349



## Appendix C: Information Sheet: Discussion Forum Participants

Discussion Forum Participant  
Eastern Institute of Technology Hawke's Bay  
Gloucester Street  
Taradale  
Napier

Dear \_\_\_\_\_

As a Principal Academic Staff Member at the Eastern Institute of Technology Hawke's Bay (EIT) I have developed an interest in attempting to identify evidence of critical thinking, particularly among students enrolled in our undergraduate degree programmes. This has arisen from a personal desire to identify ways that critical thinking skills can be encouraged and developed as students participate in classroom and online activities. I have chosen to investigate how evidence of critical thinking can be identified among discussion forum participants as part of my studies towards gaining a Doctor of Philosophy at Massey University.

The title of my research project is:

### **Automating the Measurement of Critical Thinking in Discussion Forums**

I have chosen to use action research to carry out my study so that I can build and refine a computerised system that will attempt to identify evidence of critical thinking automatically.

I am seeking your agreement to act as participants in the study.

### **Participant Involvement**

- As part of your course assessment you will be asked to participate in a discussion forum activity.

- You will be expected to make at least six discussion forum postings over a period of three weeks and the quality of your postings will be used to grade your assessment activity. Your agreement to participate (or not participate) will have no impact on the grade that you may be awarded for the discussion forum activity.
- The transcript of your discussion forum activity will be used as part of the study. It will be categorised by one or more human coders against a critical thinking measurement model and it will be imported into a computerised system that will attempt to automate the classification process.
- You may be asked to participate in the Cambridge Thinking Skills Assessment, an activity that will take approximately 90 minutes.

### **Project Procedures**

- All information will be collected in complete confidence and pseudonyms will be used in reporting.
- The transcript of the discussion forum will be categorised by one or more human coders against a critical thinking measurement model and it will be imported into a computerised system that will attempt to automate the classification process.
- All information gathered during the study will be securely stored in my home office for a period of five years, and then it will be destroyed.
- A report of the findings will be made available to all participants on request.

### **Participant Rights**

Participants will have the right to :

- decline to participate
- decline to answer any questions
- withdraw from the study at any time up until the data is analysed
- ask any questions about the study at any time during participation
- provide information on the understanding that names will not be used unless specific permission is given to the researcher
- be given access to summary findings when it is concluded
- be given the results of the Cambridge Thinking Skills Assessment

**Ethics Approval**

This project has been evaluated by peer review and judged to be low risk. Consequently it has not been reviewed by one of the University's Human Ethics Committees. The researcher named above is responsible for the ethical conduct of this research.

If you have concerns about the conduct of this research that you wish to raise with someone other than the researcher, please contact Professor Sylvia Rumball, Assistant to the Vice Chancellor (Research Ethics), telephone 06 350 5349, email [humanethics@massey.ac.nz](mailto:humanethics@massey.ac.nz).

**Contact Information**

If you have any questions about the study please contact:  
Stephen Corich, Principal Academic Staff Member, EIT Hawke's Bay  
(Pone. 06 974 8000 extn 5459, Email [scorich@eit.ac.nz](mailto:scorich@eit.ac.nz))



## Appendix D: Consent Form: Discussion Forum Participants

### **CONSENT FORM**

(Discussion Forum Participant)

Project Title:

### **Automating the Measurement of Critical Thinking in Discussion Forums**

THIS CONSENT FORM WILL BE HELD FOR A PERIOD OF FIVE (5) YEARS

We have read the Information Sheet and the details of the study have been explained to us. Our questions have been answered to our satisfaction, and we understand that we may ask further questions at any time.

I agree to participate in this research activity.

Full Name: \_\_\_\_\_ Date: \_\_\_\_\_

Signature: \_\_\_\_\_