

**DEVELOPING CRITICAL THINKING SKILLS IN  
TERTIARY ACADEMIC WRITING THROUGH THE USE  
OF AN INSTRUCTIONAL RUBRIC  
FOR PEER EVALUATION**

---

A thesis submitted in fulfilment of the requirements for the  
Degree of  
Doctor of Philosophy  
in the  
University of Canterbury  
by  
Nor Shidrah Binti Mat Daud

---

The University of Canterbury  
July 2012

## DEDICATION

*This thesis is especially dedicated to  
my husband Jaafar Md Napis  
and children*

*Along, Angah, Alieya, Fikrie & Dhina  
for their endless love, support and encouragement.*

*Also, to the memory of  
my late father, Mat Daud Zabidi and my late mother Siti Mariam Karib  
who passed on a respect for education.*

*Last but not least, my siblings and family members  
for the motivation and inspiration.*

## ACKNOWLEDGEMENTS

Pursuing a PhD is indeed a challenging endeavour. The unique journey, amidst the terrifying September 14, 2010 and February 22, 2011 earthquakes and hundreds of other aftershocks, has indeed strengthened my faith in God's will, taught me to be strong in facing challenges in life and shaped my academic and general life perspectives. This thesis would not have been possible without the personal and practical support of some wonderful people. It is to them I owe my deepest gratitude.

- Associate Professor Alison Gilmore and Dr. H. Elaine Mayo, my two lovely and esteemed supervisors. I am most grateful for your expertise, encouragement and patience
- Ministry of Higher Education Malaysia and University Teknologi MARA, Malaysia for the support and sponsorship
- College of Education, University of Canterbury for the continuous support
- Creon Upton, Stephanie Day and Mike Dickinson, advisors of the Learning Skills Centre; Dr. Keith Comer, Coordinator of Academic Development Group and Daniel Bedggood, Undergraduates Coordinator, Department of English, School of Humanities, College of Arts, UC for the constructive input to the creation of the *CAWAR*
- Professor Nuraihan Mat Daud, the Dean of the Centre for Languages and Pre-University Academic Development (CELPAD), International Islamic University Malaysia (IIUM) for the permission and support to conduct the study at the University
- Ainul Azmin Md. Zamin and Rosnani Kassim, instructors of *English for Academic Writing* (LE 4000) of IIUM for the kind cooperation given throughout the study
- Pat Coope for guiding me through the statistical analyses
- Mashitoh, Samantha, Ameer, Alice, David and other fellow postgraduates for their endless support and motivation

- The students from the class of *Writing the Academic Essay* (ENGL 117) Semester 2, 2009 of the University of Canterbury, and students from four classes of *English for Academic Writing* (LE 4000) Semester 2, 2009/2010 of the International Islamic University Malaysia for their enthusiastic participation and willingness to contribute their time to this project
- My loving husband, Jaafar Md Napis for his faithful love, encouragement and sacrifice which can never be repaid
- My wonderful children; Nurul Syafieqah, Muhammad ‘Afif Fakhrie, Nur ‘Alieya, Muhammad Fikrie ‘Aqlan and Nur ‘Idhina for the understanding and support
- Lastly, I offer my sincere thanks to all of those who supported me in any respect during the completion of the dissertation. My enormous debt of gratitude can hardly be repaid to you all.

## ABSTRACT

Critical thinking skills have been identified as learning outcomes expected of students for most courses of tertiary education in many countries including Malaysia. One of the courses where critical thinking is required is academic writing. Producing academic writing which is well argued, insightful, thought-provoking, characterised by evidence and wide reading is a challenge for undergraduate students. Not only do the students need to have a good command of the language, they also need to be critical as they examine viewpoints, facts and arguments and synthesise them. This thesis explores several approaches to developing critical thinking skills in an academic writing course for undergraduate students.

The use of a rubric or a checklist and discussion with peers were identified in the study to support the development of critical thinking. Their potency was explored in a quasi-experimental study involving undergraduate students taking *English for Academic Writing* course. The three treatments groups were: peer review where students used a checklist and discussed their ideas; peer evaluation where students used the rubric and discussed their ideas and evaluations; self-evaluation where students used the rubric but did not discuss their ideas. The level of critical thinking for each groups and a control group who received no treatment, was measured before and after learning interventions using two instruments: the Cornell Critical Thinking Test Level X (CCTT-X) and the *English for Academic Writing* term paper. In addition, students' and instructors' perspectives on the learning activities were elicited by means of questionnaires and interviews. Classroom observations were also carried out.

The rubric which was used in the peer evaluation and self-evaluation activities is called the Critical Thinking for Academic Writing Analytical Rubric (*CAWAR*). It contains 12 criteria with descriptions of the two ends of performance domains i.e. the best and the weakest points and a commentary space. The checklist used in the peer review activity, on the other hand, replicates the *CAWAR* except that it does not have the grading element.

The study found that all treatments showed some potential for fostering the development of critical thinking skills. Theoretically, it is argued that peer evaluation has the greatest potential of the three treatments provided that both teachers and students understand the value of collaborative learning and the importance of giving sufficient time for discussion. The introduction of either the rubric or checklist or promoting peer discussion has promoted critical thinking in an academic writing course.

## LIST OF ABBREVIATIONS

ABBREVIATIONS	IN FULL
CAWAR	Critical Thinking for Academic Writing Analytical Rubric
CCTT-X	Cornell Critical Thinking Test Level X
CELPAD	Centre for Languages and Pre-University Academic Development
IIUM	International Islamic University Malaysia

## LIST OF FIGURES

Figure		Page
1.	Vygotsky's Zone of Proximal Development (ZPD) as depicted by Atherton's (2011)	32
2.	Roles of Writer and Peer Evaluator in Producing Written Products	43
3.	The Development Process of the CAWAR	63
4.	Evaluating the Appropriateness of Scoring Categories to a Stated Purpose	80
5.	The Summary of Research Procedures	118
6.	The Distribution of Scores on the CCTT Pre-test (N=71) by Group of Students	143
7.	The Distribution of Scores on the CCTT Post-test (N=71) by Group of Students	143
8.	The Distribution of Scores on the CCTT-X <i>Inductive Reasoning</i> (N=23) Pre-test by Group of Students	144
9.	The Distribution of Scores on the CCTT-X <i>Inductive Reasoning</i> (N=23) Post-test by Group of Students	144
10.	The Distribution of Scores on the CCTT-X <i>Credibility of Assertions</i> (N=24) Pre-test by Group of Students	145
11.	The Distribution of Scores on the CCTT-X <i>Credibility of Assertions</i> (N=24) Post-test by Group of Students	145
12.	The Distribution of Scores on the CCTT-X <i>Deductive Reasoning</i> (N=14) Pre-test by Group of Students	145
13.	The Distribution of Scores on the CCTT-X <i>Deductive Reasoning</i> (N=14) Post-test by Group of Students	145
14.	The Distribution of Scores on the CCTT-X <i>Identifying Assumptions</i> (N=10) Pre-test by Group of Students	146
15.	The Distribution of Scores on the CCTT-X <i>Identifying Assumptions</i> (N=10) Post-test by Group of Students	146
16.	CCTT-X Pre-test and Post-test Overall Means of Scores	151
17.	Gain Scores by Group	151
18.	CCTT-X Pre- and Post-test Means of Scores for <i>Inductive Reasoning</i>	156
19.	CCTT-X Pre- and Post-test Means of Scores for <i>Credibility of Assertions</i>	156
20.	CCTT-X Pre- and Post-test Means of Scores for <i>Deductive Reasoning</i>	157
21.	CCTT-X Pre- and Post-test Means of Scores for <i>Identification of Assumptions</i>	157



22.	Students' Perception on Whether They Benefitted More as Assessors/Reviewers or Assesseees/Reviewees	168
23.	Peer Evaluation Group's Perceptions of the Peer Evaluation Activity Potential to Foster Critical Thinking Ability in Academic Writing: Before and After Interventions	172
24.	Peer Review Group's Perceptions of the Peer Review Activity Potential to Foster Critical Thinking ability in Academic Writing: Before and After Interventions	172
25.	Self-Evaluation Group's Perceptions of the Self-Evaluation Activity Potential to Foster Critical Thinking ability in Academic Writing: Before and After Interventions	173

## LIST OF TABLES

Table		Page
1.	Anderson et al.'s (2001) Taxonomy Table	66
2.	The Three Models Compared	69
3.	LE 4000 Course Objectives and Learning Outcomes	75
4.	Questions to Examine Each Type of Validity Evidence	82
5.	Pearson's Correlation Coefficients for Instructor's Scores and Average Scores of Peers' Rating of Third Draft using the CAWAR and Term Paper and CCTT-X Post-test Scores	84
6.	Draft 1 of the CAWAR - Option 1	90
7.	Draft 1 of the CAWAR - Option 2	96
8.	Draft 2 of the CAWAR	103
9.	Draft 3 of the CAWAR	104
10.	Draft 4 of the CAWAR	107
11.	The Means and Standard Deviations of the Total Scores of Two Raters Using the CAWAR	112
12.	Instructors' Background	121
13.	The Distribution of Participants	122
14.	Reliability Coefficients (Cronbach's Alpha) for CCTT-X and Subscales	126
15.	Means and Std. Deviation of Both Raters Scores of 40 Term Papers	130
16.	Pearson's Correlation Coefficients for Rater 1 and Rater 2	131
17.	Schedule of Activities	135
18.	ANCOVA- Controlling for the Effects of Faculty and English proficiency Level on the CCTT-X Post-test Scores of the Control and Peer Evaluation Groups	148
19.	ANCOVA- Controlling for the Effects of English Proficiency Level on the CCTT-X Post-test Scores of All Groups	148
20.	ANCOVA- Controlling for the Effects of Gender on the CCTT-X Post-test Scores of All Groups	149
21.	CCTT-X Pre-test and Post-test Means and Standard Deviation of Scores by Group of Students	150
22.	Post hoc Bonferroni test for pairwise comparisons for CCTT-X	153
23.	Pre-test and Post-test Means of Scores by Skill and Group of Students	154
24.	ANOVA for Individual Skill Pre-test Scores (N=99)	155
25.	A Summary of the Groups' Performances on the CCTT-X	158
26.	ANCOVA for Treatment and Post-test with Pre-test as Covariate	159

	(N=99)	
27.	Post Hoc Bonferroni Test for Pairwise Comparisons for <i>Inductive Reasoning</i>	160
28.	Post hoc Bonferroni test for Pairwise Comparisons for <i>Credibility of Assertions</i>	161
29.	Means and Standard Deviation of Scores by Group of Students for the Term Paper	162
30.	Two-Way Nested ANOVA for Groups' Performances on Term Paper by Two Raters	163
31.	Paired Samples Test for Academic Writing Skill Survey	164
32.	Wilcoxon Signed Ranks Test for Individual Survey Item	165
33.	Pearson's Correlation Coefficients for CCTT-X Performance and Academic Writing Ability by Group of Students	167
34.	T-Tests on Students' Perceptions of Whether They Thought They Benefitted More from Being the Assessor/Reviewer or the Assessee/Reviewee and Their Academic Writing and Critical Thinking Ability	170
35.	Spearman's Correlation Coefficients for Individual Group of Students' Perceptions towards the Learning Activity and CCTT-X Performance	174
36.	The Number of Students' Who Believed They Benefitted More as Assessors/Reviewers or as Assesseees/Reviewees	206

## TABLE OF CONTENTS

<b>ACKNOWLEDGEMENTS .....</b>	<b>i</b>
<b>ABSTRACT .....</b>	<b>iii</b>
<b>LIST OF ABBREVIATIONS .....</b>	<b>v</b>
<b>LIST OF FIGURES .....</b>	<b>vi</b>
<b>LIST OF TABLES .....</b>	<b>viii</b>
 <b>CHAPTER ONE: INTRODUCTION .....</b>	 <b>1</b>
1.0 Introduction.....	1
1.1 Background of the Research .....	4
1.2 Scope of the Study .....	5
1.3 Problem Statement.....	6
1.3.1 Approaches to Teaching Writing: Product vs. Process .....	8
1.3.2 The Role of Peer Evaluation and the Scoring Rubric in the Process-Oriented Approach .....	9
1.4 Research Objectives.....	11
1.5 Significance of the Research.....	12
1.6 Research Questions.....	13
1.7 Organisation of Chapters .....	14
 <b>CHAPTER TWO: LITERATURE REVIEW.....</b>	 <b>16</b>
2.0 Introduction.....	16
2.1 The Etymology and Brief History of Critical Thinking .....	16
2.2 Conceptions of Critical Thinking .....	17
2.3 Critical Thinking in Academic Writing .....	21
2.4 Fostering Critical Thinking via Assessment.....	23
2.4.1 Assessment Purposes .....	24
2.4.2 Assessment and Development of Critical Thinking.....	26
2.4.2.1 Approaches to Developing Critical Thinking .....	29
2.4.2.2 Peer Evaluation as an Approach to Develop Critical Thinking Skills in Tertiary Academic Writing .....	42

2.5 Key Issues in the Implementation of Peer Evaluation Activities to Promote	
Critical Thinking in Academic Writing .....	45
2.5.1 Validity and Reliability of Grading .....	45
2.5.2 Evaluation Ability .....	49
2.5.3 Student Engagement .....	51
2.5.4 Grouping of Students .....	53
2.5.5 Time Factor .....	59
<b>CHAPTER THREE: THE DEVELOPMENT OF THE CAWAR .....</b>	<b>61</b>
3.0 Introduction.....	61
3.1 Theoretical Perspectives and Context for Guiding the Development of the CAWAR .....	61
3.2 Development of the Critical Thinking for Academic Writing Analytical Rubric (CAWAR).....	63
3.2.1 Examining the Literature and LE 4000 Course Objectives .....	64
3.2.1.1 Critical Thinking and Cognitive Taxonomies.....	64
3.2.1.2 English for Academic Writing (LE 4000) Course Objectives and Expected Learning Outcomes .....	75
3.2.2 Parts of the CAWAR .....	78
3.2.2.1 Criteria/Dimensions .....	78
3.2.2.2 Scales .....	79
3.2.2.3 Benchmark Descriptors.....	80
3.2.3 Establishing the Validity and the Reliability of the CAWAR.....	80
3.2.3.1 Subject Matter Experts' (SME) Review of the CAWAR.....	87
3.2.3.2 Empirical Trials .....	109
<b>CHAPTER FOUR: METHODOLOGY .....</b>	<b>114</b>
4.0 Introduction.....	114
4.1 Design of the Study.....	114
4.2 Research Procedures .....	116
4.2.1 English for Academic Writing course (LE 4000) .....	118
4.2.2 Participants.....	120
4.2.3 Briefing and Training the Instructors .....	122

4.2.4 Pre- and Post-test Instruments and Administration.....	124
4.2.4.1 Cornell Critical Thinking Test Level X (CCTT –X) .....	125
4.2.4.2 Questionnaires.....	128
4.2.4.3 Term Paper.....	129
4.2.5. Training the Students .....	132
4.2.6 Interventions.....	132
4.2.6.1 Self-Evaluation Group .....	133
4.2.6.2 Peer Evaluation Group.....	133
4.2.6.3 Peer Review Group .....	134
4.2.7 Classroom Observations.....	136
4.2.8 Interviews.....	137
4.2.9 Data Analyses.....	138
4.2.9.1 Analyses of Quantitative Data .....	139
4.2.9.2 Analyses of Qualitative Data .....	139
4.3 Ethical Considerations .....	140

## **CHAPTER FIVE: RESEARCH FINDINGS AND INTERPRETATIONS:**

<b>QUANTITATIVE ANALYSES.....</b>	<b>141</b>
5.0 Introduction.....	141
5.1 Participants' Performance on the Cornell Critical Thinking Test .....	142
5.2 Checking the Possible Effects of Extraneous Variables .....	147
5.3 Developing Critical Thinking Skills Using Peer Evaluation vs. Self-Evaluation vs. Peer Review .....	149
5.3.1 Performance on the Cornell Critical Thinking Test.....	150
5.3.2 Performance on the Term Paper.....	162
5.3.3 Students' Perceptions about the Extent to Which They Had Developed Their Critical Thinking Skills in Academic Writing.....	163
5.3.4 Relationship between Students' Performance in the CCTT-X and Academic Writing.....	166
5.4 Students as Assessors vs. Assesseees .....	168
5.5 Students' Perceptions of Peer Evaluation to Foster Critical Thinking Skills in Academic Writing.....	171

5.5.1 The Potential of Activities to Foster Critical Thinking Ability in Academic Writing as Perceived by the Groups .....	171
5.5.2 Students' Perceptions of the Potential of the Activities to Develop their Critical Thinking Ability and Their Actual Performance in the Cornell Critical Thinking Test (CCTT-X).....	174
5.6 Key Findings.....	175

## **CHAPTER SIX: QUALITATIVE ANALYSES: A CROSS-EXAMINATION OF**

### **THE QUANTITATIVE FINDINGS .....177**

6.0 Introduction.....	177
6.1 The Development of Critical Thinking Skills Levels within and among the Groups.....	177
6.1.1 The Evidence of Critical Thinking from the Questionnaires and Interviews ...	177
6.1.1.1 The Peer Review Group.....	178
6.1.1.2 The Self-Evaluation Group .....	179
6.1.1.3 The Peer Evaluation Group.....	180
6.1.2 Explanations for the Promotion of Critical Thinking Skills through the Learning Activities .....	181
6.1.2.1 Learning Environment: Collaboration or Individual Activity? .....	182
6.1.2.2 Learning Instruments: The CAWAR or Checklist?.....	191
6.1.2.3 Implementation of Learning Activities .....	196
6.1.2.4 Students' Previous Experience with the Learning Activities and Motivation to Learn .....	201
6.2 Perceptions of Benefitting More as an Assessor/Reviewer or Assessee/Reviewee .	205
6.2.1 The Peer Evaluation Group.....	206
6.2.2 The Peer Review Group .....	210
6.2.3 Conclusion .....	213
6.3 Students' and the Instructors' Perceptions of the Peer Evaluation Activity Using the CAWAR to Foster Students' Critical Thinking Skills in Academic Writing....	214
6.4 Key Findings.....	219

## **CHAPTER SEVEN: DISCUSSION AND CONCLUSIONS .....224**

7.0 Introduction.....	224
-----------------------	-----

7.1 Summary of the Research Findings .....	224
7.1.1 Students' Learning Performance: Findings from Quantitative Analyses .....	224
7.1.2 Possible Explanations for Students' Performance: Findings from the Qualitative Analyses .....	225
7.1.3 Students' Perceptions of the Learning Activities: Findings from Quantitative and Qualitative Analyses .....	227
7.2 Current and Previous Findings Compared and Contrasted.....	228
7.3 Implications for Practice .....	233
7.3.1 Pedagogical Considerations .....	233
7.3.2 How Peer Evaluation Using the CAWAR Can Be Effectively Implemented ..	241
7.4 Contributions of the Study .....	243
7.5 Limitations of the Study .....	245
7.6 Recommendations for Future Research .....	247
7.7 Concluding Remarks.....	248
<b>REFERENCES.....</b>	<b>251</b>
<b>APPENDICES .....</b>	<b>283</b>



## CHAPTER ONE: INTRODUCTION

### 1.0 Introduction

In recent years, education systems in many countries have shown an interest in monitoring what their students gain from their educational experiences both for the students' own self-actualisation and as returns to the country's development. This is especially prevalent in the present rapidly changing economic, cultural, social, technological and global environment that requires one to possess more than a sheer stream of knowledge but also acquire skills which are transferable to different situations. Only then can one explore one's full potential, do or become what one is capable of and, thus be contributing members of the society and valuable assets for the nation's development. Abraham Maslow in his hierarchy of needs (Maslow, 1943) calls this self-actualisation. In the Malaysian curriculum, for example, such an expectation of learning is well-defined. *Code of practice: Quality assurance in public universities of Malaysia* (2004) states that

The challenge for universities is to teach enough factual knowledge and practical skills, and more importantly, to also encourage students to be inquiring, to critically appraise problems, evaluate and offer creative solutions as well as to develop in them a sense of professionalism and attitudes that are desirable for society, in particular Malaysia society. This requires the use of a variety of teaching-learning and assessment methods that will enable students to acquire a range of competencies that are necessary for them to participate in the education process, national and global development and the growth of their disciplines through a process of continuous improvement. (p. 19)

In the Malaysian context, the need for developing critical thinking among the students is evident. Hashim and Hussein (2003) reveal that despite the six years of primary and five years of secondary schooling, students fail to apply the content knowledge learned at school

in real-life problems. The business sector also reports employing graduates incapable of dealing with simple tasks and assignments (Hashim & Hussein, 2003).

Such a situation might be due to the probable lack of opportunities to employ critical thinking in class due to the learning and teaching approach in Malaysia. This contributes to an urge for policy makers and education stakeholders to gradually reform educational goals and objectives to include critical thinking. At the tertiary level particularly, the aim of instruction is no longer just about assessing students' learning but to teach students to assess their own thinking (Stone, 2001). A thinking-centred learning environment should be initiated to encourage students to constantly evaluate their thinking on issues of their discipline and make corrective adjustments as needed. In doing so, students reshape their thinking on their way to becoming a critical thinker.

Recognised as a skill of enduring importance, critical thinking has been listed among the key components of educational objectives of many countries across the world including the UK (Russell, 2000), Australia (Department of Education and Training, 2006), New Zealand (Ministry of Education, 2005; 2007), Singapore (Goh, 1997) and Malaysia (Radin Umar, Saleh, Wahid, Jamaluddin, Haslinda, & Nor Azizah, 2006). Since the shift of attention away from teaching information and content towards developing thinking skills is quite recent, continuing attempts are observed to find ways to infuse critical thinking into the tertiary curriculum.

The development of critical thinking skills through formal education is mainly done in two ways. The first way is through stand-alone critical thinking subjects where critical thinking is taught as a course on its own. The second way is by integrating the skills across the

curriculum. One of the courses where the teaching of critical thinking is prevalent is writing. This is because writing is a medium for self-expression and critical analysis (Lavelle, Smith, & O’Ryan, 2002). Gammil (2006) also argues that writing exercises encourage students to develop metacognitive and reasoning skills, and the ability to analyse and synthesise information. “No other exercises in the classroom generate higher thinking skills than does writing” (Gammil, 2006, p. 760). The evident relationship between writing and critical thinking is expressed by Boyer (1983) who wrote that “clear writing leads to clear thinking. Clear thinking is the basis of clear writing. Perhaps more than any other form of communication, writing holds us responsible for our words and ultimately makes us more thoughtful human beings” (p. 90). Ryder (1994) adds that “writing is a cognitive process involving the shaping of thought” (p. 215).

There are different approaches to the development of higher order thinking skills, such as question-based learning, collaborative learning, problem-based learning and e-learning (Radin Umar et al., 2006). Peer evaluation has also been identified as a potentially effective technique to produce quality writing. The effectiveness, however, lies in the different approaches adopted by instructors (what instructors have students do during peer evaluation) as to whether they are likely to accelerate the development of critical thinking. One possible method for further fostering critical thinking is proposed in this study and involves the adoption of critical thinking rubrics for students to use during the peer evaluation activity.

A rubric is “a tool that guides the production of students’ work as well as a tool for assessment by presenting expected performance criteria and levels of performance quality” (Andrade, Wang, Du, & Akawi, 2009, p. 287). It may be used to guide feedback in summative or formative assessments. Rubrics are useful not only for making consistent and

efficient evaluation of students' performances but also for informing students of the instructor's expectations and providing feedback for further learning (Peirce, 2006; Jonsson & Svingby, 2007). Recognising the potential of rubrics, this study seeks to develop a rubric for students' use to help develop critical thinking in academic writing.

### **1.1 Background of the Research**

The study was conducted at the International Islamic University Malaysia (IIUM). The University which celebrated its 28th anniversary in 2011 is one of 20 government universities in Malaysia. It operates under eight sponsoring governments and the Organisation of Islamic Conference (OIC). The main campus, covering 700 acres is situated in Gombak, a suburb of the capital city of Malaysia, Kuala Lumpur. Other campuses include (1) Kuantan Campus in Pahang where the Medical-Science, Pharmacy, Allied Health Sciences, Nursing and Dentistry faculties are located; (2) Centre for Foundation Studies at two different locations; Petaling Jaya, Selangor and a satellite campus at Bandar Baru Nilai, Negeri Sembilan prepares students for entrance to degree programmes; and (3) International Institute for Islamic Thought and Civilisation (ISTAC) at Persiaran Duta, Kuala Lumpur, a faculty offering degrees in Islamic civilisation and Islamic science. The University was established under the Company Act instead of the University Act to make possible the use of English as its medium of instruction and the language of administration. The students who come from more than 90 different countries have English as either their mother tongue, a second language or a foreign language. The University adopts an outcomes-based education approach, "a method of curriculum design and teaching that focuses on what students can actually do after they are taught" (Acharya, 2003, p. 1).

The University at which the researcher was previously a student was chosen because it offers the types of classes needed for this doctoral study. The *English for Academic Writing* course (LE 4000) is compulsory for all undergraduate students of the University and is a graduation requirement. The course objectives match the requirements of the research: the development of critical thinking. The outcomes of this research therefore, could benefit such classes. Not only might the rubric be used to develop students' critical thinking, but it could also provide a means for assessing critical thinking in the course.

### **1.2 Scope of the Study**

This study sought to develop and validate an analytical critical thinking rubric as an instructional tool for use by students in peer evaluation activities in academic writing lessons. Individually, students were required to produce drafts of a piece of writing and at each drafting stage they worked together in small groups to assess each others' work to help improve the quality of writing. Thus, each student assessed and was assessed using the rubric.

There are two broad types of rubrics. An analytic rubric contains a detailed scoring guide where scores are to be assigned to each dimension separately. A holistic rubric leads to an overall mark where a rater judges the quality of a product as a whole (Moskal, 2000). Generally both approaches have advantages and disadvantages. Specifically, the analytic rubric can indicate exact areas for improvement and provide more reliable scoring, but it is a more time-consuming assessment. For a holistic rubric, the opposite applies (Chatterji, 2003; Johnsson & Svingby, 2007). The analytic rubric has an advantage for the promotion of critical thinking skill development. The concise and detailed descriptions of performance criteria that outline expected achievement provide step-by-step directions for students about the quality of each aspect of the writing assignment. For example, criteria may include the

clarity of the thesis statement, the reliability of literature, the accuracy of making citations and the coherence of information. To be exact, assessors have to be analytical as they judge the quality of a work according to each assessment criterion while the assesseees have to also give a considerable amount of thought to their performance in relation to each skill as indicated by the grades received and later find ways to improve them. For this reason, an analytic instead of a holistic critical thinking rubric was developed for this study.

The scoring rubric is named the **C**ritical Thinking for **A**cademic **W**riting **A**nalytical **R**ubric (**CAWAR**). The rubric was based on models of critical thinking skills relevant to writing in the literature and from the expected learning outcomes of the academic writing course. The rubric was reviewed by subject matter experts (SME) chosen by the researcher. It was later field-tested before it was applied in a quasi-experimental study in an *English for Academic Writing* course aiming to help develop undergraduate students' academic writing skills. The development and use of the rubric are elaborated on in Chapter Three.

### **1.3 Problem Statement**

Tertiary writing differs from secondary school writing by requiring tertiary students to produce writing of a more critical academic nature. Objectivity and conciseness are the goals of such writing, and certain stylistic rules and guidelines need to be adhered to. All these entail tertiary students needing to critically examine and be engaged with the ideas and issues of the discipline within which they write considerably more complex and detailed texts using an appropriate academic voice (Vardi, 1999). This is a challenge for the majority of undergraduate students whose writing lessons at school have not focused on such skills. Instead their earlier writing skills are developed based on learning experiences that emphasise language accuracy and mechanics. At the undergraduate level, however, irrespective of the

language medium of instruction, the ability to write well does not only depend on the quality of language used but also how clear and well-thought out the ideas are. This scholarly writing requires considerable critical thinking by the writer and is integral to the production of insightful and thought-provoking writing.

The method for teaching of critical thinking is relevant to writing classes. The traditional approach of instructing students to transcribe language in written form, teaching spelling and grammatical conventions, and focusing on good writing style through modeling, is no longer the main objective of writing lessons (Galbraith & Rijlaarsdam, 1995).

The question then becomes what is a suitable approach to adopt to help promote critical thinking for academic writing. One approach is to use peer evaluation which requires students to be critical in assessing others' work, defend their own work, and subsequently make improvements. However, although peer assessment is a common teaching technique at tertiary level, it is not well researched (Haswell, 2005). The first comprehensive review of research on any kind of tertiary-level peer assessment by Topping (1998) found only 67 empirical studies. Thus, it is not surprising that the number of studies of its potential to develop critical thinking is also very limited. This study, therefore led to the development of an instructional critical thinking rubric to function not only as a scoring guide during peer evaluation but also as a tool to promote critical thinking as students assess their peers' written work. The rubric is deemed important due to the absence of such a rubric especially for students' use in academic writing lessons.

In this thesis, the term "evaluation" has been used specifically to mean the *grading* of academic writing guided by a set of assessment criteria. Peer evaluation also included written

and oral comments. Self-evaluation in this study did not include any input from peers. The term “review” is distinguished from “evaluation” and has been used to refer to an activity that involves peers using a set of assessment criteria to comment on a piece of writing, but without including the grading aspect.

### **1.3.1 Approaches to Teaching Writing: Product vs. Process**

Two basic approaches to teaching writing are the product-oriented and the process-oriented approaches. The former is the traditional approach emphasising the end-product. According to Badger and White (2000), “product-based approaches see writing as mainly concerned with knowledge about the structure of language, and writing development as mainly the result of the imitation of input, in the form of texts provided by the teacher” (p. 154). The product-oriented approach takes writing as a linear process following fixed stages: pre-writing, drafting, revising, editing and publishing (Kim & Kim, 2005). This is a teacher-centred approach in which the teacher plays the authoritative role in providing feedback and assessing the work (Xu, 2005).

The process-oriented approach to teaching writing began in the latter half of the 20th century (Matsuda, 2009) and has gained a strong advocacy over the product-oriented approach due to its compelling effectiveness in developing writing skills (Westervelt, 1998). It is “an approach to the teaching of writing which stresses the creativity of the individual writer, and which pays attention to the development of good writing practices rather than the imitation of models” (Tribble, 1996, p. 160). In a process approach, emphasis is on how writers work towards expressing ideas. Linguistic knowledge is given less attention compared to the product approach (Badger & White, 2000). Esperet and Piolat (1990) argue that the outcomes of writing are determined by how much one controls the text during writing.



With the shift of interest to the process approach, the focus becomes directed on the writing sub-processes of planning or pre-writing, drafting, revising and editing before publishing as a non-linear process. This view of the writing process involving a series of interactive, recursive phases is the result of the reworking of thoughts and ideas by moving back and forth among sub-processes of writing (Emig, 1971; Flower & Hayes, 1981; Zamel, 1983; Spandel, 2001). This entails complex cognitive activity requiring attention at various levels: thematic, paragraph, sentence, grammatical and lexical (Biggs, 1988). Multiple drafts would therefore be produced especially as re-writing is encouraged. In doing so, students learn to be thorough, productive, and independent as they work (Tarnopolsky & Kozhushko, 2007). The process approach meets the requirement of effective peer evaluation using a rubric which aims to provide feedback to students as they work on improving their piece of writing.

### **1.3.2 The Role of Peer Evaluation and the Scoring Rubric in the Process-Oriented Approach**

A shift from a product approach to teaching writing to a process approach has led to a change of interest in the nature and timing of feedback. With the process-oriented writing approach, formative feedback from both peers and the teacher has gained prominence over summative feedback from the teacher that mostly comes in the form of a letter grade with a brief end comment justifying the grade (Matsuda, 2003; Susser, 1994). Hyland and Hyland (2006) state that “summative feedback, designed to evaluate writing as a product, has generally been replaced by formative feedback that points forward to the student’s future writing and the development of his or her writing process” (p. 1). “The reference point of feedback also expanded from lexical and syntactic issues to genre-specific features as well as larger sociocultural contexts of writing” (Matsuda, 2009, p. 75).

The advent of the process approach to teaching writing makes it possible for formative feedback to be generated via peer evaluation, an alternative source of feedback from the teacher. Rieber (2006) defines peer evaluation as “a process in which students individually, or in groups, evaluate and assign grades to other students’ work” (p. 322). Topping (1998) agrees that as formative assessment, peer evaluation “aims to improve learning while it is happening in order to maximize success rather than merely determine success or failure only after the event” (p. 249). It can encourage students to be self-regulated learners (Nicol & Macfarlane-Dick, 2006) capable of regulating their thinking, motivation and behaviour during learning (Pintrich & Zusho, 2002). By adopting peer evaluation, the “sequence of writing and evaluation exercises provides students with an authentic writing task and places it within a social context that is ideal for the process-writing classroom... [and it] provides what students consider the most important aspect in an evaluation of their writing – a general measure of the extent to which an author’s aims are validated by audience” (Lifftig, 1990, p. 65).

Although it involves increased time and effort, the writing process that students undergo in peer evaluation activities results in a more polished product of writing (Reese-Durham, 2005). As peer evaluation is guided by a scoring rubric, the activity allows them to internalise the qualities of good writing (Johnson, 2001). Even more importantly, peer evaluation helps to develop students’ critical faculties (Evans, 2008). Through being evaluated, “students refine their thinking, consider the quality of their communication, and acquire a heightened sense of the reader’s understanding and interpretation of the writing” (Ryder, 1994, p. 215). As evaluators and feedback providers, on the other hand, students develop their judging skills, the fundamental skills for study and professional life (Brown, Rust, & Gibbs, 1994). Furthermore, as students evaluate others’ work, they are provided with insights into the

quality of their own work (Bostock, 2001). Topping (1998) argues that peer evaluation benefits assesses and assessors by: (1) triggering higher order thinking especially through questioning and reflecting; (2) increasing students' time on task, engagement and sense of accountability; (3) helping students identify and fill in knowledge gaps in their work; and (4) encouraging autonomous learning.

#### **1.4 Research Objectives**

This study comes out of concerns about the lack of critical thinking skills necessary for producing academic writing in undergraduate students. Therefore, the aim of this research was to find a way to foster students' critical thinking in academic writing classes. Hence, the researcher aimed at developing and validating an analytic scoring rubric as a means to help students become more critical as they assess each others' work during peer evaluation.

The specific aims of the study were:

1. to design a critical thinking rubric as a scoring and feedback guide during peer evaluations of critical thinking skills in academic writing classes;
2. to validate the critical thinking rubric (CAWAR);
3. to see the extent to which the rubric helps to develop students' critical thinking in peer evaluation activities in comparison with self-evaluation (i.e. students assess and grade their own work (Rauch & Fillenworth, 1993)) and peer review activities (i.e. "students submit their writing to peers who review (or edit) it but do not assign grades" (Rieber, 2006, p. 322);
4. to analyse students' and teachers' perceptions of the usefulness of the rubric in peer evaluation to foster critical thinking skills; and

5. to propose possible ways to promote critical thinking in academic writing using the rubric for peer evaluation.

It is hoped that the rubric developed in this study might act as a trigger for critical thinking development in students as they assess their peers' academic writing.

### **1.5 Significance of the Research**

Kuhn (1999) states that “enthusiasm for critical thinking as a goal of education shows no sign of waning” (p. 16). Since the development of critical thinking skills is gaining prominence in higher education in particular, the development of the skills needs to be carefully planned through appropriate teaching and learning approaches using valid instructional tools.

One approach is peer evaluation. Black, Harrison, Lee, Marshall and William (2003) stress that while assessing each other, “the students [are] expected to think, to assess themselves, to accept challenging expectations and be collaborative learners” (p. 97). These activities provide opportunities for students to work collaboratively towards attaining a learning goal. In fact, it is a way to make students actively engaged in the learning process as students play roles as both assessee and assessor. Students being assessed by their peers (the assessees) are more likely to produce the best quality work possible so as not to reveal their weaknesses to their peers. Students who assess their peers (the assessors), on the other hand, have to carefully evaluate the work before providing feedback and in doing so develop their assessment skills. Black et al. (2003) argue that peer assessment provides the practice ground for self-assessment. Peer evaluation and self-evaluation are more effective when the success criteria are made transparent to the students in advance. A rubric outlining the key criteria

and expected levels of performance is therefore, deemed to be a very important instrument for facilitating the peer evaluation process.

In this study, the *CAWAR* was designed to guide learning and to provide feedback from peers. In other words, the rubric was used as a tool in assessment *as* learning and assessment *for* learning as opposed to assessment *of* learning (discussion on the three is provided in Chapter Two). This is realising the capacity of the rubric to activate students' metacognition or "the skills associated with the learners' awareness of his or her own thinking" (Presseisen, 2001, p. 52). This includes reflecting and monitoring their progress towards future learning goals. Hence, this study hopes to provide for the mediation of learning more than an assessment of how much learning has taken place. Specifically, the *CAWAR* was designed to promote students' critical thinking skills so that they become critical readers and writers of academic texts. Therefore, it is hoped that this study can contribute to a critical thinking rubric for use by academic writing instructors wishing to adopt peer evaluation.

To achieve this, the *CAWAR* was carefully developed to ensure its validity and reliability. Following its creation, it was evaluated further in a quasi-experimental study to see its potential when used in peer evaluation activities in comparison to self-evaluation using the same rubric and peer review using a checklist. The study resulted in a clearer indication of how the *CAWAR* can be effectively used to develop critical thinking in academic writing.

## **1.6 Research Questions**

This study investigated the potential of the *CAWAR* to help promote students' critical thinking skills. Specifically the study sought to answer the following questions:

1. Is there evidence of students developing greater critical thinking skills when they use the *CAWAR* in peer evaluation exercises to assess academic writing than in self-evaluation activities, or when using a checklist in peer review activities?
2. Do students develop their critical thinking skills better when they use the rubric to assess their peers, or when being assessed by their peers?
3. To what extent do teachers and students perceive the peer evaluation activity using the *CAWAR* as capable of fostering students' critical thinking in academic writing?

### **1.7 Organisation of Chapters**

The rest of the chapters of this thesis are organised as follows:

The second chapter consists of a literature review presenting the etymology and a brief history of the idea of critical thinking, an overview of critical thinking conceptions and their implications for learning and teaching, and the role of assessment for learning enhancement particularly for critical thinking development in academic writing. The chapter also discusses issues pertaining to the use of peer evaluation activities and critical thinking skills development in academic writing especially via peer assessment. Chapter Three describes the development and validation of the *CAWAR*. The fourth chapter presents the research methodology. The fifth and sixth chapters present the findings to the research questions through the quantitative and qualitative data analyses respectively. Specifically, Chapter Five provides the findings to the analyses of (1) the performances of the peer evaluation, self-evaluation and peer review groups in the Cornell Critical Thinking Test to see any difference among them, (2) the peer evaluation group's perceptions of benefitting from being the assessors or assesses and (3) the extent to which the students' perceived the learning activities as helpful to promote critical thinking skills in academic writing. Chapter Six, in contrast, presents the evidence and possible explanations to support the quantitative findings through a

detailed examination of the qualitative materials from interviews, responses to questionnaire survey, observations and students' use of the commentary space in the *CAWAR* and the checklist. Chapter Seven summarises the findings, discusses the issues that emerged from the analyses and also discusses the implications for teaching and learning. The contributions and the limitations of the study are then presented before presenting recommendations for further research and ending the report with some concluding remarks.

## CHAPTER TWO: LITERATURE REVIEW

### 2.0 Introduction

This chapter examines the literature related to the potential of peer evaluation using a rubric to promote the development of critical thinking and rationalises the relevance of the study. The chapter first provides the etymology and a brief history of the idea of critical thinking followed by an overview of the various conceptions of critical thinking and their implications for learning and teaching. Research relevant to the development of critical thinking skills in academic writing via peer evaluation is discussed in three main parts: critical thinking in academic writing; fostering critical thinking via assessment; and key issues in the implementation of peer evaluation activities to promote critical thinking in academic writing.

### 2.1 The Etymology and Brief History of Critical Thinking

The word 'critical' derives from the ancient Greek 'kritikos' meaning discerning judgment and also 'kriterion' which means standards, thus implying the development of "discerning judgment based on standards"(Foundation for Critical Thinking, 2009, para 6).

Critical thinking originates from the Western philosophical traditions of ancient Greece.

From this ancient Greek tradition emerged the need, for anyone who aspired to understand the deeper realities, to think systematically, to trace implications broadly and deeply, for only thinking that is comprehensive, well-reasoned, and responsive to objections can take us beyond the surface. (*The Critical Thinking Community*, 2009, para 4)

Socrates and his followers including Plato and Aristotle are credited with establishing critical thinking. Socrates saw dialogue as useful even if it did not solve a problem or produce a



specific result. Dialogue which fosters critical thinking can clarify problems and bring solutions closer. "Socrates understood himself not as a teacher, but as a midwife easing the birth of critical self-reflection" (Delius, Gatzemeier, Sertcan, & Wünscher, p. 9). Socrates established the need to empower oneself by thinking profoundly over matters and not easily accepting others' thoughts, especially those in authority (The Critical Thinking Community, 2009). He gained a reputation as the ideal critical thinker through his method of questioning and cross-examination of positions (Caroll, 2004).

This research, however did not attempt to delve into Socrates' and his followers' philosophical ideas of critical thinking as the research interest was mainly on the development of the critical thinking skills via the curriculum.

## **2.2 Conceptions of Critical Thinking**

As discussed in Chapter One, the ability to think critically has been identified as one of the learning outcomes of university education. Since critical thinking is so significant in education and the real world of life, it is worth exploring the different ways the term is used. However, like all abstract concepts, critical thinking is hard to define and results in different interpretations in different contexts. As such, critical thinking is also open to definitions from multiple perspectives: philosophical, psychological and educational. Some definitions are broad and some others are narrow. Attempts to define this complex thinking began over 100 years ago and its meaning has evolved since then. One of the earliest definitions was developed by John Dewey, an American philosopher, psychologist and educator who is known as the 'father' of the modern critical thinking tradition (Fisher, 2001). He defines it as an "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, 1938, p. 9). In this definition, Dewey emphasises reflective thinking, asking oneself questions about what to believe through evaluating reasoning, and considering the implications of one's beliefs.

Edward Glaser, the co-author of one of the world's most widely used tests of critical thinking, the Watson-Glaser Critical Thinking Appraisal defines 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 one's experience; (2) knowledge of the methods of logical enquiry and reasoning; and (3) some skill in applying those methods" (Glaser, 1941, p. 5). Compared to Dewey's definition, which stresses the act of thinking, Glaser identifies three elements of critical thinking: having an attitude of being thoughtful when dealing with problems; knowing; and being able to apply the methods of logical enquiry and reasoning.

The urge for an agreed definition was one of the factors leading to the APA Delphi project being conducted. The two-year project (1988-1990) involved 46 American and Canadian panelists representing different disciplines of studies led by Dr. Peter Facione. An international expert consensus definition of critical thinking was determined and is published in the APA Delphi Report entitled *Critical thinking: A statement of consensus for purposes of educational assessment and instruction* (Facione, 1990a). The agreed definition of critical thinking is "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" (p. 2).

Despite the consensus achieved by this large group of experts and the various earlier definitions of critical thinking, attempts to define it still continue as part of the critical thinking tradition to suit the current topic pertaining to the higher order cognitive skills. A later definition by Scriven and Paul (2004) sees critical thinking as “the intellectually disciplined process of actively and skillfully conceptualising, applying, analysing, synthesising, and/or evaluating information gathered from, or generated by, observation, experience, reflection, reasoning, or communication, as a guide to belief and action” (p. 1). These scholars propose an active and skillful use of a set of information processing skills to guide belief and action.

In yet another definition, knowledge is stressed as the basis for the development of alternative ideas and assumptions. According to Yancher and Slife (2003) critical thinking is a cognitive ability that: (1) requires knowledge of the assumptions and underlying world views of a particular discipline or field of inquiry and (2) involves developing ideas and assumptions that are alternatives to existing views. Another relatively recent definition has been provided by Tsui (2003). Tsui sees critical thinking, in a university environment, as involving students’ abilities to identify issues and assumptions, recognise relationships, make correct inferences, evaluate evidence or authority, and deduce conclusions.

According to Hager, Sleet, Logan and Hooper (2003), the most widely accepted characterisation of critical thinking as incorporating abilities and dispositions is due to Robert Ennis (a co-author of other widely used critical thinking test instruments: the Cornell Critical Thinking Test (1985) and the Ennis-Weir Critical Thinking Essay Test (1985)). Underpinning these dispositions and abilities is Ennis’ conception of critical thinking as “reasonable, reflective thinking that is focused on deciding what to believe and do” (Ennis, 1985, p. 54).

This definition has been considered the generic definition by many critical thinking scholars (Huitt, 1988; Fisher, 2001; Jenicek & Hitchcock, 2005). Ennis claims that his conception of critical thinking is superior because it includes the following features (1) a focus on belief and action; (2) makes statements in terms of things that people actually do or should do; (3) includes criteria to help us evaluate results; (4) includes both dispositions and abilities; and (5) is organised in such a way that it can form the basis for a thinking-across-the-curriculum programme as well as a separate curriculum-specific critical thinking course at the tertiary level (Ennis, 1987).

While there are many other definitions of the concept of critical thinking the definitions generally show clearly that critical thinking has both cognitive and affective domains.

### **Implications of Conceptions of Critical Thinking for Learning and Teaching**

Hatcher (2000) states that a clear conception of critical thinking is integral to education. This is especially because the conception might differ according to the context in which it is used. Without a clear understanding of the concept of critical thinking, difficulties await educators who endeavor to teach and measure it. In writing, for example, Flateby (2011) argues that a clear understanding of how critical thinking applies and relates to writing is important before both critical and writing skills can be developed and assessed.

How critical thinking is conceptualised determines the content of a course and the course assessment. What to include and exclude from a course in critical thinking tends to vary if there is no agreement among teachers over what constitutes critical thinking. Assessment will also be adversely affected as the disagreement will lead to difficulty in setting standardised tests to measure how much critical thinking learning has taken place. In actuality, assessment

should reflect what is taught. If the ability to make inductive or deductive reasoning is to be taught then measurement of this specific cognitive skill should be undertaken. If critical thinking dispositions are to be developed, the measurement should be geared to measuring them and in accordance with the pre-defined context (Ab Kadir, 2007).

### **2.3 Critical Thinking in Academic Writing**

Brookfield (1997, p. 18) claims that “critical thinking is “irrevocably context bound... [and] learning to think critically is an irreducibly social process”. This statement supports what was mentioned in the earlier section regarding how the concept of critical thinking is shaped by the context to which it is applied. Critical thinking involves thinking critically about something and that ‘something’ has a direct influence on how much thinking is needed before any decision is made. Critical thinking can be further promoted when there is some interaction taking place as feedback gained about the choice made provides some indication of the quality of the thinking. The object for thinking and its function for communication contribute to the relevance of teaching critical thinking in academic writing lessons. Academic writing provides a context for developing critical thinking skills especially as it involves the expression of certain ideas to be shared with a particular audience. Communication between the writer and the reader via the text is indeed a social act. The writer shares ideas and views with an audience and the feedback obtained from the audience benefits the writer’s development of critical thinking.

A statement by Condon and Kelly-Riley (2004, p. 56) that “writing is a tool of thinking” marks the link between critical thinking and writing. Since critical thinking is the central concept underpinning the development of the Critical Thinking for Academic Writing

Analytical Rubric (CAWAR) within this study, it was crucial for this research to have a clear understanding of how critical thinking pertains to academic writing.

Sorrell, Brown, Silva and Kohlenberg (1997) clarify that writing requires one “[t]o connect ideas from internal and external sources, critically think about the ideas, and then infer a generalisation that gives the separate pieces of information a coherent verbal shape” (p. 14). Bean (2001) asserts that writing “requires analytical or argumentative thinking and is characterized by a controlling thesis or statement and a logical, hierarchical structure” (p. 17). Similarly, Schafermen (1991) explains that “writing forces students to organize their thoughts, contemplate their topic, evaluate their data in a logical fashion, and present their conclusions in a persuasive manner” (p. 7). Good writing is therefore a reflection of good critical thinking. The sources of ideas can be from across a variety of texts and those based on observation, experience and reflection (Vardi, 1999). Hence, critical thinking in academic writing is a manifestation of an author’s ability to understand and analyse the ideas, evaluate and synthesise the arguments in a variety of sources before making any conclusions, and then presenting them clearly to an audience. It entails the ability to: understand key concepts and ideas; distinguish the main ideas and arguments from the subordinate ones; judge their relevance and provide reasons; judge the credibility of sources of information; and be able to paraphrase them and later draw conclusions based on all the justifications made. Engaging oneself in all these tasks exercises thinking and heightens it. Olson (1992) argues that thinking can be refined through pre-writing, writing, revising and editing activities. This means, as a writer is engaged in the writing process, the writer is using his/her judgments to evaluate his/her own text and make any necessary changes to express his/her ideas clearly and confidently to readers. Thus, engaging students in critical thinking during [academic]

writing classes is very important but it can only be achieved if the writing assignments foster such work (Reynolds & Moskovitz, 2008).

For the expansion of students' critical thinking skills, teaching and learning methods adopted in the classroom should be able to create a conducive teaching and learning environment (Cousins & Ross, 1993; Sorell et al., 1997). Critical thinking is associated with a deep approach to learning (Gadzella & Masten, 1998) which is, in turn, likely to be adopted when writing essay assignments (Elander, Katherine, Norton, Robinson, & Reddy, 2006). Therefore, there is a need to investigate how classroom activities can be adapted to encourage deep learning. Paul and Elder (2005) contend that routine writing practice using the appropriate critical thinking tools is important to be an accomplished writer. Such practice will lead to deep learning and make the writer a more effective communicator of ideas.

#### **2.4 Fostering Critical Thinking via Assessment**

Assessment in education plays an important role in tying instruction to learning to meet certain educational goals specified by education administrators and policy makers, and to satisfy the demand of stakeholders that include students, parents, communities and employers. "The new assessment culture aims at assessing higher order thinking processes and competencies instead of factual knowledge and lower level cognitive skills, which has led to a strong interest in various types of performance assessments" (Jonsson & Svingy, 2007, p. 131). "Assessment works best when its purpose is clear, and when it is carefully designed to fit that purpose" (Earl, Katz, & Western and Northern Canadian Protocol for Collaboration in Education (WNCP) Assessment Team, 2006, p. 13).

### 2.4.1 Assessment Purposes

The three main groups of purposes of assessment are to assess *for learning*, to assess *as learning* and to assess *of learning* (Earl et al., 2006). The three are inter-related with assessment *of learning* more easily distinguished from assessment *for* and *as learning*.

Assessment *of learning* aims to check students' learning, usually at the end of a course, and is summative in nature. It summarises what students are able to do and what they know in comparison with the expected learning outcomes specified in the curriculum or how students achieve in relation to other students (Earl et al., 2006). The evidence gathered from assessment of learning may be used to monitor students' progress, compare standards among learning institutions and plan improvement procedures (Learning and Teaching Scotland, 2007; Oosterhof, 2009). The results of assessments may be accessible for public view "as evidence of achievement to parents, other educators, the students themselves, and sometimes outside groups (e.g. prospective employers and other educational institutions)" (Earl et al., 2006, p. 55).

In contrast, assessment *for learning* serves to promote learning as it progresses via the feedback gathered from the assessments. It is formative assessment. A comprehensive definition of the assessment *for learning* is given by Stiggins, Arter, Chappuis and Chappuis (2004) who refer to it as assessment conducted "throughout teaching and learning to diagnose student needs, plan our next steps in instruction, provide students with feedback they can use to improve the quality of their work, and help students see and feel in control of their journey to success" (p. 31). Assessment *for learning* is more than just ongoing assessments of students' progress. It also includes how students can benefit from the feedback to help them to further improve in learning (Black and Wiliam, 1998; The Assessment Reform Group,



1999; Chappuis, Stiggins, Arter & Chappuis, 2003; Stiggins et al., 2004). Effective assessment for learning requires high quality interactions between teachers and students, and between students and peers (Learning and Teaching Scotland, 2007). Being made aware of their responsibility to make progress, students can monitor their learning by evaluating the feedback gained from their teacher and peers.

Many authors use the term assessment *for* learning to encapsulate the idea of assessment *as* learning (as cited in Earl et al., 2006). One distinctive feature is that assessment *as* learning focuses more on the students who are expected to play an active role in their own learning and assessment rather than passive recipients of knowledge or feedback. Being informed of the learning goals and success criteria, students check their own progress by reflecting on evidence of their learning (Learning and Teaching Scotland, 2007). It is stated that “the ultimate goal in assessment *as* learning is for students to acquire the skills and the habits of mind to be metacognitively aware with increasing independence” (Earl et al., 2006, p. 42). This element of self-review or self-regulation contributes to the overlap between assessment *for* learning and assessment *as* learning. Students need to

- possess an appreciation of what high quality work is,
- have the evaluative skills necessary for them to compare with some objectivity the quality of what they are producing in relation to the higher standard, and
- develop a store of tactics or moves which can be drawn upon to modify their own work. (Sadler, 1989, p. 119)

### **2.4.2 Assessment and Development of Critical Thinking**

Assessment serves either to supply evidence that learning has taken place or to help improve students' learning. To achieve the latter in particular, assessment needs to be used as part of the learning process. Shephard (2000) comments:

To accomplish the kind of transformation envisioned, we have not only to make assessment more informative, more insightfully tied up to learning steps, but at the same time we must change the social meaning of evaluation. Our aim should be to change our cultural practices so that students and teachers look to assessment as a source of insight and help instead of an occasion for meting out rewards and punishments. (p. 10)

The idea corresponds to what was argued earlier by Falchikov (1986) that the traditional forms of assessment do not help students develop the competencies they need to face life. Instead it “tends to breed conformity in students and militates against not only personal development but also development of interpersonal skills” (p. 147). Thus, it is timely to consider changes in instruction to help students develop the knowledge and skills which are crucial for facing the life challenges through learner-centred assessment i.e. by self or peers. Brown (1997) argues that “assessment defines what students regard as important, how they spend their time and how they come to see themselves as students and then as graduates ... If you want to change student learning then change the methods of assessment” (p. 7).

As for developing critical thinking skills in academic writing, Flateby (2011) comments, “If the thinking components of writing are not addressed in assessment, writing may be viewed more as a skill set and less as a way to develop and express thought. Similarly, if writing is assigned, but only grammar and mechanics are factored into a grade, students generally will not attend to the broader aspects of writing, such as the development of ideas and audience needs” (p. 4). While classroom teacher assessment may acknowledge the importance of the

thinking aspects of writing, the use of self- and peer assessment during the production of a piece of writing may be more effective. Teacher assessment only indicates to students the areas of the writing that need improvement but self- and peer assessment potentially engage students in the thinking process itself as they try to understand, analyse, clarify, comment and defend each others' work in order to improve it. Thus, students can experience the assessment procedure as well as, more importantly, they can realise "the recursive linkage between learning processes and outcomes" (Green, Christopher, & Lam, 1997, p. 263). Besides, they see for themselves what makes quality writing as they evaluate writing on the basis of the assessment criteria and internalise them (Johnson, 2001) and develop their judging skills (Brown et al., 1994). When assessing, students judge the extent to which their own or their peers' writing has met the criteria. They need to find evidence from the text to justify the grading made. In fact, they learn to give constructive comments to their peers. The more opportunity they get to practise these activities the better their evaluation and judgment skills are enhanced. With the paradigm shift in learner autonomy, student assessment is deemed appropriate (McNamara & Deane, 1995; Green et al., 1997; Creswell, 2000). Student assessment empowers students to achieve greater learning. It allows students to take charge of their own learning and not be restricted to what is imposed on them to learn by the teacher. Hence, students learn to become independent and decisive.

Todd and Hudson (2007) conducted a study to see whether peer evaluation helped to improve students' writing skills, critical thinking ability and comprehension of material in a public relations course. A modified writing to learn (WTL) assignment was used for a peer evaluation activity. Students were first asked to find a magazine advocacy advertisement. Then they were asked to (1) evaluate the advertisement's audience(s) based on its message, publication, topic and request for action and (2) evaluate the communication or message. Two

drafts were required for each of these tasks. For each task, both the first and second drafts were evaluated by an anonymous peer evaluator using a grading rubric. This was followed by a teacher evaluation. The teacher evaluated and graded both the original draft and the comments made by the peer evaluator. Discussions of the comments from the teacher and peer evaluator followed. The student evaluator defended his/her comments and the writer challenged the peer's and teacher's comments. Corrections and modifications were made based on the feedback received before the final papers were submitted for evaluation by the teacher. They found that the peer evaluation activity benefited the students in all three aspects: improving their writing skills, critical thinking skills and knowledge about public relations.

In another study, Odom, Glenn, Scanner and Cannella (2009) aimed to promote higher levels thinking and collaboration among 30 senior nursing students in a research course focusing on how to critique a research article. A peer evaluation activity was used. The study revealed that 95% of the students reported the feedback received through the peer evaluation activity was very beneficial for them in terms of comparing viewpoints, seeing different styles of writing, and clarifying research concepts. Besides, it helped to clarify for them what should be included in a research critique.

Based on the above arguments and studies, it is therefore likely that critical thinking in academic writing would be fostered via peer and self-assessments. In this study, the potential of peer evaluation, peer review and self-evaluation to promote critical thinking in writing was explored.

Armstrong and Paulson (2008) define peer evaluation as an activity that “generally involves students rating/grading/judging their peer’s papers instead of simply responding to it” (p. 403). Strijbos and Sluijsmans (2010) add that “peer assessment is an educational arrangement where students judge a peer’s performance quantitatively and/or qualitatively and which stimulates students to reflect, discuss and collaborate” (p. 265). Peer review contrasts with peer evaluation in that peers read, review and edit someone’s work but no grading is involved (Rieber, 2006). On the other hand, self-evaluation, like peer evaluation, involves grading work but students are asked to grade their own work also using a rubric (Rauch & Fillenworth, 1993). In short, in a peer evaluation activity, students give feedback on peers’ work by grading it based on the assessment criteria listed in a rubric and provide comments to justify and explain the grading which has been made. In a peer review activity, students provide feedback to peers’ work which is also guided by assessment criteria but they are not required to assign a grade to it. In a self-evaluation activity, on the other hand, students individually evaluate their own work according to the assessment criteria listed in a rubric. A more detailed discussion on the three approaches is presented below.

#### **2.4.2.1 Approaches to Developing Critical Thinking**

Peer evaluation, peer review and self-evaluation, which either share the collaborative learning environment or the assessment tool, are all potentially very important for helping to develop skills that should benefit students throughout their studies and professional life (Brown et al., 1994). Most importantly, they help develop students’ critical faculties (Black et al., 2003; Kay, Li, & Fekete, 2007; Evans, 2008). Through peer evaluation and peer review, students are more reflective of their own strengths and weaknesses as they can see the performance of others. While checking their peer’s work, students develop their self-assessment abilities as they reflect and think critically about the content of the work. When self-assessing, the

students are reflecting on their learning process and progress. In other words, students engage in deliberate thoughts about how they are learning and what they are learning. This promotes autonomous learning (Srimavin & Darasawang, 2004; Cassidy, 2006; Saltourides, 2006; Evans, 2008). Guided by their developing critical faculties, students learn to become independent learners.

Both self- and peer assessment require students to be active participants. This activity contributes to their learning (Weimer, 2002; McCombs & Miller, 2008; Murdoch & Wilson, 2008; Blumberg & Weimer, 2009; Campbell, 2010). The evaluative experience helps students develop an evaluative expertise (Sadler, 1989). In fact “inviting students into the shared experience of marking and moderating should also enable more effective knowledge transfer of assessment process and standards” (Rust et al. 2003, p. 152). According to Elander et al. (2006), “critical evaluation has a special status in the context of improving student performance, and familiarising students with the criteria that are applied to their work, and providing opportunities for them to apply those criteria themselves, may be an especially effective method to promote critical and evaluative thinking more generally” (p. 78). The assessment activities also expose and focus students’ attention on the learning objectives being measured (Orsmon, Merry, & Reiling, 2000; Nicol & Macfarlane-Dick, 2006; Lee, 2006).

The three approaches, however, vary in terms of the learning setting provided, the feedback gained and the assessment tools used.

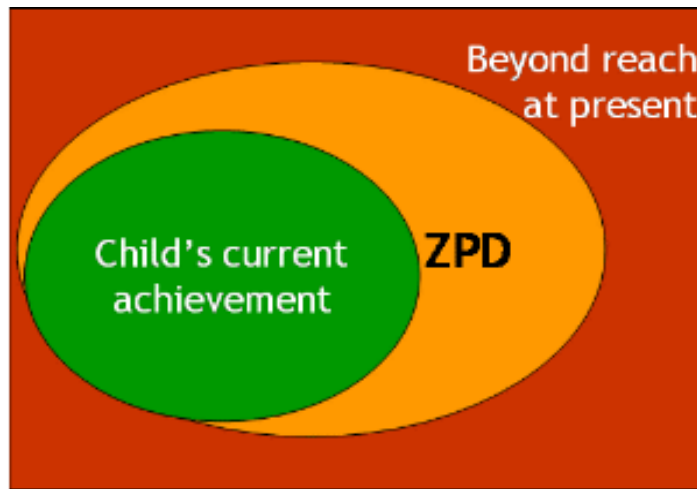
### **Learning Setting**

Working collaboratively entails students either working in pairs or groups to share the learning experience and they are made responsible for each others' learning success. It is a learning environment made possible through peer evaluation and peer review activities.

The terms “collaborative”/“collaboration” and “cooperative” learning connote students working together in a group to improve learning. Thus, the literature of both collaborative and cooperative learning was referred to. However, collaborative learning was mostly referred to as peer evaluation and peer review in the context of this study were more closely linked to the nature of this learning approach. Cooperative learning is structured and teacher directed, focuses on the end product and is targeted at mastery of foundational knowledge. Collaborative learning, on the other hand, distrusts structure, empowers students and aims at personal and social development (Kagan, 1989; Matthews, Cooper, Davidson, & Hawkes, 1995; Oxford, 1997). Besides, as stated by Smith and MacGregor (1992), “Collaborative learning” is an umbrella term for a variety of educational approaches involving joint intellectual effort by students, or students and teachers together” (p. 9). One of the approaches is cooperative learning.

Collaborative learning is influenced by a “social constructivist” philosophy which “views learning as the construction of knowledge within a social context and which therefore encourages acculturation of individuals into a learning community” (Oxford, 1997). The social nature of peer evaluation and peer review fits Vygotsky’s Zone of Proximal Development (ZPD) which is illustrated in Figure 1.

Figure 1. Vygotsky's Zone of Proximal Development (ZPD)  
as Depicted by Atherton's (2011)



In his theory of social development, Vygotsky argues that socialisation affects the learning process in an individual. The individual's learning potential is expanded with the help of a teacher or (more experienced) peers further than what he or she is able to do alone (Vygotsky, 1978). The ability of peer evaluation and peer review to promote learning is characterised by this feature. The supportive feedback provided by peers helps accelerate one's learning.

On the other hand, self-evaluation engages students in an individualised learning environment to assess their own learning. However, self-assessment is believed to be able to “promote intrinsic motivation, internally controlled effort, a mastery goal orientation, and more meaningful learning” (McMillan & Hearn, 2008, p. 40).

Compared to working individually, working in teams helps students to achieve higher levels of thought. Information is also held longer (Johnson and Johnson, 1986; Slavin, 1990). This is supported by Totten, Sills, Digby and Russ (1991) and Gokhale, (1995) who assert that collaborative learning enables students to become critical thinkers. Romney's (1996) list of



advantages of collaborative learning can help explain what enables the promotion of critical thinking during collaborative work. This includes students' willingness to share their difficulties with others in small groups with whom they are familiar, and later gaining confidence when they can solve problems. They learn to accept criticism, as they are also allowed to criticise. Furthermore, during discussions, students learn to tolerate the viewpoint of others. All these are integral to effective learning taking place. Johnson, Johnson and Smith (1991) argue that collaborative learning can be useful "whenever the learning goals are highly important, mastery and retention are important, the task is complex and conceptual, problem solving is desired, divergent thinking or creativity is desired, quality performance is expected and higher level reasoning strategies and critical thinking are needed" p. 40). In addition, Slavin (1995) states that students' motivation to learn may affect the time spent on a task. He reported that most studies measuring time on-task reveal higher proportions of engaged time for students in cooperative learning groups compared to the students working individually.

A few studies were found that compared the effect of group and individual work for promoting critical thinking. Group work was found to be more effective than individual work in promoting the acquisition of high-level cognitive reasoning strategies in a study by Skon, Johnson and Johnson (1981). The study involved 86 first grade students who were randomly assigned to work in one of three learning conditions. They found that students in the collaborative condition consistently achieved more highly than students in the competitive and individualistic conditions on all measures of the given tasks (i.e. free-recall, spontaneous retrieval, categorisation strategy and awareness of strategy for categorisation and retrieval task; explanation for metaphor interpretation; and equations for story problem task).

Gokhale (1995) also found that collaborative learning was more effective than individual learning for enhancing critical thinking. The study involved two groups of 48 undergraduate students enrolled in a Basic Electronics course at Western Illinois University, Illinois. One group was randomly assigned a task to complete individually and the other group completed the task in small groups (the group members were self-selected). A pre- and post-test of critical thinking, developed by the researcher, were administered to all the students. The findings of the study revealed that the students who learned collaboratively performed significantly better than the students learning individually.

In a more recent study, Quitadamo et al. (2009) investigated the effect of Peer-Led Team Learning (PLTL), a specific form of a small group learning method that promotes discourse and creative problem solving, on critical thinking in undergraduate science courses. To determine students' level of critical thinking, the California Critical Thinking Skills Test (CCTST) by Facione (1990b) was used. Results from a quasi-experimental pre-test/post-test control group design revealed that a significantly higher gain in critical thinking was observed for the PLTL students compared to the non-PLTL group.

A neutral result was found in Garside's (1996) study comparing the effectiveness of traditional lecture methods of instruction to group discussion methods of instruction in developing critical thinking skills. One hundred and eighteen students enrolled in an introductory interpersonal communication course participated in the study. The results indicated no significant difference in students' learning via the two methods.

Despite the studies comparing the effect of collaborative learning versus individual learning on the development of critical thinking skills, the search for studies comparing the effect of

peer assessments (peer evaluation and peer review) and self-evaluation on the promotion of critical thinking skills using the two most used education online databases (i.e. EBSCOHost and ProQuest) did not bring out any results. This absence is not surprising given that there are few studies on the use of peer evaluation to promote critical thinking skills. This study addresses the need to offer insights into the possible effects of peer assessments compared to self-assessments.

### **Feedback**

The next aspect that differentiates the activities is the nature of the feedback students receive. Feedback on a student's performance is essential to help students be self-regulated learners. Self-regulation according to Butler and Winne (1995) is

a style of engaging with tasks in which students exercise a suite of powerful skills: setting goals for upgrading knowledge; deliberating about strategies to select those that balance progress toward goals against unwanted costs; and, as steps are taken and the task evolves, monitoring the accumulating effects of their engagement. (p. 245)

The relationship between feedback and self-regulated learning is explained by Nicol and MacFarlane-Dick (2006) below.

Intelligent self-regulation requires that the student has in mind some goals to be achieved against which performance can be compared and assessed. In academic settings, specific targets, criteria, standards and other external reference points (e.g. exemplars) help define goals. Feedback is information about how the student's present state (of learning and performance) relates to these goals and standards. Students generate internal feedback as they monitor their engagement with learning activities and tasks and assess progress towards goals. (p. 200)

The main source of feedback in classrooms is commonly the teachers. This is especially true when the teacher is the authoritative figure in the classroom and when students work

individually to assess their own performance. However, when peer evaluation or peer review is adopted in the classroom, not only do students get feedback from the teacher, they also get it from their peers. Peer feedback is normally more immediate than the teacher's feedback. Getting a rapid response is a benefit that students report as being the main determiner for course effectiveness (Wiggins, 1993). To clarify, in writing classrooms where the teacher assesses the work of all students, students will often have to wait for some time before they get some feedback on their performance due to the time their teacher has to take to mark a large number of essays. The delayed feedback will not benefit the students as much as when rapid feedback is received. After a delay, the students become less attached to the task, and even the joy of getting complements for good work might be reduced. Even worse, when necessary amendments are required, the students find it more difficult to make the corrections as they have become less engaged with the task and might have forgotten some details. In cases where immediate feedback from the teacher is feasible, the feedback might not be as detailed as the students might want. Yet giving thorough comments on students' work is time consuming for the teacher and would result in further delayed feedback on the work of other students.

In addition, by having peers check their work, students are provided with more perspectives on their performance as there are more people reading the work. For academic writing especially, a better-crafted piece of writing is then produced (Todd & Hudson, 2007; Chu-yao, 2008; Saddler & Andrade, 2004). Through peer evaluation and peer review activities, the multiple readers as reviewers and assessors indicate their individual level of satisfaction for the writing and make suggestions for improving the work. Students engaged in a self-evaluation activity with no interaction with peers, however have to rely on feedback only from the teacher.

Reese-Durham (2005) conducted a study to find out (1) the extent to which peer feedback is meaningful and effective; (2) the lessons (if any) peer evaluators learn from a peer evaluation activity; and (3) the extent to which the peer evaluation process results in better research papers for student researchers and peer evaluators. In the study, a graduate educational research class of 19 teachers gave a copy of their partial research paper to two self-selected classmates to review and comment on using a provided evaluation form. The sections of the form corresponded to the required sections in the paper to have been completed at the midpoint of the course including the problem statement, review of related literature, hypothesis and descriptions of the participants. Specific questions for the sections were included. Discussion of the comments with the reviewers then followed. Having experienced the learning activities, all the students agreed that the feedback was helpful, constructive, clear and understandable. The evaluators understood their role as assessors well. The activity allowed students to evaluate their own understanding of what was learnt.

A study by Li, Steckelberg and Srinivasan (2008) investigated student perceptions of a web-based peer assessment system. The participants were 38 undergraduate teacher education students. These students were asked to develop a WebQuest proposal, an instructional strategy developed by Dodge and March in early 1995 aimed at involving users in higher-order learning processes, such as analysis, synthesis and evaluation. In this study, the peer assessment process followed five steps: namely, discussing marking criteria provided by the instructor, submitting their projects for random peer review, assessing peers' projects and providing feedback, viewing and rating peer feedback, and improving the project. A questionnaire was then distributed to each student to solicit their general perception of peer assessment. Results of the students' responses generally revealed that "peer assessment can be effective in minimising peer pressure, reducing management workload, stimulating student

interactions, and enhancing student understanding of marking criteria and critical assessment skills” (Li et al., 2008, p. 8).

A more recent study by Praver, Rouault and Eidswick (2011) examined 86 Japanese university students’ attitudes toward peer evaluation in an intensive English as a Foreign Language (EFL) Reading programme (intermediate proficiency). In particular, the students’ preferences toward peer evaluation for the components of ease, nervousness, embarrassment, and usefulness, when using a numerical scale only (NSO) and a numerical scale together with written comments (NWC) were investigated. The findings revealed that the students were more nervous and embarrassed when engaging in evaluation with NWC and believed that NSO was easier than NWC when providing feedback. However, most students felt that NWC was more useful than NSO. Notwithstanding the perceived difficulty, anxiety, or embarrassment, the students considered that with the additional element of written comments “NWC was the form more likely to aid in their development, more likely to guide future improvement in performance, and to help them gain confidence in the areas that were done well” (Praver et al. 2011, p. 95).

Other studies highlighting the positive effects of peer feedback for learning enhancement via peer evaluation or peer review include Ertmer, Richardson, Belland, Camin, Connolly, Couthard and Mong (2007), Basheti, Ryan, Woulfe and Bartimote-Aufflik (2010) and Jurado (2011). Ertmer et al.’s (2007) study revealed that using peer feedback as an instructional strategy increased the quality of students’ online postings. Similarly, Basheti et al.’s (2008) study indicated that pharmacy students’ anonymous peer assessment of a medication management review (MMR) was perceived useful by the students. In Jurado’s (2011) study, interior design studio classes were also found to benefit from peer feedback by helping them

to focus on their own design, motivating them to work harder, providing viewpoints to improve their design, and sharing of information.

### **Assessment Tools**

Both peer evaluation and self-evaluation use a rubric to guide evaluation. Peer review, on the other hand, does not use a rubric as it does not involve assigning a grade to a task. Rieber (2006) listed three types of peer review. First, open-ended peer review for advanced students who “have the knowledge, ability, and confidence” (p. 323) to check their peer’s work and minimal reliance on the teacher is needed. Second, is guided peer review where students rely on “a list of general questions [that]... typically summarize the directions and ask the reviewer to consider specific aspects of the peer’s writing” (p. 323). Third, is directed peer review in which a thorough review can be initiated by providing a checklist. Directed peer review is also useful for students who may have limited writing skills. When used in collaborative peer review, all reviewers will use the same criteria. Directed peer review was used in this doctoral study but is referred to as peer review.

The checklists and scoring rubrics used in any peer review, peer evaluation or self-evaluation activities support the process approach to writing by outlining the criteria or expectations for a particular piece of work. However, checklists differ from rubrics because they do not provide a measure of quality of performance (Depka, 2007). When the criteria are made available as measures against which assessment is made, learning becomes more focused and self-directed. The scoring guide or rubric which defines the assessment criteria improves the reliability and validity of marking, contributes to a more transparent and explicit assessment, and most importantly, actively engages students with the criteria (Elander et al., 2006). In fact, the rubric is capable of quantifying students’ performance objectively and students can

see the specific learning areas that need improvement and mastery (Arthur & McTighe, 2000; Groeber, 2007). Latucca (2005) argues that “this type of criterion-based grading approach alleviates student and instructor fears about the subjective nature of grading and banishes concerns about grading on a curve” (p. 249). Lattuca (2005) further explains that a rubric allows an instructor to “break a complex performance into discrete components that can be individually assessed against a standard” (p. 248). The established criteria and defined standards help to direct students to excel as the use of the rubric informs them of their strengths and weaknesses. This is obvious as the purpose of a rubric is to give students feedback about their progress as well as detailed evaluations of their final products (Andrade, 2000).

Using assessment criteria during peer evaluation, peer review and self-evaluation provides an avenue for deep learning. Assessing the quality of a learning performance against the criteria immerses students in the learning process. The assessment criteria guide and trigger students’ thinking as they make judgments about the quality of the learning performance. Entwistle (2001) explains that “the influence of assessment on deep learning is clear-cut. Assessment techniques that encourage students to think for themselves... shift students toward a deep approach. Assessment perceived by students as requiring no more than accurate reproduction of information lets students rely on a surface approach” (p. 16). Entwistle (2001) further elaborates that deep learning can be promoted through assessment by

- Focusing and understanding performance, using tasks to develop and demonstrate understanding and feedback to clarify and stress understanding;
- Using techniques to tap understanding, including more open-ended questions and less reliance on multiple-choice questions; and



- Grading in relation to levels of understanding, using qualitative criteria to boost validity. (p. 16)

Bailin, Case, Coombs, and Daniels (1999) argue that “critical thinking is not promoted simply through the repetition of “skills” of thinking, but rather by developing the relevant knowledge, commitments and strategies and, above all, by coming to understand what criteria and standards are relevant” (p. 280). The use of a rubric as a scoring guide paves the way for the development of critical thinking. Andrade (2000) argues that, “thinking-centered rubrics seemed to help students to think more deeply” (p. 16). Particularly in writing, rubrics help to channel students’ focus onto critical thinking (Elander et al., 2006). “Importantly, the top level of the rubric communicates what exemplary work should look like and, as such, involves the student in constructive learning and self-evaluation” (Hafner & Hafner, 2003, p. 1510). According to Saddler and Andrade (2004), for academic writing instruction, the learning goal is “to help students develop the self-regulation skills needed to successfully manage the intricacies of the writing process” (p. 49). They urge that this can be achieved by the use of an instructional rubric.

Jonsson and Svingy (2007) reviewed studies on using scoring rubrics in performance assessments. They wanted to find out the extent to which using the scoring rubrics (1) increased consistency of marking; (2) facilitated valid judgments of complex competencies; and (3) promoted learning. Through an analysis of 75 relevant studies they concluded that: (1) using rubrics enabled reliable marking especially when they were analytic, topic-specific, and accompanied with exemplars and/or rater training; (2) rubrics in themselves did not help valid judgment of complex competencies unless a comprehensive framework of validity was used when validating the rubric; (3) rubrics supported learning and/or improved instruction as

“rubrics make expectations and criteria explicit, which also facilitates feedback and self-assessment” (p. 130).

Studies comparing the effect of using a rubric in peer and self-evaluation activities or the effect of using a rubric or a checklist to develop critical thinking skills in peer evaluation and peer review, respectively, have not yet been found. The majority of the studies available compared the reliability and validity of self- and peer evaluation with that of teacher evaluation (e.g. Chaves, Baker, Chaves, & Fisher, 2006; Lindblom-Ylänne, Pihlajamäki, & Kotkas 2006; Şahin, 2007; Mat Daud, Abu Kassim, & Mat Daud, 2011). A few other studies looked at the potential of self- and peer assessment to support students’ motivation and promote learning (e.g. Johnson & Winterbottom, 2011; Wenji & Shuyi, 2011; Wiley & Gardner, 2010).

This thesis sheds some light on the possibility of developing critical thinking skills using peer evaluation guided by a rubric in comparison to self-evaluation using the same rubric and peer review using a checklist.

#### **2.4.2.2 Peer Evaluation as an Approach to Develop Critical Thinking Skills in Tertiary Academic Writing**

The literature discussed above suggests that peer evaluation encourages active learning among students. Both the writer and the evaluator have roles to play and both are learning through interaction. Ammer (1998) explains that working with peers provides opportunities for a student to “(a) question the present quality of his or her draft, (b) seek out restatement or elaboration of something that was recently learned, (c) receive constructive correction for misuse of a structural aspect of writing, and (d) listen to redirection advice regarding a work

in progress without the stigmatism of failure that frequently accompanies such assistance directly from a teacher” (p. 268). The roles of student writers and peer evaluators in producing written products are compared in Figure 2.

Figure 2. Roles of Writer and Peer Evaluator in Producing Written Products (Ammer, 1998, p. 268)

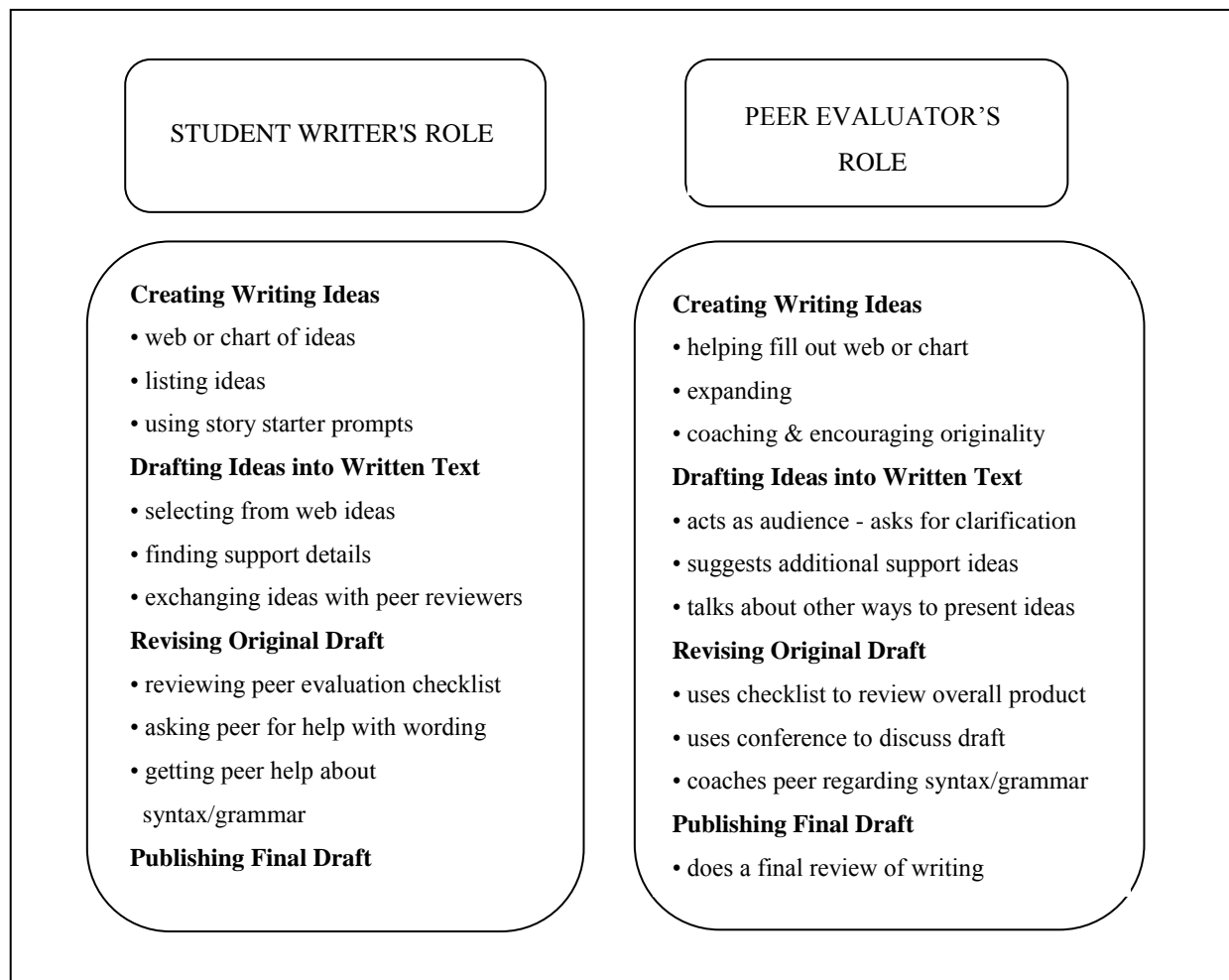


Figure 2 shows a student writer's role around the writing task beginning with brainstorming of ideas, drafting them and revising the drafts. Before getting the work published, a peer evaluator can help improve it by judging what is written at various stages of the writing process. Evaluating the peer's work allows students to utilise their content knowledge and exercise their evaluation and justification skills (Topping, 1998). For example, in the planning and drafting stages, peer evaluators are expected to provide their ideas on how to

improve the work. While in the revising stage, they are to evaluate the overall product and give comments to improve the paper before its publication. The exercise benefits the development of students' critical thinking skills. These roles of student writer and peer evaluator were adapted in the current study as the participants in the peer evaluation group were expected to individually write their essay and exchange it with their peers for evaluation before making improvement on their essay at each stage of writing.

The collaborative work above is possible through peer evaluation and peer review activities. Whatley, Bell, Shaylor, Zaitseva and Zakrzewska (2005) argue that “[c]ollaborative and cooperative learning approaches are examples of social learning where learners communicate with the tutor and other learners as they undertake tasks or projects in which learning and cognition can be situated” (p. 34). Both offer feedback from multiple readers including the teacher, which helps students be self-regulated learners (Nicol & MacFarlane-Dick, 2006). Peer evaluation has an added advantage over peer review in that it requires the evaluator to assess the quality of their peer's performance. The exercise of grading the work of a peer requires a careful analysis of the writing performance against a set of expected performance criteria.

Self-evaluation does not offer as great a potential to promote critical thinking skills in academic writing as peer evaluation. Although it is guided by a rubric, this activity does not allow input or interaction with peers. Hence, students rely only on the teacher to guide them to Vygotsky's Zone of Proximal Development (ZPD) and on the rubric to indicate important performance criteria. Peer evaluation, on the other hand, allows students to work with peers and receive feedback which is beneficial in helping them advance in their learning in addition to the guidance received from the teacher.

## **2.5 Key Issues in the Implementation of Peer Evaluation Activities to Promote Critical Thinking in Academic Writing**

Despite the apparent strengths of peer evaluation, implementing it has its own challenges. These include the validity and reliability of the grading, students' ability to evaluate, the grouping of students, the amount of time available, and engaging students in the activity. These are discussed in detail below.

### **2.5.1 Validity and Reliability of Grading**

Validity and reliability are important issues for peer evaluation (Brown et al., 1994; Bostock, 2001; Cho, Schunn, & Wilson, 2006). This is especially so when the grades are to be used for summative evaluation. Studies investigating validity, particularly criterion-related validity, frequently measure the agreement between student and teacher assessments (Topping, 1998; Falchikov & Golfinch, 2000; Cho et al., 2006). Falchikov and Golfinch (2000) and Cho et al. (2006) argue that the criterion-related validity of peer assessment can be measured by determining the level of agreement between peers' ratings and the teacher's ratings. That is, teachers' ratings are used as the criterion for determining the validity of peers' ratings. The reliability of peer assessment, on the other hand, can be measured by the similarity between the marks given by peers (as cited in Bouzidi & Jaillet, 2009).

Sahin's (2008) study, for example, showed very high validity of peer assessments. In this study, the validity of peer assessments of "Specific Teaching Methods I" course project by undergraduates of Gazi University, Turkey was studied. The Pearson correlation coefficient between the students' and teacher's scores of the course project was found to be very high ( $r=0.99$ ). In another study, Bouzidi and Jaillet (2009) examined the validity of online peer assessment among engineering students by studying the correlation between marks awarded

by peers with those of the teacher, and also between marks awarded by peers incorporating self-assessment with those of the teacher. The correlation coefficients between the teacher's marks and those of the peers for a course assessment done in two consecutive years were also found to be very high ( $r=0.90$ ) for "exams requesting simple calculations, some mathematical reasoning, short algorithms, and short texts referring to the exact science field (computer science and electrical engineering)" (p.257). Saito and Fujita (2004) investigated the validity of peer assessment of EFL writing by examining the similarity between peer, self- and teacher ratings. The Pearson correlation coefficient of students' and the teacher's scores was high and statistically significant  $r=0.72$  ( $p<0.01$ ) unlike very low and insignificant correlation coefficients between self-rating and teacher rating ( $r=0.07$ ). A meta-analytic study by Falchikov and Golfinch (2000) involving 48 quantitative studies comparing peer and teacher assessment showed close alignment of peer marks with teacher marks. Despite these studies, maintaining the validity of peer evaluation at an acceptable level is difficult (Dochy, Segers, & Sluijsmans, 1999). Some studies found low validity for peer evaluation when checked against teacher evaluations. A study of peer assessment validity in a study in high school setting by Chang, Tseng, Chou and Chen (2011), for example, revealed low validity ( $r$  values for seven assessment criteria ranged between 0.03 to 0.29). In another study by Ryan, Marshall, Porter and Jia (2007), peer evaluations of class participation, using a forced-normal distribution pattern, were found not predictive of faculty evaluations of class participation. Chen and Warren's (1999) and Mowl and Pain's (1995) studies of peer evaluation also showed low validity ( $r=0.29$  and  $r=0.22$ , respectively). Lack of practical experience in assessing tasks was found to be a possible explanation for the low validity in both studies. Mowl and Pain also indicated the low validity of peer evaluation in their study could be due to the subjective nature of the assessment task.

Many studies have found peer assessment to be reliable. A study by Haaga (1993), for example, investigated the reliability of peer assessments made by graduate students majoring in psychology when the identity of the assessor and writer were concealed from each other. A reliability check against a list of criteria used for assessing the students' term papers, using the Pearson product-moment correlation between pairs of students assessing common papers revealed modest reliability ( $r=0.55$ ). In a recent study, Marin-Garcia, Miralles and Marín (2008) examined the reliability of the peer evaluation of oral presentations. The results showed the reliability of peer evaluation based on the average scores of ratings across a list of nine assessment criteria developed together with the students, was high ( $r=0.90$ ) when there were more evaluators (more than 10 per presentation) compared with the estimated reliability of the marks given by one student marker which was modest ( $r=0.47$ ). Similar results were found in Xiao and Lucking's (2008) study involving undergraduate students enrolled in a foundations course of teacher education. The peers' ratings of drafts of an article produced by the students for inclusion in an online textbook called *Wikibook* were done for two rounds. An intra-class correlation analysis of the peer assessments revealed statistically significant coefficients for the first round assessment of three raters ( $r=0.62$ ,  $p<.005$ ) and for the second-round assessment of twenty raters ( $r=0.75$ ,  $p<.001$ ). Hafner and Hafner's (2003) 3-year study of peer-group rating involving 107 college biology students indicated moderate inter-rater reliability using Spearman's rank order correlation coefficient of approximately 0.40–0.50. Negative results, however, were found in other studies including by Chang et al. (2011). In that study, peer assessment of Web-based portfolios produced individually by 72 senior high school students was done. The inter-rater reliability investigation based on Kendall's coefficient of concordance revealed insignificant coefficients for all the six raters per group assessment. Results for the six groups ranged from 0.05 to 0.94. Low reliability of peer evaluation was also found in Gopinath's (1999) study involving MBA students. Analysis

using Kendall's coefficient of concordance on the consistency of peers' scoring on two groups of students assessments of their mid-term and final exams revealed weak consensus (i.e. Class 1- midterm=0.44, final=0.49; Class 2: midterm=0.25, final=0.24).

Studies on the validity and reliability of peer evaluation or assessment show a range of findings, some with high validity and reliability and some low. But the literature generally shows that peer assessment is relatively valid and has moderate reliability. Bias in marking (Newstead & Dennies, 1990), unfamiliarity with the assessment criteria (Mowl & Pain, 1995; Falchikov & Golfinch, 2000; Freeman, 1995) and the number of raters are some possible explanations for low validity and reliability of peer evaluation (Sung Sung, Chang, Chang, & Yu, 2010).

Some suggestions to improve the validity and reliability of peer evaluation therefore include: training and familiarising the students with the grading criteria (Mowl & Pain, 1995; Falchikov & Golfinch, 2000; Freeman, 1995; Campbell, Mothersbaugh, Brammer, & Taylor, 2001; Stanton, 1999), and increasing the number of evaluators (Sung et al., 2010). It was also suggested that over time, when students are given more opportunities to evaluate, it helps to increase the reliability of peer evaluation (Marin-Garcia & Miralles, 2008; Mat Daud et al., 2011a).

In this study, peer evaluation was used for formative evaluation purposes to provide feedback to students about the quality of their academic writing from their peers. The rubric used included assessment of how 'critical' the students had been in presenting their ideas. The main purpose was to trigger students' critical thinking when evaluating the work of their peer, and when understanding and negotiating their peers' evaluation of their own work. In



other words, its use in the study was more as a learning mechanism than an assessment mechanism. Despite this however, the reliability and validity of the peer evaluation using the *CAWAR* was investigated for two reasons. One reason was that a valid rubric would help to ensure that students would consider the important elements of what was expected of them when they were assessing, providing feedback, and considering the feedback given by peers. The second reason was, although the intended aim of the developing *CAWAR* was to help provoke critical thinking during peer evaluation, an acceptable level of reliability and validity would enable the *CAWAR* to also be used as an assessment tool.

### **2.5.2 Evaluation Ability**

Differing levels of skill among evaluators leads to issues of the accuracy and reliability of feedback given during peer evaluation. Liu and Carless (2006) affirm that “we cannot reliably assess something, if we do not know what we are trying to assess or by what means we come to a judgment” (p. 285). Evaluation ability affects students’ acceptance of the peer evaluation activity both as the assessor and assessee. Peer feedback may not be accurate and effective (Ferris, 2008) and, therefore, students might be misleading or misled by peers (Jarvis & Gibson, 1997). Feedback provided without the assessor having strong content knowledge, critical ability, and evaluation experience can be harmful to learning. This is further discussed below.

As mentioned earlier, two of the advantages of peer assessment are that it helps students to develop evaluation and justification skills, and to utilise content knowledge (Topping, 1998). However, low mastery of the skills and knowledge about how to evaluate academic writing can affect the quality of the peer evaluation. Falchikov (2001) comments that reflective criticism of a peer’s work is required before feedback can be supplied. Those being assessed

therefore, should be able to expect constructive comments from their peers. In some instances, students may resist peer evaluation when students believe that peers lack credibility for evaluating their learning performances. Instead, the teacher is regarded the most reliable and qualified to do the evaluation as revealed in studies by Sengupta (1998) and Cheng and Warren (2003). In both studies which were conducted in Hong Kong, the students had some resistance towards peer evaluation. In Sengupta's (1998) study in a secondary school writing class, students' responses to interviews indicated that the students believed they were not fit to evaluate. This was related to their perceived inability to correct grammatical errors which was shaped by their experience with error corrections. Similarly, Cheng and Warren's (2003) study involving undergraduates taking an English for Academic Purpose (EAP) course in the Hong Kong Polytechnic University revealed that the students involved in the study still doubted they would be able to peer assess effectively even after some training was provided.

One possible explanation for the findings above is that eastern cultures empower teachers in knowledge transmission and assessment. The students therefore, are deprived of the experience to peer evaluate. Peer evaluation is alien to students' educational experience, which has been passive and receptive due to the long-practised approach of knowledge and skills transmission rather than transformation (Hassan, Jamaludin, Sulaiman, & Baki, 2010). Asking the students to peer evaluate also means challenging long-held notions about the teacher as sole knowledge provider and assessor.

Notwithstanding the above demands of students to possess some level of content knowledge and critical and evaluative skills, Kagan and Kagan (2010) assert that the issue of 'the blind leading the blind' should not be a barrier to implementing collaborative work like peer

evaluation. They explain that this can be dealt with if sufficient input and guidance is available from the teacher. Thus, the presence of the teacher to facilitate the activity is crucial. This means that the teacher's role is not undermined by peer evaluation as the teacher is required to play an active role to ensure the effectiveness of a peer evaluation activity. The teacher is expected to facilitate and monitor the learning process (Hiltz & Benbunan-Fich, 1997; Ingleton, Doube, Rogers, & Noble, 2000). Webb (2009) states that teacher's role in collaborative learning includes "preparing students to collaborate, forming groups, structuring group work to guide or require students to engage in certain processes, and engaging in certain types of discourse with groups and the class" (p. 6).

### **2.5.3 Student Engagement**

Poor student engagement creates a threat to effective peer evaluation activities. Slavin (1995) asserts that group members' contributions are vital for group work. Smith and McGregor (1992) agree that effective learning requires students to be actively working with information, ideas or skills. Thus, a lack of student commitment to peer evaluation activities may affect their learning. Not only do students grade their peers' work on a less than thorough assessment of the work (Leki, 1990), quality of the feedback was also seen to be a problem (Acton, 1984).

As stated above, students' evaluation ability could be one factor. Students believe that assessing peers who are less capable is easier than assessing those who are more able (Falchikov, 2001). Therefore, it is common that students are reluctant to evaluate peers they know are more able than them. Better students might also feel less respectful and appreciative of evaluations from peers who are weaker than them.

‘Loss of face’ may also inhibit active engagement in peer evaluation. In a study by Miller and Ng (1994) on the peer assessment of oral language proficiency among Chinese tertiary level students, many of the students who participated in the study did not favour being assessed by their classmates. The students who were used to teacher evaluation regarded it as embarrassing to have peers comment on their work. They preferred not to let their performance be so vulnerable to their peers’ comments. These students also did not feel comfortable assessing others’ work. One student commented that the equal status assumed among the classmates was threatened when peer evaluation was practised.

Students’ reluctance to peer evaluate may also be due to the effort required of them. Effective peer evaluation requires dedication and hard work. Besides the grading task following a careful examination of the given work; negotiations of ideas, defending one’s work and asking for clarification all contribute to the load. Students who have limited experience with learning activities requiring critical thinking especially will feel this most. Halx and Reybold (2006) explain that “when student first begin to think critically, they often experience discomfort because critical thinking calls for students to reflect; set aside their established assumptions; and consider other, sometimes counter, perspectives” (p. 295). This may result in some degree of pressure which in turn leads to students disengaging or withdrawing from the peer evaluation activity.

Furthermore, students’ traditional attitudes to authority could influence the extent of students’ engagement in the peer evaluation activity (Sengupta, 1998). For some students, the teacher is the sole knowledge provider and is the person responsible for responding to their work and determining the quality of the work. When students are required to be active and be in control of their learning this creates a mismatch of the traditional practice and the current practice.

This conflict of practice has affected students' readiness to participate in peer evaluation activities especially in the initial stage they are introduced to it.

All the above are generally true among Malaysian students who have no or little experience peer evaluating which therefore makes them less confident and less comfortable with evaluating others and being evaluated by others.

#### **2.5.4 Grouping of Students**

When carrying out a classroom activity, one of the concerns is finding the most appropriate grouping of students that is able to maximise learning. In the literature, discussion on grouping of students revolves around the size of the group, the selection of group members and the duration of the group (e.g. Johnson, Johnson, & Holubec, 1994; Murdoch, & Wilson, 2004; Arends & Kilcher, 2010). Detailed discussion of each follows below. Although group formation for cooperative learning has been extensively discussed in the literature, it has not received much research attention (Potosky & Duck, 2007).

Determining the best group size is essential for effective learning. The decision is commonly influenced by the purpose and nature of the task (Murdoch & Wilson, 2004), students' previous experience of working in groups, the availability of resources and space, and the duration of time given for the activity (Johnson et al., 1994). Generally, the ideal size is believed to be four to six people per group as larger groups normally are not able to function well as students tend to disengage from the learning activities (Arends & Kilcher, 2010; Murdoch & Wilson, 2004). Johnson et al. (1994) add that, it is hard to monitor students' effort and contribution to the group activity when the group is too big. In contrast, Rau and Heyl (1990) argue that smaller groups have "less diversity; and may lack divergent thinking

styles and varied expertise that help to animate collective decision making” (as cited in Gokhale, 1995, p. 25). However, it is advisable that pairs are formed for students who do not have much experience working in groups. Laughlin, Hatch, Silver and Boh (2006) studied the effects of group size on solving letters-to-numbers problems among 760 students at the University of Illinois. The participants were randomly assigned to solve the problems either individually, or in groups of two, three, four or five people. The results indicated that all those working in groups of three, four and five performed significantly better than those working individually or those working in pairs. The groups of three, four, and five people did not differ from each other. Thus, the study concluded that groups of three are sufficient to solve intellectual problems but suggested further research needed to be done to determine the appropriate group size for other problem-solving tasks such as solving survival problems.

In terms of the selection of group members; age, gender, academic ability, interest and learning style are among the main characteristics considered. The questions are whether to form heterogeneous or homogeneous groups randomly or non-randomly selected by the teacher or by the students themselves. A careful selection of group members can help optimise the possible learning benefits that will be gained through collaborative learning. Arguments on the best choice are discussed below.

Many researchers believe that heterogeneous groups are likely to produce better academic and cooperative results than homogeneous groups (e.g. Johnson et al., 1994; Murdoch & Wilson, 2004; Arends & Kilcher, 2010). With the opportunity to work with a wider range of people, students are exposed to “more elaborative thinking, more frequent giving and receiving of explanations and greater perspective-taking in discussion material, all of which increase the depth of understanding, the quality of reasoning and the accuracy of long-term

retention” (Johnson et al., 1994, p. 26). This supports the proposition that working collaboratively enhances critical thinking development. One issue raised is whether the high achieving students benefit from heterogeneous groupings. Kagan and Kagan (2010) assert that working with lower achieving students enables the higher achieving ones to develop social and emotional skills which are more useful to securing a job than IQ or academic success.

Despite the claims that heterogeneous groupings produce better effects on learning than homogeneous groupings, there is research that suggests that heterogeneity is not the determining factor for effective learning (e.g. Cheng, Lam, & Chan, 2008; Moody & Gifford, 1990). Studies finding the opposite (i.e. homogeneous groupings are better than heterogeneous grouping) were also found. Adodo and Agbayewa (2011), for example, conducted a study comparing the effect of homogeneous and heterogeneous ability level grouping class teaching on students’ learning outcomes in Integrated Science. The research participants were 60 junior secondary school students who came from two schools. Two groups of 30 students (15 males and 15 females) each were formed. The study revealed that the homogeneous ability group performed better in the achievement test in integrated science (ATIS), science oriented attitudinal scale (SOAS) and science vocational interest inventory (SVII) than the heterogeneous ability group. Results from the survey questionnaire also showed that the students were in favour of working in homogeneous groups. The findings from this study corroborate the research-based information on timely topics (RBITT) magazine on ability grouping (2002) which reports that having students with similar ability in groups resulted in better learning gains than those with mixed abilities.

Besides deciding whether to opt for homogeneous or heterogeneous grouping, another aspect to consider is whether to form random or non-random groups selected by the teacher or students. Johnson et al. (1994) argue that teacher-selected random group assignment is the easiest, most effective and involves no preparation. Besides, students also see this selection as “fair” (Barkley, Cross, & Major, 2005). Kagan and Kagan (2010) support that it can be fun working with those non-immediate classmates. Wagaman (2008) adds that forming groups randomly helps avoid: students labeling each other especially the slow ones, troublesome students from being together, and cliquing among students. Random group assignment, however, is not recommended for long-term stable base teams (Johnson et al., 1994; Kagan & Kagan, 2010). Chapman, Meuter, Toy, and Wright (2006) also warn that “although random assignment to groups has some advantages, it leaves the process of group composition purely to chance, and groups may or may not come together well” (p. 560). Not only this, random group selection might not lead to heterogeneous group formation especially when heterogeneity is expected within the groups. Although it appears to be fair, Bacon, Stewart and Anderson (2001) argued that it is actually quite unfair and groups might not have good skill sets or diversity. Instead Johnson et al. (1994) suggest teachers use stratified random group assignment especially when a teacher seeks to have groups with similar characteristics work together to achieve certain learning objectives. This means students are stratified for certain characteristics important for the activity, for example, achievement level, personal interest and/or learning style before being assigned to groups randomly.

The non-random student grouping can be formed either by the students or the teacher. Kagan and Kagan (2010) argue that allowing students to select their own group members enables students to work comfortably with each other. However, Johnson et al. (1994) does not recommend student-selected groups especially due to the tendency for the creation of



homogeneous groups. This is especially critical for low-ability students who will be deprived of the opportunities to learn and be adversely affected motivationally due to peer, personal and teachers' expectations of poor performance (Farrar, Evans, & Kirk, 2003). Brown and Thomson (2000) also add that "[s]tudent-selected groups often have powerful social agendas that take up their time and attention and results in much "off-task" behavior" (p. 64). Kagan and Kagan (2010) agree with this as they argue that the shared interests beyond schoolwork can lead to off-task behavior. In fact, student-selected group assignment may cause students to remain with the same group members for other tasks (Graham et al., 2004).

Non-random group selection by the teacher is an alternative to student group selection. It, however, requires that the teacher has some knowledge of the students' characteristics. Having teachers decide group composition, "at-risk" students especially can be assigned to work with those who are better and supportive (Johnson et al., 1994). Besides, it is claimed that teacher-selected groups help enhance between-group homogeneity and within-group heterogeneity (Johnson, 1998; Kagan, 1994; Slavin, 1995). This is because fair competition between homogeneous groups and constructive interaction among heterogeneous group members with diverse perspectives and skills can be initiated.

Johnson et al. (1994) suggest that one way to deal with the threats of random group assignment whether by teacher or students, is to give the opportunity for students to list the names of those they would prefer to work with and teacher to pick one from the list and select the other members.

It is apparent that both heterogeneous and homogeneous groupings whether randomly or non-randomly selected by the teacher or by students have got their own potency and downsides.

Teacher familiarity with the students, time allowed for the activity and the nature of the task and the expected learning objectives are among the main factors that determine the choice for the selection.

The duration for groups working together can also vary. Johnson et al. (1998) argue that a group should stay together long enough for it to be effective. Basically, two factors determine the duration: the difficulty level of the assignment, and familiarity with working in groups (Jacobs, 2006; Kagan, 1994).

The difficulty level of a task is measured by the complexity of the process involved in dealing with the task and the clarity of the task goal (Waern, 1982). For an assignment demanding multiple tasks, for example, a longer period is expected. The depth of the particular topic also influences the duration of group. In addition, groups who do not understand clearly what they need to do will require a longer time to work on a given task.

Whether or not the group members have experienced working in groups may also influence the length the group members need to stay together. For those who have had no or very little experience working in groups, have “more reason to overcome difficulties they may have in working with certain group mates if they know their group will exist for weeks or months” (Jacobs, 2006, p. 35). However, those who are used to working in groups require shorter time to work comfortably with others.

The formation of the peer evaluation and peer review groups in this study was made by taking into account the most benefits the grouping would offer to students’ development of critical thinking skills. Non-random teacher-selected mixed ability groups of three were

initially sought. However, due to the instructors' unfamiliarity with the students' ability at the point of the formation of the groups, random student-selected groups of three or four were instead made especially when the groups confronted time constraint to choose their group members due to their late attendance for the peer evaluation and peer review activities. The threat of the formation of homogeneous groups which were not good for weaker students was therefore averted.

### **2.5.5 Time Factor**

Peer evaluation requires more time than self-evaluation and peer review. Not only is more time needed to grade peers' writing after carefully analysing it, but also to discuss the feedback. Romney (1996) argues that working collaboratively is slower than traditional methods of learning. When the teacher is the decision maker, there is not much room for arguments. However, when peer evaluation is adopted, students need to clarify, defend and suggest ideas. Despite this, she asserts that the discussion itself is worth holding. Gokhale (1995) explains that, "The peer support system makes it possible for the learner to internalize both external knowledge and critical thinking skills and to convert them into tools for intellectual functioning" (p. 28). During the discussion, students exchange ideas, are more sensitive to others' views, think deeper about the issue, thus making them more active in the learning process.

Second, additional time is crucial especially at the initial exposure to the activity. The reason being that to familiarise students with the new learning experience and to get those who are not familiar to work comfortably with each other takes time as discussed above.

Students also perceive peer evaluation activities as time consuming (Topping et al., 2000; Falchikov, 2001; Odom et al., 2009). This is particularly of concern when coverage of the syllabus is at the heart of the course. Liu and Carless (2006) corroborate that “when under pressure to cover a certain amount of content within a specific module, many lecturers may perceive peer feedback as an unwanted extra” (p. 286).

Albeit the greater time consumption, Knight and Steinbach (2011) argue that the benefit gained should be of paramount consideration. Stone (2001) points out that if we expect students to show their best thinking, sufficient time for them should be provided.

Having analysed the strengths and challenges of peer evaluation in the literature, this study therefore sought to throw some light into the possibility of developing critical thinking skills through the use of the Critical thinking for Academic Writing Analytical Rubric (CAWAR) designed for undergraduate students’ use in peer evaluation activities.

## **CHAPTER THREE: THE DEVELOPMENT OF THE CAWAR**

### **3.0 Introduction**

Chapter Three deals with the detailed description of the development of the *CAWAR*. A number of steps were gone through from (1) identifying appropriate elements of critical thinking from the literature and the LE 4000 course objectives to (2) peer review and (3) trialing of the instrument. The chapter begins with a brief discussion of the theoretical perspectives and context guiding its development. This is followed by an examination of its validity and reliability as part of the processes of developing the *CAWAR* are actually inter-related with establishing its validity and reliability.

The development of the *CAWAR* is explained in detail because there was no such rubric available for use in peer evaluation activities of tertiary academic writing. Hence it is a new territory for exploration. Besides, it provides readers with evidence of the rigour of its development for use with the undergraduate students. By doing this, a better understanding can be gained about what the *CAWAR* is and why it looks the way it does.

### **3.1 Theoretical Perspectives and Context for Guiding the Development of the CAWAR**

To develop critical thinking, classroom activities should allow team work to solve problems and should trigger metacognition (Ryder, 1994). This means students' active engagement in the learning activities is integral. Learning should therefore be learner-centred because "teacher-centered instruction does not teach students to think for themselves" (Temple, 2005, p. 16). When an instruction is teacher-centred, the teacher plays the dominant role in directing learning and less effort is normally generated by students. Since the students' role is more passive, less thinking is required of them. To be a critical thinker however, one cannot

simply watch others think critically but one should be an active participant in the learning process (van Gelder, 2005).

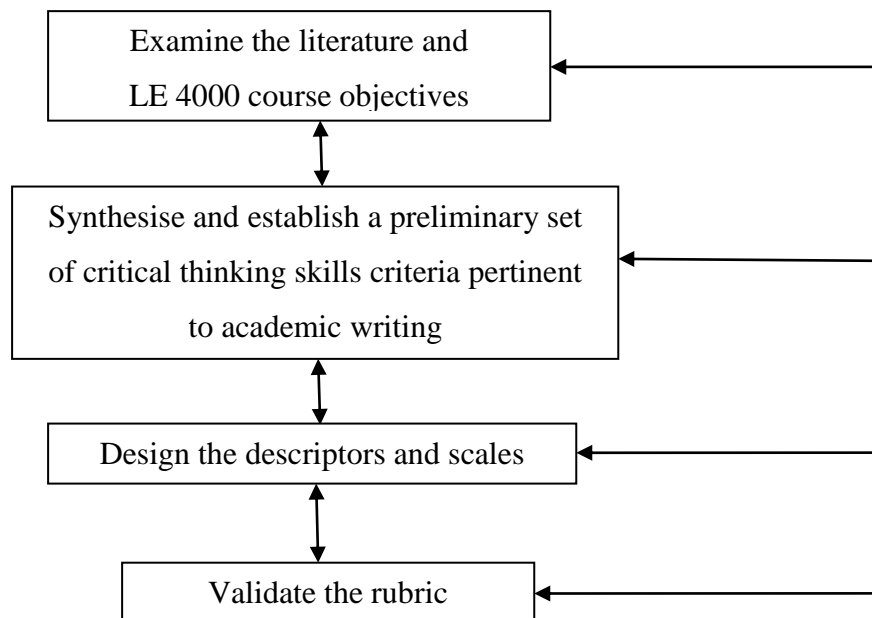
Peer evaluation engages students in a collaborative learning environment where students, either working in pairs or in groups, share the learning experience and are made responsible for each others' learning success. The exercise allows students "to practise the [metacognitive] skills needed for life-long learning (particularly, evaluation and critical thinking skills) by evaluating other students and observing how others evaluate the results of their learning" (Omelicheva, 2004, p. 2). To do this, the evaluation activity relies on a rubric as a guide to help more objectively quantify students' performance by providing more detailed feedback about the specific learning areas that need further attention (Groeber, 2007).

Writing provides a context for critical thinking development. Writing is open-ended where students' have more control over the length and content (Chatterji, 2003). Such a task enables elicitation of students' higher order thinking (Jonsson & Svingby, 2007). Academic writing in particular, offers an effective avenue for the promotion of critical thinking. This is due to its nature that requires critical analysis of sources before they can be coherently put together to produce new insights. Ryder (1994) elaborates that "critical writing is writing that displays thought involving such complex actions as drawing inferences, recognizing and creating relationships, and synthesizing large amounts of data to generate principles or global themes" (p. 211).

### 3.2 Development of the Critical Thinking for Academic Writing Analytical Rubric (CAWAR)

According to Tomlinson (2003), “In order to ensure that materials are devised, revised, selected and adapted in reliable and valid ways, we need to ensure that material evaluation establishes procedures which are thorough, rigorous, systematic and principled” (p. 5). The procedures involved in constructing and validating the CAWAR in this study were adapted from a compilation of various sources on designing and validating rubrics (Moskal, 2000; Wright, Burt, & Strongman, 2006; Allen & Knight, 2009). They are as follows in Figure 3.

Figure 3. The Development Process of the CAWAR



The following detailed account of the development of the rubric follows the pattern described in Figure 3 but the process was recursive in nature. The CAWAR was not developed following fixed steps. While working at one stage, the earlier phase/s were revisited in the process of refining the rubric.

### **3.2.1 Examining the Literature and LE 4000 Course Objectives**

In this study, the construct or attribute under study is unobservable. For observers to assess unobservable constructs like critical thinking, the constructs first need to be defined in operational terms. The “operational definitions specifically outline the actual responses, actions, tasks, or behaviors that will serve as observable evidence of a construct” (Chatterji, 2003, p. 9). This helps to establish construct validity of a rubric. To measure a construct not directly measureable, multiple items are often used (Gliem & Gliem, 2003). Thus, to identify the ‘grounded attributes’ of critical thinking for academic writing appropriate for academic writing criteria, a careful scrutiny of the literature relating to several cognitive models and taxonomies, and the *English for Academic Writing* course objectives was done.

#### **3.2.1.1 Critical Thinking and Cognitive Taxonomies**

It is generally agreed that human cognitive ability varies from one person to another. The differences are due to the quality of the individual’s mental interactions (Presseisen, 2001). Bloom’s Taxonomy of Educational Objectives of the cognitive domain (Bloom & Krathwohl, 1956) serves as one basis for understanding the order of thinking ability. The taxonomy, hierarchical in nature, suggests that one would perform the lower levels before they move to the higher levels. The six levels of thinking from lower to higher levels are: knowledge, comprehension, application, analysis, synthesis and evaluation. Among the six, the last four require critical thinking and are known as the higher-order thinking skills while the first two act as the base to the higher level ones (Bloom, 1994).

After its development in 1956, the Bloom’s taxonomy has been reinterpreted in different ways resulting in the construction of other taxonomies including Marzano’s model (1988) which expands the original taxonomy to eight (focusing, information gathering,



remembering, organising, analysing, generating, integrating and evaluating). Later, the North Carolina curriculum team (1994) reduced the first three levels of Marzano's model to a category called 'knowledge' (as the original first level in Bloom's taxonomy), re-included the level of applying which was dropped by Marzano from Bloom's model and retained the other levels in Marzano's model so that the North Carolina curriculum team's model read as: knowledge, organising, applying, analysing, generating, integrating, evaluating.

Another more recent taxonomy is the product of five years' work (1995-2000) labeled as Anderson and Krathwohl's taxonomy. Anderson and Krathwohl's taxonomy arose out of the collective work of a team that included cognitive psychologists, curriculum theorists and instructional researchers, and testing and assessment specialists. The taxonomy contrasts with Bloom's as it is two-dimensional involving cognitive processes and knowledge (Anderson, Krathwohl, & Airasia, 2001). The revised taxonomy is differentiated by not only the listings, rewordings (from nouns to verbs), renaming of some of the components, repositioning of the last two levels of the cognitive dimension (remembering, understanding, applying, analysing, evaluating and creating (formerly known as 'synthesis' in Bloom's taxonomy) but most importantly the way "the taxonomy intersects and acts upon different types and levels of knowledge – factual, conceptual, procedural and metacognitive" (Wilson, 2006, para. 5). The intersection between the two dimensions is represented in the Taxonomy Table (Anderson et al., 2001, inside front cover) as in Table 1:

Table 1. Anderson et al.'s (2001) Taxonomy Table

THE KNOWLEDGE DIMENSION	THE COGNITIVE PROCESS DIMENSION					
	1. REMEMBER	2. UNDERSTAND	3. APPLY	4. ANALYZE	5. EVALUATE	6. CREATE
A. FACTUAL KNOWLEDGE						
B. CONCEPTUAL KNOWLEDGE						
C. PROCEDURAL KNOWLEDGE						
D. METACOGNITIVE KNOWLEDGE						

Other than the aforementioned cognitive taxonomies, there are also critical thinking taxonomies, including one produced by Cambridge Assessment personnel and four critical thinking experts (Black, 2008). The taxonomy comprises five skills: analysis, evaluation, inference, synthesis/construction, and self-reflection and self-correction. Another well-accepted critical thinking taxonomy is one developed by Ennis (in Hager et al., 2003). Ennis has been refining his list of critical thinking abilities for decades based on the critiques of others and his own experience working with the abilities. Beginning with the earliest version of 1987, he improves the list to the most recent in 1991 which comprises four clusters (as cited in Hager et al., 2003, p.307-308):

The first five items involve *clarification*:-

1. Identify the focus: the issue, question, or conclusion
2. Analyse arguments
3. Ask and answer questions of clarification and/or challenge
4. Define terms, judge definitions, and deal with equivocation

5. Identify unstated assumptions

The next two involve the *basis* for the decision:-

6. Judge the credibility of a source

7. Observe, and judge observation reports

The next three involve *inference*:-

8. Deduce, and judge deductions

9. Induce, and judge inductions

a. to generalizations, and

b. to explanatory conclusions (including hypotheses).

10. Make and judge value judgments

The next two involve *supposition* and *integration*:-

11. Consider and reason from premises, reasons, assumptions, positions, and other propositions with which they disagree or about which they are in doubt

12. Integrate the other abilities and dispositions in making and defending a decision.

The Delphi Project of the American Philosophical Association also defined critical thinking as having two dimensions: cognitive skills and affective dispositions. The cognitive skills and sub-skills that the group included are (Facione, 1990a):

1. interpretation; sub-skills:

- categorization
- decoding significance
- clarifying meaning

## 2. analysis; sub-skills:

- examining ideas
- identifying arguments
- analysing arguments

## 3. evaluation; sub-skills:

- assessing claims
- assessing arguments

## 4. inference; sub-skills:

- querying evidence
- conjecturing alternatives
- drawing conclusions

## 5. explanation; sub-skills:

- stating results
- justifying procedures
- presenting arguments

## 6. self-regulation; sub-skills:

- self-examination
- self-correction

Despite the fact that the taxonomies described above are distinctive, there exist some similar abilities and similar patterns of positioning the different thinking levels. To develop the Critical Thinking for Academic Writing Analytical Rubric (CAWAR), three of these models were chosen. The choice was made due to their similarities and immediate relevance to academic writing. The models are compared in Table 2.

Table 2. The Three Models Compared

Anderson et al. (2001, p. 67-68)	Cambridge Assessment as cited in Black (2008, p. 9-10)	Facione (1998, p. 5-7)
<b>REMEMBER</b> - Recognizing (Identifying) - Recalling (Retrieving)		<b>INTERPRETATION</b> - recognizing a problem and describing it without bias - distinguishing a main idea from subordinate ideas in a text constructing a tentative categorization or way of organizing something you are studying - paraphrasing someone's ideas in your own words - clarifying what a sign, chart, or graph means - Identifying an author's purpose, theme, or point of view
<b>UNDERSTAND</b> - Interpreting (Clarifying, paraphrasing, representing, translating) - Exemplifying (Illustrating, instantiating) - Classifying (Categorizing, subsuming) - Summarizing (abstracting, generalizing) - Inferring (Concluding, extrapolating, interpolating, predicting) - Comparing (Contrasting, mapping, matching) - Explaining (Constructing models)		
<b>APPLY</b> - Executing (Carrying out) - Implementing (Using)		
<b>ANALYZE</b> - Differentiating (Discriminating, distinguishing, focusing, selecting) - Organizing (Finding coherence, integrating, outlining, parsing,	<b>ANALYSIS</b> -Recognising and using the basic terminology of reasoning - Recognising arguments and explanations - Recognising different	<b>ANALYSIS</b> - examining ideas - detecting arguments - analyzing arguments - identifying the similarities and differences between two approaches to the solution of a given problem

structuring) - Attributing (Deconstructing)	types of reasoning - Dissecting an argument - Categorising the component parts of an argument and identifying its structure. - Identifying unstated assumptions - Clarifying meaning	- picking out the main claim made in a newspaper editorial and tracing back the various reasons the editor offers in support of that claim - identifying unstated assumptions - constructing a way to represent a main conclusion and the various reasons given to support or criticise it - sketching the relationship of sentences or paragraphs to each other and to the main purpose of the passage - graphically organizing a chapter, knowing its purpose
<b>EVALUATE</b> -Checking (Coordinating, detecting, monitoring, testing) - Critiquing (Judging)	<b>EVALUATION</b> - Judging relevance - Judging sufficiency - Judging significance - Assessing credibility - Assessing plausibility - Assessing analogies - Detecting errors in reasoning - Assessing the soundness of reasoning within an argument - Considering the impact of further evidence upon an argument	<b>EVALUATION</b> - judging an author's or speaker's credibility - comparing the strengths and weaknesses of alternative interpretations determining the credibility of a source of information - judging if two statements contradict each other - judging if the evidence at hand supports the conclusion being drawn - recognizing the factors which make a person a credible witness regarding a given event or a credible authority with regard to a

		<p>given topic</p> <ul style="list-style-type: none"> <li>- judging if an argument's conclusion follows either with certainty or with a high level of confidence from its premises</li> <li>- judging the logical strength of arguments based on hypothetical situations</li> <li>- judging if a given argument is relevant or applicable or has implications for the situation at hand</li> </ul>
	<p><b>INFERENCE</b></p> <ul style="list-style-type: none"> <li>- Considering the implications of claims, points of view, principles, hypotheses and suppositions</li> <li>- Drawing appropriate conclusions</li> </ul>	<p><b>INFERENCE</b></p> <ul style="list-style-type: none"> <li>- querying evidence</li> <li>- conjecturing alternatives</li> <li>- drawing conclusions</li> <li>- seeing the implications of a position someone is advocating</li> <li>- drawing out or constructing meaning from the elements in a reading</li> <li>- identifying and securing the information needed to formulate a synthesis from multiple sources</li> <li>- after judging that it would be useful to resolving a given uncertainty if you knew certain facts, deciding on a plan which would yield clear knowledge regarding those facts</li> <li>- when faced with a problem, developing a set of options</li> </ul>

		for addressing it - conducting a controlled experiment scientifically and applying the proper statistical methods to attempt to confirm or disconfirm an empirical hypothesis
<b>CREATE</b> - Generating (Hypothesizing) - Planning (Designing) - Producing (Constructing)	<b>SYNTHESIS/CONSTRUCTION</b> - Selecting material relevant to an argument - Constructing a coherent and relevant argument or counter-argument - Taking arguments further - Forming well-reasoned judgments - Responding to dilemmas - Making and justifying rational decisions	<b>EXPLANATION</b> - stating results - justifying procedures - presenting arguments - constructing a chart which organises one's findings, - writing down for future reference your current thinking on some important and complex matter - citing the standards and contextual factors used to judge the quality of an interpretation of a text - stating research results and describing the methods and criteria used to achieve those results - appealing to established criteria as a way of showing the reasonableness of a given judgment - designing a graphic display which accurately represents the subordinate and super-ordinate relationship among concepts or ideas



		<ul style="list-style-type: none"> <li>- citing the evidence that led you to accept or reject an author's position on an issue</li> <li>- listing the factors that were considered in assigning a final course grade</li> </ul>
	<p><b>SELF-REFLECTION AND SELF-CORRECTION</b></p> <ul style="list-style-type: none"> <li>- Questioning one's own preconceptions</li> <li>- Careful and persistent evaluation of one's own reasoning.</li> </ul>	<p><b>SELF-REGULATION</b></p> <ul style="list-style-type: none"> <li>- monitoring and correcting an interpretation you have offered</li> <li>- examining and correcting an inference you have drawn</li> <li>- reviewing and reformulating one of your own explanations</li> <li>- examining and correcting your ability to examine and correct yourself</li> <li>- examining your views on a controversial issue with sensitivity to the possible influences on your personal biases or self-interest</li> <li>- monitoring how well you seem to be understanding or comprehending something</li> <li>- separating your personal opinions and assumptions from those of the author of a passage or text</li> <li>- double checking yourself by recalculating the figures</li> <li>- varying your reading speed and method according to the type of material and your purpose for reading</li> </ul>

		<ul style="list-style-type: none"> <li>- reconsidering your interpretation or judgment in view of further analysis of the facts of the case</li> <li>- revising your answers in view of the errors you discovered in your work</li> <li>- changing your conclusion in view of the realization that you had misjudged the importance of certain factors when coming to your earlier decision</li> </ul>
--	--	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

The sub-skills listed in the taxonomies were used to guide the match of cognitive skills among the three taxonomies. The ordering of the skills followed the hierarchy of thinking from the lower to higher levels. A direct match was made on two thinking levels of all the taxonomies specifically the *analysis* and *evaluation* skills. Cambridge Assessment (as cited in Black, 2008) and Facione (1998) also shared the *inference* and *self-reflection/self-regulation* skills but these two skills were absent in Anderson et al.'s (2001) cognitive taxonomy. Unlike the other two however, Anderson et al. (2001) included the *application* skill in their model which was a lower thinking skill than *analysis* but higher than the *interpretation* skill. Anderson et al.'s (2001) first two levels of thinking namely *remember* and *understand* were leveled with the *interpretation* skill in Facione's (1998) because the first two overlap with the *interpretation* skill. In other words, the *interpretation* skill requires the ability to recognise and understand a task or issue. Finally, Anderson et al.'s (2001) *create*, which refers to the same thinking process as Cambridge Assessment's (2008) *synthesis/construction* and Facione's (1998) *explanation* skill were put at the same level. The *explanation* skill was

leveled with *create* and *synthesis* as it involves the ability to justify and present one's own thinking based on well-reasoned judgments of others' arguments or ideas.

After some comparisons and contrasts among the critical thinking skills and sub-skills of the three models, the skills which were relevant to academic writing and were measurable were chosen for inclusion in the rubric assessment domain. At this point, the *interpretation*, *application*, *analysis*, *evaluation* and *synthesis* skills together with their sub-skills relevant to academic writing were selected.

To affirm their relevance and to gather the appropriate skill criteria, the *English for Academic Writing* course objectives and learning outcomes were analysed. Moskal and Leydens (2000) support the need to use the stated purpose and objectives as reference documents to guide a rubric development to establish the validity of a rubric.

### 3.2.1.2 English for Academic Writing (LE 4000) Course Objectives and Expected Learning Outcomes

The LE 4000 course objectives and expected learning outcomes are presented in Table 3:

Table 3. LE 4000 Course Objectives and Learning Outcomes

Course Objectives	Learning Outcomes
1. Enable the students to <i>synthesize</i> kulliyah* related information from academic and Islamic primary sources in 2500 word written arguments  * Kulliyah means faculty.	1. Compose a well organized and a well supported argumentative academic research paper
2. <i>Evaluate</i> kulliyah related information from academic and Islamic primary sources	2. Appraise views on kulliyah related topics and justify own stance

in 2500 word written arguments	
3. <i>Evaluate</i> appropriate techniques in citing academic and Islamic primary sources	3. Use relevant APA citation techniques in supporting academic research arguments
4. <i>Synthesize</i> kulliyah related information, language forms and language functions in presenting oral arguments	4. Construct an argumentative academic research paper using correct grammar and tone
5. <i>Evaluate</i> relevant academic and Islamic perspectives in relation to stance	5. Justify arguments with relevant academic and Islamic perspectives
6. Develop confidence in being responsible for the management of one's own continuous process of learning	
7. Critically appraise one's own understanding on the focused research topic	

From the list of course objectives, particularly objectives 1 to 5, it is apparent that *evaluation* and *synthesis* skills are emphasised. Objective 6 relates to *self-regulation* skill and objective 7 relates to knowledge of the research topic (i.e. content knowledge). Since they were intrapersonal skills unobservable to assessors, they could be excluded from being considered for inclusion in the CAWAR.

The learning outcomes, on the other hand, present the criteria of skill achievement expected to be developed by the students. These are among the specific skills relevant for incorporation in the rubric i.e. able to compose a well organised and a well supported paper, evaluate different views and justify one's own stance, use relevant citation techniques, correct grammar and tone, and justify arguments. Thus far, the preliminary set of sub-skills for each

of the five skills chosen based on the three models of cognitive taxonomies and the LE 4000 course objectives and learning outcomes were as follows:

1. Interpretation
  - a. identifying the problem/s
  - b. distinguishing the main idea from subordinate ideas
  - c. paraphrasing
  - d. identifying an author's purpose, theme or point of view
  - e. clarifying meaning
2. Application
  - a. using the correct citation techniques
  - b. using correct grammar
  - c. using correct argumentative tone
3. Analysis
  - a. classifying information, arguments, knowledge or perspectives
  - b. dissecting problem and arguments for critical evaluation
  - c. presenting arguments coherently
  - d. identifying unstated assumptions
4. Evaluation
  - a. judging relevance of arguments
  - b. assessing credibility of sources
  - c. judging sufficiency of arguments
  - d. judging significance of arguments
  - e. assessing the soundness of reasoning within an arguments
5. Synthesis
  - a. generating new insights from different perspectives

- b. drawing conclusions
- c. justifying own stance

Having identified the skills and the sub-skills, the study then concentrated on the different parts of the rubric.

### **3.2.2 Parts of the CAWAR**

Rubrics generally consist of three parts:

1. the criteria/dimensions which indicate the areas for assessment
2. the scales which indicate the level of performance
3. the benchmark descriptors that identify the standards of performance

At this stage the rubric also has an additional column for students to write comments to explain the grade they assign by setting out the specific strengths or weaknesses of the paper.

#### **3.2.2.1 Criteria/Dimensions**

Based on the researcher's conceptualisation of critical thinking in academic writing and after analysing, evaluating and synthesising the three models and the LE 4000 course objectives and learning outcomes, a preliminary list of five cognitive processes was generated to measure the critical thinking construct in academic writing. They are *interpretation*, *application*, *analysis*, *evaluation* and *synthesis*. The sub-skills, which were identified above, made up the operational definition of these skills in the context of academic writing. The sub-skills formed the criteria for evaluation in the CAWAR. The initial list however, was refined and then reduced in number after the face validity, construct validity and content validity were checked as discussed later in this chapter.

### 3.2.2.2 Scales

A six-point rubric labeled by a scale of *emerging*, *developing* and *mastering* as used in the Washington State University New Critical and Integrative Thinking Rubric (Center for Teaching, Learning & Technology at Washington State University, 2006) was preferred over a *poor* to *excellent* scale as they are more valid reflections of skills development “suggesting a continuum rather than a divide, providing a more educative and nuanced approach than a dualistic system can offer” (Center for Teaching, Learning & Technology at Washington State University, n.d., para 3). An *Immature* to *professional* scale is another alternative but the two ends, *immature* and *professional* are not appropriate terms to indicate skill development after some instruction. *Immature* carries with it a negative connotation of one’s capability while *professional* is an attainment level too high to reach after following a limited period of instruction. The word *emerging*, on the other hand, is more positive indicating initial progress in learning while *mastering* suggests progress approaching the targeted level of attainment in performance. Asmus (1999) reminds us that “the terms selected to describe the various performance levels should be chosen so that they do not have negative connotations if the purpose is to inform with an eye on future improvement” (para 7).

The rubric was specifically designed to also cater for ESL (English as a Second Language) or ESOL (English for Speakers of Other Languages) undergraduate students’ use. Thus the language complexity was targeted for this group. A check of the possible linguistic difficulty of the chosen scale was conducted with two undergraduate students one of whom was Egyptian and the other Pakistani. No comprehension difficulties were reported.

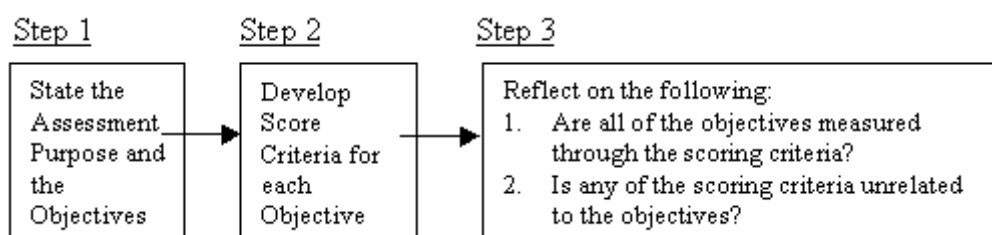
### 3.2.2.3 Benchmark Descriptors

The first draft of the *CAWAR* included benchmark descriptors which describe each level of performance represented by each scale of a rubric. However, after some consultation and considerations of the use of the *CAWAR* within group discussions, the descriptors were merged together in the criteria/ dimensions making it more economical and therefore, less tiring for students' use.

### 3.2.3 Establishing the Validity and the Reliability of the CAWAR

The establishment of the reliability and validity of the *CAWAR* followed the guide for establishing the validity and reliability of a scoring rubric provided by Moskal and Leydens (2000). To establish the validity of a rubric, what is hoped to be learnt about the responding students (i.e., the purpose) “and how the students will display these proficiencies (i.e., the objectives)” must first be determined and “[t]he teacher should use the stated purpose and objectives to guide the development of the scoring rubric” (Moskal & Leydens, 2000, para 12). The process of developing a rubric is therefore, as in Figure 4.

Figure 4. Evaluating the Appropriateness of Scoring Categories  
to a Stated Purpose (Moskal & Leydens, 2000)



The purpose of the construction of the *CAWAR* was to guide assessment of students' critical thinking in academic writing and the objective was to measure all the relevant critical thinking skills pertaining to academic writing. Having identified the assessment purpose and objectives, the score criteria for each objective were developed. This



was followed by a reflection of whether the scoring criteria provided the measurement of all the objectives and if there were any irrelevant criteria included. Checking the criteria in this way provides evidence to support the validity of the rubric, that is, whether it measures what it is intended to measure (Constantine & Ponterotto, 2006). Moskal & Leydens (2000) comment that

If the intention of an assessment instrument is to elicit evidence of an individual's knowledge within a given content area, such as historical facts, then the appropriateness of the content-related evidence should be considered. If the assessment instrument is designed to measure reasoning, problem solving or other processes that are internal to the individual and, therefore, require more indirect examination, then the appropriateness of the construct-related evidence should be examined. If the purpose of the assessment instrument is to elicit evidence of how a student will perform outside of school or in a different situation, criterion-related evidence should be considered. (para 14)

The three above (i.e. content-related, construct-related and criterion-related validity) are the most common types of validity of an assessment instrument (Brown, 2000; Moskal & Leydens, 2000; Onwuegbuzie, Jiao, & Bostick, 2004). Content validity is “the systematic examination of the test content to determine whether it covers a representative sample of the behaviour domain to be measured” (Anastasi & Urbina, 1997 p. 114). A construct is, “an attribute, proficiency, ability, or skill that happens in the human brain and is defined by established theories” (Brown, 2000, p. 9). Construct validity is “to assess the extent to which the test measures a theoretical construct or trait” (Groth-Marnat, 2009, p. 19). Criterion-related validity is observed when a test has successfully predicted the criterion or indicators of a construct (Cherry, 2011).

Moskal and Leydens (2000) provide questions to guide the examination of each type of validity evidence of a rubric in Table 4.

Table 4. Questions to Examine Each Type of Validity Evidence  
(Moskal & Leydens, 2000)

Content	Construct	Criterion
<ol style="list-style-type: none"> <li>1. Do the evaluation criteria address any extraneous content?</li> <li>2. Do the evaluation criteria of the scoring rubric address all aspects of the intended content?</li> <li>3. Is there any content addressed in the task that should be evaluated through the rubric, but is not?</li> </ol>	<ol style="list-style-type: none"> <li>1. Are all of the important facets of the intended construct evaluated through the scoring criteria?</li> <li>2. Is any of the evaluation criteria irrelevant to the construct of interest?</li> </ol>	<ol style="list-style-type: none"> <li>1. How do the scoring criteria reflect competencies that would suggest success on future or related performances?</li> <li>2. What are the important components of the future or related performance that may be evaluated through the use of the assessment instrument?</li> <li>3. How do the scoring criteria measure the important components of the future or related performance?</li> <li>4. Are there any facets of the future or related performance that are not reflected in the scoring criteria?</li> </ol>

For this study, all the three evidence types of validity were investigated. An assessment instrument is deemed content valid when it contains adequate samples of the content domain and the students' responses reflect the specific intended content of knowledge (Moskal & Leydens, 2000). Thus, for a writing evaluation instrument to have good content validity, it must be checked that "it evaluates writers' performance on the kind of writing task they are

normally required to do in the classroom” (Jacobs, Zinkgraf, Wormuth, Hartfield, & Hughey, 1981, p. 74). The content validity of the CAWAR was established through the Subject Matter Experts’ (SME) review and empirical trials as explained below in section 3.2.3.1 and 3.2.3.2 respectively.

The construct being assessed was critical thinking skills for academic writing. In order to unravel the construct of critical thinking skills in academic writing, the literature and the *English for Academic Writing* course objectives were referred to as explained above in section 3.2.1.

Criterion-related validity is observed when a test has successfully predicted the criteria of a construct (Cherry, 2011). Two types of criterion-related validity are concurrent and predictive validity. Concurrent validity refers to “the measurements taken at the same time, or approximately the same time as the test (Groth-Marnat, 2009, p. 17). “This indicates the extent to which the test scores accurately estimate an individual’s current state with regards to the criterion” (Cherry, 2011, para 6). Predictive validity, on the other hand, occurs when the criterion measures against which a test score is to be correlated with are obtained at a later time in the future (Cherry, 2011). In this study, concurrent validity rather than predictive validity was investigated. Predictive validity was impossible to check due to time constraints. The scores of the CAWAR were collected at the end of the semester hence, correlations between the scores on the CAWAR and some other criterion measure to be collected later could not be done.

To check for evidence of criterion-related (concurrent) validity of the CAWAR, *English for Academic Writing* (LE 4000) of IIUM instructor’s and peers’ ratings of the third draft of the

term paper for the course using the *CAWAR* were analysed. The instructor's and peers' ratings provide multiple indicators of the *CAWAR* criterion-related validity. Only 22 scripts out of 27 had complete scorings, thus used in the analysis. The third draft instead of the first or the second draft was used because it was scored using 11 out of 12 criteria of the *CAWAR* (except criterion 1, "Clearly states the thesis"), thus giving almost a complete coverage of the *CAWAR*. For the analysis, the instructor's scores and the average scores of the peers' rating were correlated with the term paper (rated using the criteria set by the Centre for Languages and Pre-University Academy Development (CELPAD) as listed in Chapter Four, section 4.2.4.3), International Islamic University Malaysia (IIUM) and the Cornell Critical Thinking Test Level X (CCTT-X) post-test scores as they were also two measures of critical thinking skills. Table 5 presents the results.

Table 5. Pearson's Correlation Coefficients for Instructor's Scores and Average Scores of Peers' Rating of Third Draft using the *CAWAR* and Term Paper and CCTT-X Post-test Scores

		Instructor	Students
Term	Pearson Correlation	0.42	0.13
	Sig. (2-tailed)	0.05	0.58
CCTT-X	Pearson Correlation	0.16	0.29
	Sig. (2-tailed)	0.48	0.19

The results indicate that the relationship between the instructor's scores using the *CAWAR* and their scores on the term paper using the rubric provided by the CELPAD, IIUM was close to significant, but no significant relationship was observed between the instructor's scores using the *CAWAR* and the CCTT-X ( $p=0.05$  and  $p=0.48$  respectively). In contrast, the relationships between the students' scores using the *CAWAR* and both the term paper and the CCTT-X were found not significant. This suggested two important points:

1. The instructor was more likely able to apply the criteria in the *CAWAR* more consistently than the students because they had done the assessment on the completed term paper, thus had a better interpretation of it compared to the students who only assessed parts of the term paper depending on the drafting stages. Due to this, it was expected that the correlation would be higher for the instructor's scores.
2. The *CAWAR* was more strongly related to the term paper than the CCTT-X. One reason was, given that the *CAWAR* was designed partly based on the LE 4000 course objectives, the correlation between the scoring done using the *CAWAR* and using the rubric set by the CELPAD, IIUM was than expected to be higher. Comparing and contrasting the two rubrics, it was found that the 11 criteria in the rubric set by the CELPAD, IIUM used to assess the term paper represented what the *CAWAR* was trying to assess. Second, the CCTT-X on the other hand, having a modest internal consistency (showing that measuring critical thinking itself is difficult) as revealed in the study and the test manual (discussed in section 4.2.4.1), might have some impact on this analysis (i.e. the weak evidence of criterion-related validity of the *CAWAR* via correlation with the CCTT-X scores). Besides, there was less overlap between the skills being assessed by the *CAWAR* and the skills being assessed by the CCTT-X and this contributed to the insignificant relationship between the two measures.

From this analysis, it was concluded that there was modest evidence of criterion-related validity of the *CAWAR* as a measure of critical thinking in academic writing. However, considering the fact that the *CAWAR* was not designed as a test instrument to generate scores on students' critical thinking skills in academic writing but rather to stimulate critical thinking as they assess the drafts of their academic papers during the self- and the peer

evaluation activities, the modest evidence of criterion-related validity was therefore did not impede its use for the mentioned purpose.

Face validity which is concerned with whether the *CAWAR* looks like it measures what it is intended to measure (Hughes, 2003) was also checked but it was not determined through formal procedures (Shuttleworth, 2009). The examination indicates that the *CAWAR* is face valid.

An assessment instrument is reliable when evidence shows that a test instrument provides consistent information (Moskal & Leydens, 2000). There are various ways of measuring reliability including test/retest reliability, equivalent-forms reliability, split half reliability and rational equivalence reliability. However, these “...are used to establish consistency of student performances within a given test or across more than one test” (Moskal & Leydens, 2000, para 19). When it has to do with the consistencies in assessment scores, the inter-rater reliability or teacher judgments rather than scores of the students are checked.

The validity and reliability of the *CAWAR* were rigorously established through subject matter expert (SME) reviews and empirical trialing. This included investigating whether (Moskal, 2000; Wright et al., 2006; Allen & Knight, 2009):

- a. The rubric items covered all the skills to measure the intended construct i.e. the critical thinking skills for academic writing;
- b. The levels were easy to distinguish between good and weak performance;
- c. There was an appropriate number of levels;
- d. The qualitative wordings used in the descriptors were appropriate and clear;
- e. High scores were consistent with good work and the low scores with poor work and;

- f. The rubric had acceptable inter-rater reliability.

These are explained in detail below.

### **3.2.3.1 Subject Matter Experts' (SME) Review of the CAWAR**

In developing instruments to assess a particular construct, Chatterji (2003) states that they are justifiable if they are based on “established, formal knowledge about the characteristic in question, and the consensus of opinion among experts about its occurrence in actual context” (p. 10). Chatterji (2003) continues that “knowledgeable experts help to validate content of an assessment by systematically reviewing and verifying the match between the assessment ‘operations’ (i.e. “the content of items or tasks, their structure and format, and the conditions under which the assessment is administered and scored) with its domain and the theoretical underpinnings” p. 61).

Subject matter experts (SME) were consulted to help check the relevance, clarity and conciseness of the CAWAR. The experts were selected based on their relevant training, experience and expertise in the field (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1985). Grant and Kinney (1992) suggest criteria in selecting content experts, which are: a history of publications in refereed journals, national presentations and research on the phenomenon of interest. Scheele (1975) suggested that experts must be selected from stakeholders who will be directly affected, experts with relevant experience, and facilitators in the field under study.

For this study, the validation procedure began with input from eight experts (three academic writing advisors from the University of Canterbury Learning Skills Centre, the Academic Development Group Coordinator, University of Canterbury who is an expert in rubric

development, the University of Canterbury *Writing the Academic Essay* (ENGL 117) course coordinator and three academic writing ‘practitioners’ i.e. two academic lecturers from College of Education, University of Canterbury and, an *English for Academic Writing* (LE 4000) course instructor who had almost ten years experience teaching the course and was also formerly the course coordinator from the International Islamic University, Malaysia representing the stakeholders.

Consultation with the individual experts and consequent revisions of the rubric took about three months to complete. Comments were invited from the experts independently and, based upon the subsequent feedback the researcher iteratively refined the rubric. The researcher also made the final decision regarding modifications of the rubric. The objective was to produce a valid and user-friendly rubric. Therefore, some alterations were made on the layout and the content of the rubric. Some parts were either retained, added, dropped, modified or reworded. Towards the end of the consultation period the experts’ feedback suggested that the rubric was constructed for what it was designed to assess and would be beneficial especially for undergraduate students’ use.

### **Drafts of the CAWAR**

This section describes the refinement process of the CAWAR involving three drafts before the fourth which was the final.

Initially there were two options of the first draft as shown in Tables 6 and 7. Both options listed five main skills; *interpretation*, *application*, *analysis*, *evaluation* and *synthesis* while 20 sub-skills made up the criteria. A commentary space was also provided for each sub-skill. The main differences between the options were that in option 1 there were detailed



benchmark descriptors to guide evaluation along the scale while option 2 indicated the benchmark of skill performance only at the best points. Secondly, option 1 had the performance scale incorporated in the table while in option 2 the scale was set in a band form each with a description of attainment level. The two approaches were designed to see which was more appropriate for use by undergraduate students within the approximately 1½ hour lesson.

Table 6. Draft 1 of the CAWAR - Option 1

Read the essay and rate each sub-skill by circling the appropriate mark.

Skills	Sub-skills	Criteria of Quality					
		Mastering		Developing		Emerging	
		6	5	4	3	2	1
Interpretation	1. Identifying problem	A very clear and accurate articulation of the problem.		Some attempt to identify the problem but vaguely expressed.		No attempt or failure to identify the issue.	
		Comments:					
	2. Distinguishing the main idea from subordinate ideas	Clearly distinguishes the main points from the subordinate ideas.		A vague distinction between the main idea and the subordinate ones.		No distinction between the main idea and the subordinate ones.	
		Comments:					
	3. Paraphrasing	An accurate use of own words to express others' ideas without changing the original meaning.		A fair use of own words to replace the original ones while meaning retains.		Most of the original words are retained with minor changes in the sentence structure or an attempt that greatly suffers the original meaning.	
		Comments:					
	4. Identifying an	A precise and accurate interpretation		Partially accurate interpretation of		Major misinterpretation of others'	

	author's purpose, theme, or point of view	<i>of others' point/s of view or arguments.</i>	<i>others' point/s of view or arguments.</i>	<i>point/s of view or arguments.</i>
		Comments:		
	5. Clarifying meaning	<i>A clear and accurate explanation of difficult ideas, concepts, arguments, facts etc.</i>	<i>A vague explanation of difficult ideas, concepts, arguments, facts etc.</i>	<i>No attempt to clarify of difficult ideas, concepts, arguments, facts etc.</i>
		Comments:		
<b>Application</b>	1. Using correct APA citation techniques	<i>An accurate application of APA citation.</i>	<i>Some errors in using APA citation techniques.</i>	<i>Serious errors in applying APA citation.</i>
		Comments:		
	2. Using correct grammar	<i>A consistent use of correct grammar.</i>	<i>Some grammatical errors.</i>	<i>A lot of major grammatical errors distracting comprehension of the essay</i>
		Comments:		
	3. Using correct argumentative	<i>A distinctive and consistent academic argumentative tone.</i>	<i>Lacks in academic argumentative tone.</i>	<i>Absence of academic argumentative tone.</i>
		Comments:		

<b>Analysis</b>	tone			
		Comments:		
	1. Classifying information, arguments, knowledge, or perspectives	<i>A clearly appropriate classification of information into meaningful categories e.g. evidence, examples, reasons, arguments or perspectives.</i>	<i>Some inappropriate classification of information into meaningful categories e.g. evidence, examples, reasons, arguments or perspectives</i>	<i>Failure to classify most information into meaningful categories.</i>
		Comments:		
	2. Dissecting problem and arguments for critical evaluation	<i>An evident break down of the problem or arguments into constituent parts for detailed analysis.</i>	<i>Lack in depth analysis of the constituent parts of the problem or arguments.</i>	<i>No attempt to dissect the problem.</i>
		Comments:		
	3. Presenting arguments coherently	<i>A clear sequencing of arguments across the text by relating sentences and also paragraphs to each other and to the main purpose of the writing task (using appropriate conjunctions and linking words) contributing to smooth flow of thought.</i>	<i>Evidence of attempts at structuring content through the grouping of ideas within and across sentences and paragraphs.</i>	<i>Generally very weak organisation of ideas at sentence or paragraph level.</i>

		Comments:		
	4. Identifying unstated assumptions	<i>An accurate identification of facts, beliefs, principles which are essential to the argument but have not been explicitly presented in the literature.</i>	<i>Partially accurate identification of facts, beliefs, principles which are essential to the argument but have not been explicitly presented in the literature.</i>	<i>Failure to identify facts, beliefs, principles which are essential to the argument but have not been explicitly presented in the literature.</i>
		Comments:		
<b>Evaluation</b>	1. Judging relevance of arguments	<i>An appropriate choice of arguments relevant to the issue.</i>	<i>Lack of relevant arguments.</i>	<i>Arguments are generally irrelevant to the issue.</i>
		Comments:		
	2. Assessing credibility of sources	<i>Accurate choices of trustworthy and reliable sources and rejection of those presenting inconsistent and false arguments.</i>	<i>Some sources lack credibility.</i>	<i>All or most sources are unreliable.</i>
		Comments:		

	3. Judging sufficiency of arguments	<i>Issue attended to with sufficient depth and breadth for the given length of writing.</i>	<i>Issue attended to with inadequate depth and breadth for the given length of writing</i>	<i>Issue superficially attended to.</i>
		Comments:		
	4. Judging significance of arguments	<i>A consistent reliance on only significant arguments when addressing the issue.</i>	<i>A mixture of significant and less significant arguments is included.</i>	<i>Arguments are generally insignificant.</i>
		Comments:		
	5. Assessing the soundness of reasoning within an argument	<i>A consistent use of strong, logical and solid points in reasoning inspiring confidence in the argument.</i>	<i>Arguments are moderately and not really confidently reasoned.</i>	<i>Arguments are very weakly or not confidently reasoned.</i>
		Comments:		
<b>Synthesis</b>	1. Generating new insights from different perspectives	<i>Accurate formulation of new ideas or insights by thoughtful and in-depth assessments of the relevant information from multiple sources.</i>	<i>Less accurate formulation of new ideas or insights by superficial assessments of the relevant information from multiple sources.</i>	<i>Very weak or no attempt to generate new ideas.</i>

		Comments:		
	2. Drawing conclusions	<i>A clear and concise formulation of a conclusion supported by appropriate and sufficient information or facts.</i>	<i>Draws a conclusion that is supported by insufficient and weak data.</i>	<i>Draws a conclusion that is not supported by data or fails to reach a conclusion.</i>
		Comments:		
	3. Justifying own stance	<i>Accurate and clear explanation of own stance on the issue with specific reasoned evidence.</i>	<i>Demonstrate stance but poorly reasoned.</i>	<i>No clear stance on the issue.</i>
		Comments:		

Table 7. Draft 1 of the CAWAR - Option 2

Read the essay and rate each criterion by writing a mark from 1-6.

**KEY:**

5-6 marks (Mastering) : Displays a well-developed ability to undertake the task

3-4 marks (Developing) : Demonstrates satisfactory attainment/mastery level

1-2 marks (Emerging) : Demonstrates a weak ability to do the task requiring a lot of attention and much practice

Skills	Criteria/Qualities	Score	Comments
<b>Interpretation</b>	1. A very clear and accurate articulation of the problem.		
	2. Clearly distinguishes the main points from the subordinate ideas.		
	3. An accurate use of own words to express others' ideas without changing the original meaning.		
	4. A precise and accurate interpretation of others' point/s of view or arguments.		
	5. A clear and accurate explanation of difficult ideas, concepts, arguments, facts etc.		



<b>Application</b>	1. An accurate application of APA citation.		
	2. A consistent use of correct grammar.		
	3. A distinctive academic argumentative tone.		
<b>Analysis</b>	1. A clearly appropriate classification of information into meaningful categories e.g. evidence, examples, reasons, arguments or perspectives.		
	2. An evident break down of the problem or arguments into constituent parts for detailed analysis.		
	3. A clear sequencing of arguments across the text by relating sentences and also paragraphs to each other and to the main purpose of the writing task (using appropriate conjunctions and linking words) contributing to smooth flow of thought.		

	4. An accurate identification of facts, beliefs, principles which are essential to the argument but have not been explicitly presented in the literature.		
<b>Evaluation</b>	1. An appropriate choice of arguments relevant to the issue.		
	2. Accurate choices of trustworthy and reliable sources and rejection of those presenting inconsistent and false arguments.		
	3. Issue attended to with sufficient depth and breadth for the given length of writing.		
	4. A Consistent reliance on significant arguments only when addressing the issue.		
	5. A Consistent use of strong, logical and solid points in reasoning inspiring confidence in the argument.		

<b>Synthesis</b>	1. Accurate formulation of new ideas or insights by thoughtful and in-depth assessments of the relevant information from multiple sources.		
	2. A clear and concise formulation of a conclusion supported by appropriate and sufficient information or facts.		
	3. Accurate and clear explanation of own stance on the issue with specific, reasoned evidence.		

The two options were first shown to three University of Canterbury Learning Skills Centre advisors for comments. A one-to-one consultation was held with each advisor. Based on the feedback gathered, the second option was preferred over the first for its simplicity as it was less wordy thus would not tire its users yet still have all the information together. The portrait page setup also made it more convenient to handle.

Other modifications made included:

1. Noun phrases were changed into verb phrases as the rubric was meant to evaluate actions.  
For instance, *A consistent use of correct grammar* became *Use correct grammar consistently*
2. Instead of providing space for comments on each skill, one commentary space was provided at the end of the rubric so that they could freely comment on any skill or skills at their own discretion

3. Instead of a specific benchmark descriptor for each criterion, a general descriptor for each point of the six-point scale under the labels of *emerging*, *developing* and *mastering* following a continuum of development was created to apply to all the criteria
4. A more careful choice of words was made to make the rubric more understandable and relevant for all types of academic writing (e.g. *Using correct argumentative **tone*** became *Use appropriate academic **register*** – an easier word to understand; *A very clear and accurate articulation of the **problem*** to *Clearly identifies the **task*** – the word ‘task’ was more applicable to all disciplines than ‘problem’; *An accurate application of **APA** citation* to *Cites the literature accurately* – other citation formats might be used)
5. The five main skills were deleted as they did not have obvious significance to users
6. The list of criteria covering all the five main skills were renumbered 1 to 20 consecutively after the deletion of the five main skills as below:
  - (1) A very clear and accurate articulation of the problem.
  - (2) Clearly distinguishes the main points from the subordinate ideas.
  - (3) An accurate use of own words to express others’ ideas without changing the original meaning.
  - (4) A precise and accurate interpretation of others’ point/s of view or arguments.
  - (5) A clear and accurate explanation of difficult ideas, concepts, arguments, facts etc.
  - (6) An accurate application of APA citation.
  - (7) A consistent use of correct grammar.
  - (8) A distinctive academic argumentative tone.
  - (9) A clearly appropriate classification of information into meaningful categories e.g. evidence, examples, reasons,

- arguments or perspectives.
- (10) An evident break down of the problem or arguments into constituent parts for detailed analysis.
  - (11) A clear sequencing of arguments across the text by relating sentences and also paragraphs to each other and to the main purpose of the writing task (using appropriate conjunctions and linking words) contributing to smooth flow of thought.
  - (12) An accurate identification of facts, beliefs, principles which are essential to the argument but have not been explicitly presented in the literature.
  - (13) An appropriate choice of arguments relevant to the issue.
  - (14) Accurate choices of trustworthy and reliable sources and rejection of those presenting inconsistent and false arguments.
  - (15) Issue attended to with sufficient depth and breadth for the given length of writing.
  - (16) A Consistent reliance on significant arguments only when addressing the issue.
  - (17) A Consistent use of strong, logical and solid points in reasoning inspiring confidence in the argument.
  - (18) Accurate formulation of new ideas or insights by thoughtful and in-depth assessments of the relevant information from multiple sources.
  - (19) A clear and concise formulation of a conclusion supported by appropriate and sufficient information or facts.
  - (20) Accurate and clear explanation of own stance on the issue with specific, reasoned evidence.

7. The 20 items were then reduced to 12 by:

- a. Combining overlapped items together (e.g. item 19 (*A clear and concise formulation of a conclusion supported by information or facts*) and item 20 (*Accurate and clear explanation of own stance on the issue with specific reasoned evidence*) were

- replaced with *Conclusion is strong with a clear stand taken on the issue*; item 3 (*An accurate use of own words to express others' ideas without changing the meaning*) and item 4 (*A precise and accurate interpretations of others' point/s of view or arguments*) were changed into *Paraphrases other people's ideas accurately*.
- b. Deleting items as they were already embedded in another item like item 2 (*Clearly distinguishes the main points from the subordinate ideas*) and item 9 (*A clearly appropriate classification of information into meaningful categories e.g. evidence, examples, reasons, arguments or perspectives*) which were expressed in item 10 (*An evident break down of the problem or arguments into constituent parts for detailed analysis*)
8. One criterion was added i.e. *Fulfils all the requirements of the task*
9. Long, complex sentences were shortened and made precise (e.g. item 11 (*A clear sequencing of arguments across the text by relating sentences and also paragraphs to each other and to the main purpose of the writing task (using appropriate conjunctions and linking words) contributing to smooth flow of thought*) was simplified to *Has organised ideas and/or information coherently*; item 14 (*Accurate choices of trustworthy and reliable sources and rejection of those presenting inconsistent and false arguments*) was replaced with *Cites the literature accurately*).
10. The rubric was compressed into a page.

The second draft as in Table 8 was produced as a result of the changes made. This refined draft was then taken to a lecturer at the College of Education, University of Canterbury for further advice.



After examining the draft, further suggestions were made and led to the following changes:

1. The instruction was rewritten to be more precise
2. The rubric was reduced to contain the two ends of performance domains i.e. the best and the weakest points to avoid wordiness and possible tiredness
3. The performance scale and labels were included in the rubric, thus deleting the earlier key to the rubric rating
4. The wordings and sentence constructions were further improved.

The following draft in Table 9 was then produced.

Table 9. Draft 3 of the CAWAR

### **The Critical Thinking in Academic Writing Analytical Rubric (CAWAR)**

**Writer:** \_\_\_\_\_ **Assessor:** \_\_\_\_\_

Please read the essay and then rate how well each skill is demonstrated in the essay. Use the 1-6 scale from the skill 'emerging' through 'developing' to the skill being 'mastered'. Circle one number on the scale. You are welcome to use the available space to write any comments.

<div><div><i>Emerging</i></div><div><i>Developing</i></div><div><i>Mastering</i></div></div> <div></div>								
1.	Does not clearly identify the task	1	2	3	4	5	6	Clearly identifies the task
2.	Explains difficult terms, concepts, facts and/or ideas poorly	1	2	3	4	5	6	Explains difficult terms, concepts, facts or/and ideas clearly



3.	Doesn't break down the issue into parts for detailed analysis	1	2	3	4	5	6	Properly breaks down the issue into parts for detailed analysis
4.	Makes many grammatical, spelling and punctuation errors	1	2	3	4	5	6	Writing is free from grammatical, spelling and punctuation errors
5.	Has not used the appropriate academic register	1	2	3	4	5	6	Has used the appropriate academic register
6.	Does not cite the literature accurately	1	2	3	4	5	6	Cites the literature accurately
7.	Does not support arguments using relevant literature	1	2	3	4	5	6	Supports arguments well using relevant literature
8.	Does not use reliable literature	1	2	3	4	5	6	Uses only reliable literature
9.	Paraphrases other people's ideas poorly	1	2	3	4	5	6	Paraphrases other people's ideas accurately
10.	Has not organised ideas and/or information coherently	1	2	3	4	5	6	Has organised ideas and/or information coherently
11.	Conclusion is weak with a vague stand taken on the issue	1	2	3	4	5	6	Conclusion is strong with a clear stand taken on the issue

12. Fulfil none of the requirements of the task	1	2	3	4	5	6	Fulfil all the requirements of the task
<b>Comments:</b>							

This draft was then shown to the Academic Development Group Coordinator and the *Writing the Academic Essay* (ENGL 117) course coordinator, University of Canterbury for further comments. Their suggestions which were accepted to help improve the rubric were:


1. To reorder the criteria of assessment in a way that they follow the flow of writing keeping the technical/ mechanical skills towards the end (e.g. item 6 went down to 11th, item 4 became the 12<sup>th</sup> and item 10 went up to 6<sup>th</sup>)
2. To delete criterion 12 (*Fulfil all the requirements of the task*) as it was subsumed under all other criteria listed in the rubric
3. To separate criterion 11 (*Conclusion is strong with a clear stand taken on the issue*) into two: *Concludes the essay strongly* and *Demonstrates a clear stand on the issue* as they were considered two different criteria that could be treated separately
4. For a repeated use of the rubric for writing in stages by the same user(s), the irrelevant items should be concealed or faded from view to help user(s) focus on the relevant criteria and avoid tiredness having to look at the unrelated ones.

The refined draft is presented in Table 10.

Table 10. Draft 4 of the CAWAR

**The Critical Thinking in Academic Writing Analytical Rubric (CAWAR)****Writer:****Assessor:****Section :****Draft :**

Please read the essay and then rate how well each skill except the shaded one (item no. 9) is demonstrated in the essay. Use the 1-6 scale from the skill ‘emerging’ through ‘developing’ to the skill being ‘mastered’. Circle one number on the scale. You are welcome to use the available space to write any comments.

<div><div><i>Emerging</i><i>Developing</i><i>Mastering</i></div><div></div></div>								
1.	Does not clearly identify the task	1	2	3	4	5	6	Clearly identifies the task
2.	Explains difficult terms, concepts, facts and/or ideas poorly	1	2	3	4	5	6	Explains difficult terms, concepts, facts or/and ideas clearly
3.	Doesn't break down the issue into parts for detailed analysis	1	2	3	4	5	6	Properly breaks down the issue into parts for detailed analysis
4.	Does not support arguments	1	2	3	4	5	6	Supports arguments well
5.	Does not use reliable literature	1	2	3	4	5	6	Uses only reliable literature



The CAWAR was finally checked by the instructor of *English for Academic Writing* (LE 4000) course instructor from the International Islamic University, Malaysia. She proposed the word ‘task’ in criterion 1 (*Clearly identifies the task*) change into ‘thesis’ and the word ‘register’ in criterion 10 (*Has used the appropriate academic writing register*) into ‘tone’.

In brief, with economy and effectiveness in mind, avoiding wordiness and possible tiredness, the rubric was reduced to contain 12 criteria with descriptions of the two ends of performance domains i.e. the best and the weakest points and a commentary space.

### 3.2.3.2 Empirical Trials

The CAWAR was then trialed with several groups. It was trialed with:

1. Four PhD students of University of Canterbury (three from the College of Education – a Chinese, an Iranian and a Sri Lankan, and one from the Department of Computer Science – a Malaysian).
2. A class of 15 students taking the *Writing the Academic Essay* (ENGL 117) course at the English Department, College of Arts, University of Canterbury. Most of the students were undergraduates and a few were postgraduates.
3. Four groups of social science undergraduate students doing *English for Academic Writing* (LE 4000) course at the International Islamic University Malaysia who participated in a quasi-experimental study.

The rubric was first tested on the PhD students to get some initial feedback on its clarity and usefulness before it was used with the actual targeted group, the undergraduates. The PhD

students were university lecturers at their home country with an average of 10 years of teaching experience. Two academic essays and two copies the *CAWAR* were distributed to the students. They read the essays and rated them using the *CAWAR*. The feedback gathered from the PhD students showed that they had no problem understanding the criteria listed and predicted that undergraduates would also not have problem with it. The postgraduates also rated the essays accordingly; one was better than the other.

Some constructive comments were given by the students. One suggested ‘depth of thinking’ and ‘scope of knowledge’ be included in the list of criteria. These however were not in the final list for three reasons. One, the existing 12 items of the rubric encapsulated ‘depth of thinking’ and ‘scope of knowledge’ with regard to academic writing. Second, one’s ‘depth of thinking’ and ‘scope of knowledge’ were not easily measured by looking at just one piece of work. Third, the terms were too broad to measure. Another student pointed out the possibility of modifying the rubric if it was to be used for evaluating an academic research paper. It was a point to consider but, since the rubric was meant for the general academic writing, the rubric was then retained. The other two students suggested that the rubric would be more useful and clear if students understood the concepts e.g. ‘thesis’, ‘reliable literature’ and ‘citation’. The researcher took note of this suggestion and later found them not problematic for the undergraduates taking the academic writing course as the terms were taught in academic writing courses.

The *CAWAR* was then distributed to a class of students taking the *Writing the Academic Essay* (ENGL 117) course at the English Department, College of Arts, University of Canterbury. They represented the types of participants who would be involved in the quasi-experimental study.

This group of students was also given the same two academic essays as the PhD students above and asked to rate them using the CAWAR.

Comments on the rubric were sought from these students. A lot of positive comments were received including the following written comments:

*The rubric is helpful because it gives a clear and quick way to mark an essay against the selected criteria, making it fair when marking a lot of essays.*

*Easy to use, good scale in the three parts (Emerging 1-2, developing 3-4, Mastering 5-6), easy to answer – clear questions.*

*The rubric has a good coverage of the important skills in academic writing.*

*The rubric way of marking is a great idea, and works well with the initial readings of an essay.*

Telephone conversations done randomly with two of the students to elicit more detailed views towards the CAWAR revealed that the CAWAR did not tire its user when evaluating as it was not wordy and the assessor did not have to read much when applying the rubric which helped him to stay focused. The rubric was also reported as well-formatted, containing well-ordered relevant criteria, using clear easy to understand wordings, and easy to rate scales (emerging to mastering).

A few comments suggested that some further considerations were necessary. The students wrote:

*The rubric gives little consideration to the style of writing.*

*Clear rubric – easy to use and understand! Could prove beneficial in essay writing but could easily be dismissed by the essay writer reading it.*

*The rubric helps to review particular areas of writing, but does not give examples of what constitutes a good or bad essay.*

The first comment led to an alteration of the wordings used in one of the criteria which addresses writing style but was not clear to the student. Thus, ‘academic writing style’ was used replacing ‘academic register’. The second comment reminded the researcher that the CAWAR is most suitable when used with discussions following the rating. This was not a problem for this study, as the rubric application in the study was meant to be followed by peer discussion to help generate more thinking among the participants. Having discussions following the grading task also would benefit both the assessor and the assessed. The final comment suggested that the assessor is required to have sound knowledge of the subject, writing styles and citation technique required of an academic paper. Some lessons on the academic writing skills and training using the CAWAR prior to using the rubric would thus be important.

After the refinement of the CAWAR draft 4 based on the comments made by the PhD students and the students taking the *Writing the Academic Essay* (ENGL 117) course, the CAWAR was checked for its inter-rater reliability. Two lecturers of the International Islamic University Malaysia (IIUM) rated a total of 23 English for Academic Writing (LE 4000) term papers. The results are as in Table 11 below.

Table 11. The Means and Standard Deviations of the Total Scores  
of Two Raters Using the CAWAR

	Mean	N	Std. Deviation
Rater 1	54.17	23	11.023
Rater 2	42.91	23	12.05



As seen above, Rater 1 was more lenient than Rater 2 when scoring using the *CAWAR*. To see if the difference between their scorings were statistically significant a paired sample t-test was then conducted. The result revealed a significant difference between the two raters ( $t=3.96$ ,  $p=0.001$ ). This suggested that one marker was significantly more lenient in their marking than the other. One possible reason for this could be due to the different length of teaching experience between the two raters. The more experienced rater who had a more severe scoring seemed to have more stringent views about critical thinking skills than the other rater.

The inter-rater reliability was then examined using a correlation coefficient. The result was  $r=.303$ , ( $p<.05$ ). This means that although one marker was consistently stricter than the other, the two raters rated the term papers in a similar order. Thus, the results showed that there was some evidence of moderate inter-rater reliability for the *CAWAR*.

At this point, the researcher was confident that the *CAWAR* possessed reasonable clarity, was easy to apply and could be interpreted reliably by students. The final version is attached as Appendix 16. Further study of its validity was provided from analyses of its use in the quasi-experimental study, described in Chapter Five and Six. Description of the study methodology is described in the next chapter (Chapter Four).

## CHAPTER FOUR: METHODOLOGY

### 4.0 Introduction

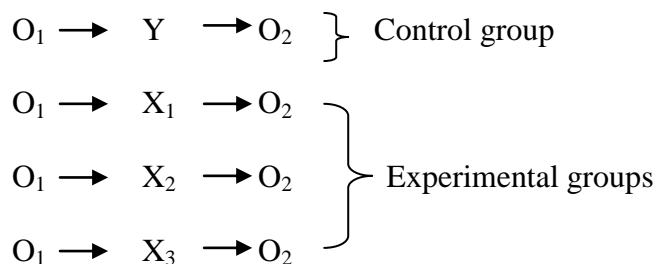
This chapter describes the quantitative and qualitative methods used in the study. The mixed-method approach was employed for several reasons: for methodological triangulation, to help validate the *CAWAR* and to gain a comprehensive understanding of the potential of peer evaluation through the use of the analytical rubric to foster critical thinking in academic writing. The detailed description of the research design, research procedure, data collection and data analyses is presented below.

### 4.1 Design of the Study

A quasi-experimental study was conducted to see whether the use of the rubric as a scoring guide for peer evaluation would help improve students' critical thinking skills.

The design used non-equivalent pre-post-test groups. This design is common in educational studies and was selected for its ability to control the extraneous variables that can threaten the internal validity of the study. It was also chosen due to its ability to justify post-test outcomes and demonstrate the extent of changes in students' critical thinking skills after the treatment. Mertens (1998) argues that in non-equivalent pre-post-test group design, the initial difference between the groups can be dealt with by the use of the pre-test. This helps to counter an argument such as by Trochim (2006) who worries that convenience sampling in quasi-experimental research would not make the groups as comparable as groups that are randomly assigned.

Four classes of students were each assigned to one treatment. In the design, the control group learned about academic writing the traditional way. This meant that students received writing instruction from the instructor and then produced essays to be checked and graded by their instructor. The control group allows the researcher to see if there is any difference in students' performance in academic writing without a treatment. The three experimental groups underwent different treatments in addition to writing instruction: peer evaluation and self-evaluation both guided by the *CAWAR*, and peer review guided by a checklist. As indicated in Chapter Two, directed peer review was used in the study because it was useful for students who might have limited writing skills, but by providing a checklist, a thorough review could be initiated. In the study, this group is referred to as the 'peer review' group. The one-page checklist that was developed based on the same criteria listed in the *CAWAR* for the peer review group's use is attached as Appendix 1. The 12 assessment criteria in the *CAWAR* were transferred into a list of 12 reviewing points for the students and underneath the list, a commentary space was provided for the students' use. No grading task was required. The list of 12 criteria was used by students as they reviewed their peers' work. The peer review group was included to allow comparisons with the peer evaluation group on the peers' critical thinking development without and with the *CAWAR* respectively. The self-evaluation group provided evidence as to whether critical thinking skills could be accelerated faster when the rubric was used for peer rather than self-evaluation. The detailed description of the treatments is provided at Section 4.2.6 below. The design of the study is as in the following notation:



$O_1$  – pre-test

$O_2$  – post-test

$Y$  – control group with no intervention/standard condition

$X_1$  – peer review with a checklist

$X_2$  – self-evaluation with rubric

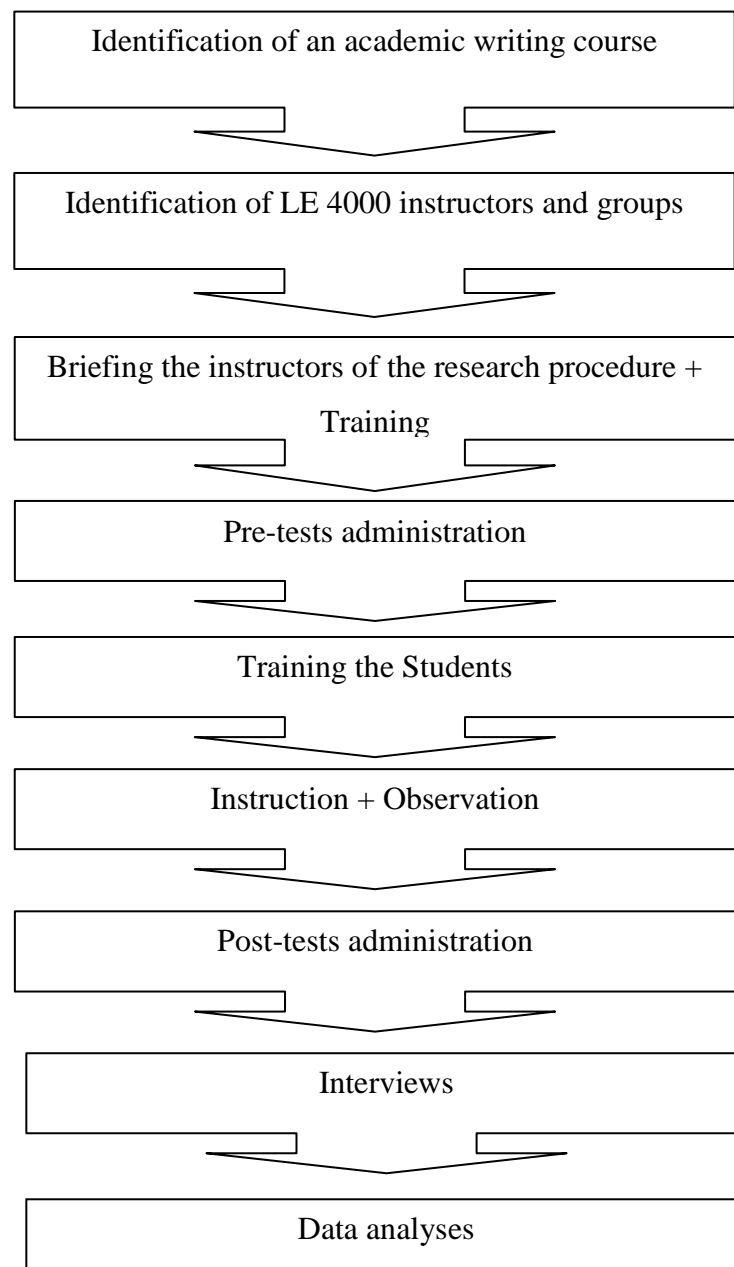
$X_3$  – peer evaluation with rubric

## 4.2 Research Procedures

To find out whether peer evaluation activities using the *CAWAR* would foster critical thinking, an academic writing course with critical thinking as one of the learning objectives was first identified. This was followed by the identification of two instructors and four homogenous groups to be the participants of the study. The identification of the research participants was done in the first half of the semester. Consent was sought from the participants by providing them with the information letters and consent forms as shown in appendices 7 to 14. Then, at the end of first half of the semester the instructors were briefed on the study including training them on how to adopt the rubric for peer and self-evaluation and the checklist for peer review. In the classrooms, the students were also trained on how to use the rubric or the checklist to help increase their confidence (Langan, Shuker, Cullen, Penney, Preziosi, & Wheeler, 2008).

The study was then started at the beginning of the second half of the semester with two pre-tests; Cornell Critical Thinking Test followed immediately by a survey questionnaire generally probing the extent to which the students perceived the peer evaluation, peer review and self-evaluation activities as important for assisting their learning. These pretests were immediately followed by the intervention during the drafting stage of the course term paper, two months before the end of the course. During the intervention, the researcher observed the activities carried out in all four classes to collect evidence on important aspects influencing the effectiveness of the learning activities. Towards the end of the semester, after the term paper final drafts had been submitted, the post-tests using the same instruments used in the pre-tests were administered. Interviews were then conducted with both the instructors and some selected students. Later, data analysis followed. The procedures are summarised in Figure 5. The detailed description of each step follows.

Figure 5: The Summary of Research Procedures



#### 4.2.1 English for Academic Writing course (LE 4000)

The *English for Academic Writing* course (LE 4000) guides students in writing an academic research paper. The course is offered by the Centre for Languages and Pre-University Academic

Development (CELPAD) that has the responsibility of providing and coordinating language classes to all faculties at the IIUM. Critical thinking is reflected in the course objectives to:

- a. Enable the students to synthesise kulliyah-related information from academic and Islamic primary sources in 2500 word written arguments;
- b. Evaluate kulliyah-related information from academic and Islamic primary sources in 2500 word written arguments;
- c. Evaluate appropriate techniques in citing academic and Islamic primary sources
- d. Synthesise kulliyah-related information, language forms and language functions in presenting oral arguments;
- e. Evaluate relevant academic and Islamic perspectives in relation to stance; and
- f. Develop confidence in being responsible for the management of one's own continuous process of learning, and critically appraise one's own understanding on the focused research topic.

According to the course coordinator (N. A. Abdul Ghani, personal communication, March 15, 2010), the semester-long course is offered in semester one and two with three contact hours per week. It is taught by about 35 to 38 instructors each semester. A great number of students register for the course each semester. In Gombak Campus (main campus) alone, approximately 1200 students take the course per semester. Of that number, 88% are final year students, 10% are third year students and 2% are second year students. The students make up 38 to 40 classes per semester. Some classes are homogeneous, according to kulliyah (faculty), but most classes are made up of students from different kulliyahs.

Instructions are delivered via tutorials (1½ hours per session, 2 sessions per week) and through a Learning Management System (LMS) which is a software that replaces mass lectures and is accessible to all the LE 4000 instructors and students for managing teaching and learning. Assessments of students' performance were based on an oral presentation of the term paper outline (10%), a mid-semester exam (20%), a term paper (30%) and a final exam (40%). Only the scores of the term paper were used for analyses and they were described in section 4.2.4.3.

#### **4.2.2 Participants**

The International Islamic University Malaysia (IIUM) was especially selected for the known standard of English proficiency among the students because the ability to communicate in the target language (English) has been identified as an important factor to the success of peer evaluation activities (Zaitseva, Bell, Whatley, & Shaylor, 2004). This was particularly sought since at the IIUM, students were expected to use English in class as English is the medium of instruction of the university and the course itself was an English course. At the university, prior to their enrolment for bachelor courses including the *English for Academic Writing* course (LE 4000), the students are required to have a minimum overall grade of 6 for IELTS (International English language Testing System) or its equivalent (Mat Daud, Mat Daud, & Md Zamin, 2011).

The participants of this study were two instructors and four classes of students of the *English for Academic Writing* course. To minimise the threat of non-comparability by extraneous variables, a careful selection was made of the two instructors and the four classes of students. Comparable gender, age, teaching experience and qualifications were the criteria sought for the instructors to minimise the influence of teacher factors on the results of the study. Having identified these



criteria, the Dean of the Centre for Languages and Pre-University Academic Development (CELPAD) was asked to nominate an instructor. Following the nomination, the instructor was approached via email and telephone and her consent was sought to participate in the study. She was then asked to suggest another instructor having a similar profile to be the second instructor. Table 12 below shows the detailed background of the instructors.

Table 12. Instructors' Background

<b>Instructor</b>	<b>Gender</b>	<b>Qualifications</b>	<b>Age</b>	<b>Past Teaching experience</b>	<b>LE 4000 Teaching experience</b>
1	Female	B.Ed TESL Masters of Management	40	High school teacher – 7 years	2001 - present
2	Female	B. Ed TESL M. Ed TESL	40	High school teacher – 6 years	2000 - present

The homogeneity of the groups was optimised by selecting four intact classes of students who were in approximately the same year of study and from the same discipline of study. With the advice of the Centre for Languages and Pre-University Academic Development (CELPAD), students in their third or fourth year of study in the social sciences were selected as participants.

The four intact classes contained 140 students. However, only 99 students did both the pre- and post-tests and were present for at least two out of the three intervention sessions were included in the study. The distribution of the four groups of participants according to faculty, nationality and gender is presented in Table 13.

Table 13. The Distribution of Participants

		GROUP				TOTAL
		CG	PR	SE	PE	(N=99)
		(N=24)	(N=24)	(N=24)	(N=27)	
FACULTY	Economics	20	0	0	18	38
	Human Science	0	0	24	0	24
	Law	0	24	0	0	24
	Accountancy	4	0	0	9	13
NATIONALITY	Malaysian	21	23	21	22	87
	Other*	3	1	3	5	12
GENDER	Male	4	1	4	10	19
	Female	20	23	20	17	80

Note:

CG = Control Group

PR = Peer Review

SE = Self-evaluation

PE = Peer Evaluation

\*International students from Montenegro, Brunei, Singapore, Indonesia, Maldives, India, Kenya and Yemen.

#### 4.2.3 Briefing and Training the Instructors

A week before the classroom activities started, two meetings were organised with the instructors. The first was done with both instructors present to check on the instructors' class schedules and weekly teaching plans and to set up and determine the intervention period. The second meeting was held with each instructor separately for more detailed instruction on the treatment (as explained in Section 3.3.6). Instructor 1 was assigned to two experimental groups: self-evaluation and peer evaluation. The reason was not to expose the rubric to the other instructor.

Instructor 2 was in charge of the control and peer review classes. The assignment of the treatments and control groups between the two groups per instructor was done randomly.

The drafting stage for the course term paper began at the start of the second half of the semester which was on the first week of February 2010. Students were expected to write four drafts:

1<sup>st</sup> – Introduction + 1<sup>st</sup> argument

2<sup>nd</sup> – 2<sup>nd</sup> argument + Counter-argument + Refutation

3<sup>rd</sup> – Conclusion, Abstract + Bibliography

4<sup>th</sup> (Final draft) – the whole paper

The intervention only involved the first three drafts. The final draft was the final paper that was submitted to the instructors for assessment.

The instructors were told that the criteria in the rubric or checklist would be highlighted progressively according to requirement of each draft to help students focus on the specific criteria to evaluate and review during the specific drafting stage yet made them aware of other criteria that would be evaluated and reviewed at the later stages. All the rated rubrics with the written comments and all the checklists with the reviewers' notes on them for the 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> drafts were submitted together with the final draft (the 4<sup>th</sup>) at the end of the semester.

For the peer review and peer evaluation groups, the instructors were reminded to ask the students to find the strengths in their peers' work before searching their weaknesses. This was considering the reminder made by Chisholm (1991) that "...most writers respond positively to positive feedback. When the first words the writer hears are words of genuine praise, they sound so

delicious that they make the writer's ear receptive to less positive comments that are sure to follow" (p. 14).

Cho, Schunn and Charney (2006) claim that "novices tend to just accept feedback when feedback givers have higher status, whereas they tend to challenge feedback from peers or less-knowledgeable people" (p. 285). To address the possibility of students' reluctance to accept feedback from their peers whom they regarded as less capable, Lu and Bol (2007) suggest that the identity of assessors be blinded. However, this was not possible in the present study as the students were expected to orally discuss each of the drafts they assessed. In fact, for the classes of students selected for the study, they had no or very small knowledge of each other's ability in academic writing for two reasons. First, the students had had only one assessment which was an oral presentation for the course prior to the intervention. That was not sufficient to inform them of the peers' capabilities. Second, some of the students also did not know their classmates well as they were from different departments and year of study. Thus, the problem of students hesitating to accept the feedback from peers was not a major issue in the study.

#### **4.2.4 Pre- and Post-test Instruments and Administration**

Two instruments were used in the pre- and post-tests (Cornell Critical Thinking Level X (CCTT-X) and questionnaire). A third instrument was a measure of student' academic writing which was the term paper collected at the end of the semester. Each is described below.

#### **4.2.4.1 Cornell Critical Thinking Test Level X (CCTT –X)**

Critical thinking was investigated using the Cornell Critical Thinking Test Level X (CCTT–X) developed by Robert Ennis and Jason Millman in 1985. The test was developed based on the earlier discussed conception of critical thinking (in Chapter Two) by one of its authors, Robert Ennis. The test was especially chosen for its wide recognition as a reliable test of critical thinking. It has reliability estimates with various populations which ranged from 0.67 to 0.90 (Ennis, Millman, & Tomko, 2005). It is also known for its objectivity in scoring (Rollins, 1990) and is suitable for use with any groups irrespective of specific disciplines (Royalty, 1995). The CCTT-X consists of 71 multiple-choice items (a 50-minute test) and is meant for 4<sup>th</sup> through to 14<sup>th</sup> graders but can be used with undergraduates who are less sophisticated (Ennis et al., 2005). Four skills are tested in CCTT-X which are:

- Deduction
- Induction
- Credibility of assertions
- Identification of assumptions

Although the skills tested in this test were not explicitly manifested in *CAWAR*, they were embedded in the assessment criteria. It was therefore used mainly to determine whether students' general critical thinking abilities had improved with the use of the rubric and was the dependent variable in the study.

The CCTT-X was administered in each class before the drafting stage began and again after the final draft of the term paper was submitted (as pre- and post-tests). The students were first asked to write down their names and other particulars on the answer sheets prepared by the researcher.

They were then briefed on the four sections of the test. Next, the students were asked to read the instructions on the back of the front cover of the test booklet to themselves. When they were finished and had no question to ask, the students were instructed to begin the test. They were given one hour to do the test.

Students' answers to the CCTT-X were hand-marked. At the end of the semester after the classroom interventions and the interviews were completed, both the pre- and post-test results were keyed into an Excel spreadsheet programme and revealed to the students via the Learning Management System (LMS). The results were disclosed by class and students were identified only by their student number.

To determine the internal consistency of the subscales and overall scale of the CCTT-X for the sample of 99 students, a reliability analysis was conducted using Cronbach's alpha (Santos, 1999). The results are displayed in Table 14.

Table 14. Reliability Coefficients (Cronbach's Alpha) for CCTT-X and Subscales

Scale	No. of Items	Test	Alpha Value
Inductive reasoning	23	Pre-test	0.53
		Post-test	0.54
Credibility of assertions	24	Pre-test	0.50
		Post-test	0.48
Deductive reasoning	14	Pre-test	0.61
		Post-test	0.67
Identifying assumptions in an argument	10	Pre-test	0.30
		Post-test	0.30
<b>Overall CCTT-X Cronbach Alpha Coefficients</b>	<b>71</b>	<b>Pre-test</b>	<b>0.66</b>
		<b>Post-test</b>	<b>0.70</b>

As seen in Table 14, the overall Cronbach's Alpha Coefficients of the whole test instrument (71 items) were 0.66 and 0.70 for the pre-test and post-test respectively. The coefficients are close to the lower end of the range of reliability coefficients published for the test (0.67 to 0.90). By reason of the marginal difference, the scores were taken as reasonably reliable. As for the subscales, the low range of the reliability coefficients (0.30 to 0.67) was expected. According to Ennis et al. (2005), it is due to "the overlap between sections, the moderate number of items in the parts [especially the last scale, *Assumption Identification*], and the probable heterogeneity of critical thinking" (p. 17). A study by Goldson (1990) on the relationship between cognitive style, critical thinking, and moral reasoning among 196 eighth and ninth graders for example, reported the reliability coefficients of the subscales ranged from 0.55 to 0.71. Another study by Mat Daud and Hussin (2004) on the possibility of developing critical thinking skills in computer-aided extended reading classes among undergraduates had the subscales alpha values ranging from 0.33 to 0.62. Because of the low range of the reliability coefficients for the different scales of the test and support from findings from earlier studies using the test instrument, it was therefore concluded that the subscales' reliability estimates for this study were within the reported range and thus the scores were reasonably reliable to use in the following analyses. However, since the alpha reliability coefficient was rather low for "ability to identify assumptions in argument skill" (0.30 for both pre-and post-test), the results of analysis for this subscale should be interpreted with caution. Thus, in the following analyses of students' performance of the CCTT-X subscales, any significant findings were treated with caution.

#### 4.2.4.2 Questionnaires

The questionnaire was divided into two parts as shown in Appendix 2. The first part consisted of 12 items related to the aspects of critical thinking relevant to writing listed in the *CAWAR*. It asked the students to rate themselves on the aspects along a four-point Likert scale from “poor” to “excellent”.

The second part of the questionnaire required the students to assess the extent to which they perceived peer evaluation, peer review and self-evaluation activities as being important for their learning. The students rated their perceptions of benefitting from the activities along a four-point Likert scale from “not at all” to a “large extent”.

To allow a pre-post-test comparison, the questionnaire was administered before the drafting stage began and again after the final draft of the course project was submitted. Before the students started answering the questionnaire, they were asked if they had any difficulties with the questions. After the students were clear on the instructions, they were given 10 minutes to mark their answers.

The same procedure was applied during the post-test. However, during the post-test, each student was also given an envelope containing the questionnaire they had answered earlier. The students were only allowed to open it after they had finished the post-test. In the envelope, students were provided with an instruction to compare their initial ratings with the later ones. Written explanations were sought on any differences between ratings made from the initial ones. Students in the peer evaluation and peer review groups were also asked whether they felt they



benefitted more when being evaluated/reviewed by peers or when evaluating/reviewing their peers' work.

#### **4.2.4.3 Term Paper**

Students were required to produce a term paper as part of the course performance assessment. The term paper carried a total of 30 marks and assessment was done according to criteria determined by the department which were:

1. abstract- (1 mark)
2. introduction - (3 marks)
3. argumentation (argument 1) - (2 marks)
4. argumentation (argument 2) - (2 marks)
5. argumentation (counter-argument) - (2 marks)
6. argumentation (refutation) - (2 marks)
7. conclusion - (2 marks)
8. bibliography – (2 marks)
9. technique - (3 marks)
10. format- - (1 mark) and
11. language – (10 marks)

The term papers for all four classes were marked by their own course instructor. To check the inter-rater reliability of the two instructors' marking, 10 papers were randomly selected from each class and graded by the other instructor independently. This gave 40 papers which both instructors had graded. The inter-rater reliability estimates using Cronbach's Alpha was then

done. The inter-rater reliability of  $r=0.61$  was statistically significant ( $p<0.01$ ). Thus, both instructors were grading the papers reasonably similarly.

Another analysis was done to further check on the inter-rater reliability. The average scores and the standard deviation of the 40 papers for the four groups as rated by the two instructors were calculated. The results are as in Table 15.

Table 15. Means and Std. Deviation of Both Raters Scores of 40 Term Papers

Group		Rater 1	Rater 2
CONTROL	Mean	23.10	22.55
	N	10	10
	Std.	2.48	2.76
	Deviation		
PEER REVIEW	Mean	23.75	23.25
	N	10	10
	Std.	1.21	2.53
	Deviation		
SELF-EVALUATION	Mean	23.80	20.25
	N	10	10
	Std.	1.70	2.57
	Deviation		
PEER EVALUATION	Mean	22.50	20.45
	N	10	10
	Std.	1.27	4.65
	Deviation		
Total	Mean	23.29	21.63
	N	40	40
	Std.	1.76	3.39
	Deviation		

The above results showed that basically rater 1 was more lenient than rater 2. She scored all the groups higher than rater 2. Rater 2, on the other hand, had a wider range of scores compared to Rater 1. Pearson Correlation Coefficients between the two raters across the four groups were then checked. The results are as in Table 16.

Table 16. Pearson's Correlation Coefficients for Rater 1 and Rater 2

		Rater 1	Rater 2
Rater 1	Pearson Correlation	1	0.54**
	Sig. (2-tailed)		0.00
	N	40	40
Rater 2	Pearson Correlation	0.54**	1
	Sig. (2-tailed)	0.00	
	N	40	40

\*\* $p < 0.01$  level

From the table, it was observed that the ratings made by the two raters were highly significantly correlated ( $p < 0.01$ ), thus supporting the inter-rater reliability of the term paper. It was therefore acceptable to collapse the term paper scores of all the students in the four groups in the subsequent analyses.

The term paper marks were used as a measure of the students' academic writing ability. It was also used as another measure of critical thinking. Although the set of criteria used to assess the term paper were realised to be different from the set of criteria of the *CAWAR* and the *CCTT-X*, they were all instruments for assessing critical thinking. The *CCTT-X* was used as a general measure of the critical thinking while the *CAWAR* and the term paper criteria reflected the critical thinking relevant to academic writing.

#### **4.2.5. Training the Students**

Prior to the evaluation and reviewing activities, the students in the experimental groups were trained on how to use the *CAWAR* or checklist. This was essential because the *CAWAR* and the checklist were not developed by the students themselves. Therefore, the initial exposure was deemed to be important to familiarise students with the listed criteria.

For this, copies of a term paper produced by one of the instructors' previous students were distributed to the classes during their *English for Academic Writing* period at the start of the second half of the semester. In their respective classes, students in the peer and self-evaluation groups were asked to read the paper and individually use the rubric to rate the work and write their comments. The peer review group was also asked to read the paper and later write their comments on the work based on the criteria listed in the provided checklist. After this was done, a whole class discussion was held to get students to share their opinions. Any uncertainties on how to use the rubric or checklist were dealt with during this period.

#### **4.2.6 Interventions**

The intervention sessions took place during the second half of the semester. The first half of the semester had been spent on lessons on writing an academic paper i.e. formulating a thesis statement, developing topic sentences, making claims etc. Students had also been asked to identify a topic for their term paper, visited the library and had been guided on how to find resources. They then had had to prepare an outline and present orally the self-chosen topic at the end of the first half of the semester.

During the second half of the semester, the control group did not receive any intervention. Students in the group prepared the drafts individually and the evaluation of the drafts was done solely by the instructor. The treatment conditions for the experimental groups however, were as follows:

#### **4.2.6.1 Self-Evaluation Group**

After preparing each draft, each individual student was given the *CAWAR* to self-evaluate their performance and identify their own strength/s and weakness/es. The rated *CAWAR* for each draft was kept by the writer for reference as they individually consulted the instructor for feedback and then make revisions. The instructor also evaluated each of the students' drafts using a separate sheet of the *CAWAR* which then enabled the students to compare their own ratings and the instructor's ratings.

#### **4.2.6.2 Peer Evaluation Group**

The students in the peer evaluation group were divided into groups of three or four. The students were expected to stay in the same group until the end of the peer evaluation sessions. However, latecomers or absences caused the group members to change. In the group, each member read each other's work and evaluated it using the *CAWAR*. In other words, each student's paper was evaluated by all of the group members. The reason was to provide the writer with multiple feedback for comparisons. Students wrote their feedback on the rubric. The allocation of time for this activity was approximately 10-15 minutes for each draft. Later the students were put in a discussion session (5-10 minutes for each paper) to share their comments based on the notes prepared earlier and for the writer to respond to the comments. The rubrics with comments on

them from all evaluators were then kept by the writer for reference as they improved the draft. The same procedure was applied for all the drafts. The instructor also rated and commented the students' work using a separate copy of the *CAWAR*.

#### **4.2.6.3 Peer Review Group**

The treatment for the peer review group followed the procedures undergone by the peer evaluation group only that instead of the *CAWAR*, a checklist was used to guide the reviewing activities but no grading was done by the students. The instructor also provided comments on individual students' drafts based on the same criteria listed in the checklist. No grade was given until the final draft was submitted.

In the peer evaluation and peer review activities both written and spoken feedback were required. Chisholm, (1991) comments on the advantages of having written and spoken peer review feedback which is also relevant to peer evaluation feedback. Asking students to write down their comments benefits both the reader and the writer. The reader has time to give more thought on the work while the writer is provided with a written account of the feedback. The spoken comments, on the other hand, "...with its flexibility and give and take, will more likely stimulate ideas. In addition, comments that may seem harsh or cold in writing may be made personal and warm when spoken" (p. 12).

The in-class activities including the pre- and post-tests and training took seven weeks. Evaluation and review of the drafts began on the third week of the second half of the semester. Schedule of the activities is presented in Table 17.

Table 17. Schedule of Activities

Week	Meeting	Activities	Action
A week before interventions	1	- Briefing and training the instructors	- Researcher and Instructors
1	1	- Pre-tests (CCTT-X, Questionnaire)	- Students
	2	- Lesson - Assignment of the 1 <sup>st</sup> draft task	- Instructors - Instructors
2	1	- Mid-term exam	- Students
	2	- Lesson - Training of students on rating using <i>CAWAR</i> and a short discussion	- Instructors - Researcher
3	1	- Evaluation/Review 1: First Draft (Introductory paragraph + First argument paragraph) - Classroom observation	- Students - Researcher
	2	- Lesson - Assignment of the 2 <sup>nd</sup> draft - Classroom observation	- Instructors - Instructors - Researcher
4	1	- Evaluation/Review 2: Second draft of Evaluation/Review 1 + the Second and Third argument paragraphs - Classroom observation	- Students - Researcher

	2	<ul style="list-style-type: none"> <li>- Lesson</li> <li>- Assignment of the 3<sup>rd</sup> draft</li> <li>- Classroom observation</li> </ul>	<ul style="list-style-type: none"> <li>- Instructors</li> <li>- Instructors</li> <li>- Researcher</li> </ul>
5	1	<ul style="list-style-type: none"> <li>- Evaluation/Review 3: Third draft of Evaluation/Review 1 + Second draft of Evaluation/Review 2 + Concluding paragraph</li> <li>- Classroom observation</li> </ul>	<ul style="list-style-type: none"> <li>- Students</li> <li>- Researcher</li> </ul>
	2	<ul style="list-style-type: none"> <li>- Lesson</li> </ul>	<ul style="list-style-type: none"> <li>- Instructors</li> </ul>
6	1	<ul style="list-style-type: none"> <li>- Submission of the final draft of the term paper</li> </ul>	<ul style="list-style-type: none"> <li>- Students</li> </ul>
	2	<ul style="list-style-type: none"> <li>- Revision</li> </ul>	<ul style="list-style-type: none"> <li>- Instructors</li> </ul>
7		<ul style="list-style-type: none"> <li>- Post-tests (CCTT-X, Questionnaire)</li> <li>- Interviews</li> </ul>	<ul style="list-style-type: none"> <li>- Students</li> <li>- Researcher and Students</li> </ul>

#### 4.2.7 Classroom Observations

Observations of the classroom activities were carried out with the permission of the instructors and the students of each group.

The ‘critical event’ observation technique (Wragg, 1999) was adopted by looking for particular classroom behaviours that could be indicative of something. According to Wragg (1999), critical events “are simply things that happen that seem to the observer to be of more interest than other events occurring at the same time, and therefore worth documenting in greater detail, usually



because they tell a small but significant part of a larger story” (p. 70). The observations were done throughout the drafting stages from the beginning until the end on all four classes. A written record of the observations was made for analysis. In particular, differences in the four classroom practices and activities especially the students’ participation and reactions to the different interventions, the running of the activities and the possible instructional inequities for different groups of students that might influence the learning outcomes were noted.

#### **4.2.8 Interviews**

Following the study, investigations were conducted on the perceptions of the students in the experimental groups and instructors about the extent to which the rubric helped them to achieve the learning outcomes (critical thinking skills). For this purpose, five students from each group were randomly selected from the course list. Every fifth person in the list were invited to engage in individual semi-structured face to face interviews to get a more detailed feedback on the adoption of the *CAWAR* and the different activities (peer evaluation, peer review and self-evaluation) in the *English for Academic Writing* course. The following person in the list was then invited for the interviews if the earlier identified student/s could not participate in the interviews. The interview questions were as in Appendix 3. Interviews were conducted for two weeks after the interventions. The time for the interviews was set by the students. They were carried out in a sound proof room to avoid distractions thus, providing a comfortable setting for the interviews. Each interview took 10 minutes to 20 minutes. The students were also allowed to use Malay if they preferred to pave the way for a better expression of views and ideas thus providing as much feedback as possible.

The two instructors were also interviewed separately at a private meeting room. In the 10 to 15-minute interview, the instructors' views on the potential of the different learning activities especially peer evaluation using the *CAWAR* were sought.

All the interviews were recorded using a digital voice recorder with the participants' permission to help the researcher focus on the interview. Only short written notes were taken to keep the researcher on track.

The recordings of the interviews were transcribed by a transcriber and later read through and checked for any spelling errors and missing words. To ease analyses, points were separated by dividing lines on the transcriptions. Ideas were then highlighted for easy reference later.

For writing the analyses, the original transcriptions which contained language and structural errors were rewritten into standard writing to ease comprehension. Malay transcriptions were also first given literal translation before being rewritten into the standard writing. For some transcriptions, pauses are retained by using the hyphen symbol (-). As shown in Appendix 4 the researcher has to mark off an interposed explanatory remark and inserted words. Parentheses ( ) and square brackets [ ] were used respectively.

#### **4.2.9 Data Analyses**

Both the quantitative and qualitative data analyses aligned with the three research questions. The analyses methods are described below.

#### **4.2.9.1 Analyses of Quantitative Data**

The data gathered from the study were analysed using the SPSS programme. Various statistical analyses were conducted on the pre- and post-tests data from the CCTT-X results and questionnaire, and the final marks of the term papers. The Cronbach's alpha reliability analysis was done on CCTT-X results to determine the internal consistency of the subscales and overall scale of the test. The descriptive analyses included analyses of performance distribution, and variability of the groups' performance using mean, variance and standard deviation. Inferential and multivariate statistics that were also used to examine differences in the students' performance and to identify relationships between variables included t-test, analysis of variance (ANOVA), analysis of covariance (ANCOVA), Pearson's correlation coefficient and Spearman's correlation coefficient.

For the statistical analyses, the actual significance levels are reported in the tables. When discussed in the text, the critical significance levels are referred to (i.e.  $p < 0.05$ ,  $p < 0.01$ ,  $p < 0.001$ ). Significance levels are reported to two decimal points except when probability is less than 0.001. Where this rounding results in the significance level becoming equal to one of the critical levels (e.g.  $p = 0.009$  becomes  $p = 0.01$ ), this is indicated in the table as  $< 0.01$ .

#### **4.2.9.2 Analyses of Qualitative Data**

A thematic analysis of the interview transcriptions and the written responses from the questionnaire was carried out to identify any commonalities, differences or/and relationships among the data pertaining to the interview questions (Gibson & Brown, 2009). In order to do

this, relevant details were sorted out and collated according to the themes derived from the patterns emerging from the data.

Identification of students' and instructors' responses to interviews and surveys were made using the following notation:

(Group/Instructor-Student/Instructor identification number, Interview/Survey)

Examples:

(PE-12, Survey)

(Instructor-1, Interview)

#### **4.3 Ethical Considerations**

Ethical approval to undertake this study was granted by the University of Canterbury Educational Research Human Ethics Committee. Consent was also sought from the Dean of CELPAD, instructors and students of the LE 4000. The information and consent letters were as in Appendices 5 to 14. Ethical requirement regarding confidentiality, anonymity, protection from discomfort was emphasised. The participants were provided with written and verbal explanations of the study.

## CHAPTER FIVE: RESEARCH FINDINGS AND INTERPRETATIONS: QUANTITATIVE ANALYSES

### 5.0 Introduction

This chapter presents and discusses the findings from the quantitative analyses of the data gathered from the quasi-experimental study on four groups of undergraduate students taking the *English for Academic Writing* (LE 4000) course at the International Islamic University Malaysia, Malaysia. The findings specifically helped to further verify the validity and usefulness of the Critical thinking for Academic Writing Analytical Rubric (CAWAR) by answering the three research questions which were:

1. Is there evidence of students developing greater critical thinking skills when they use the CAWAR in peer evaluation exercises to assess academic writing than in self-evaluation activities, or when using a checklist in peer review activities?
2. Do students develop their critical thinking skills better when they use the rubric to assess their peers, or when being assessed by their peers?
3. To what extent do teachers' and students perceive the peer evaluation activity using the CAWAR as capable of fostering students' critical thinking in academic writing?

Only the student perceptions are analysed in this chapter. The two instructors' perceptions of the potential of the rubric to foster critical thinking skills in peer evaluation activities for academic writing were investigated in the qualitative analyses.

In order to answer the above research questions, the following analyses were conducted. Analyses of the distribution of students' performance in the CCTT-X and possible effects of

extraneous variables were first done to ensure the data were fit to use. Then, analyses to address the first research question were carried out. These included the examination of students' performances in the CCTT-X and term paper followed by their perceptions of how much they had developed their critical thinking in academic writing. Next, the second research question was addressed by comparing the peer evaluation and peer review groups' perceptions of the extent to which they believed they benefitted from being an assessor or assessee. The relationship between these findings was then investigated in relation to the students' performances in the CCTT-X and the term paper. Later, the examination of the third research question was conducted. The students' perceptions towards the potential of the particular learning activity they experienced in promoting critical thinking was analysed. The results were then tested for any correlation with the students' performance in the CCTT-X. An additional analysis was carried out to find out the impact of the different intervention for each group on their academic writing ability through examining the correlation between their performance in the CCTT-X and their term paper.

### **5.1 Participants' Performance on the Cornell Critical Thinking Test**

The distribution of students' performance on the pre-test and post-test of the CCTT-X (71 items) for each group are displayed in Figures 6 and 7.

Figure 6. The Distribution of Scores  
on the CCTT-X Pre-test (N=71) by Group  
of Students

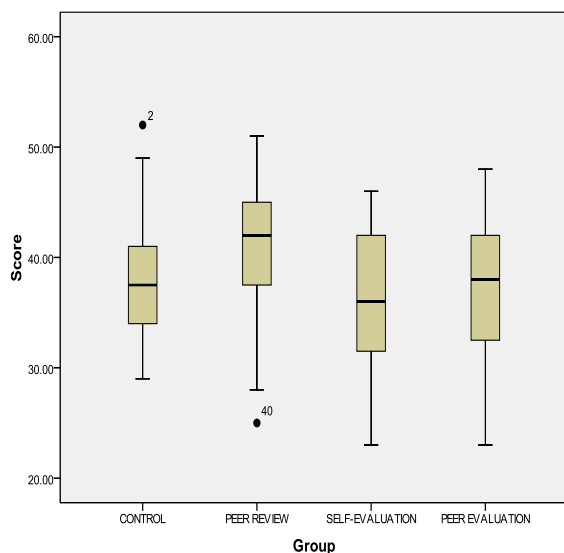
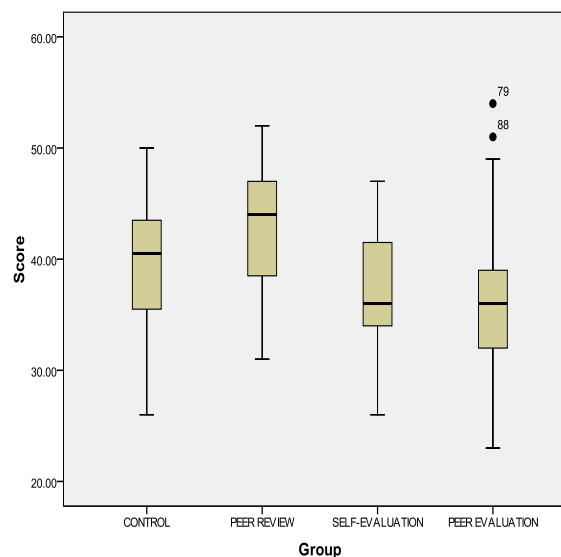


Figure 7. The Distribution of Scores  
on the CCTT-X Post-test (N=71) by Group  
of Students



The above figures indicate that the peer review group seemed to outperform the other groups in both the pre- and post-test ( pre-test  $\bar{x} = 40.58$  and post-test  $\bar{x} = 42.71$  compared to the control, self-evaluation and peer evaluation groups' pre-test  $\bar{x} = 8.33$ ,  $36.25$  and  $36.63$  and post-test  $\bar{x} = 39.96$ ,  $36.83$  and  $36.52$  respectively). This was evident from the group's inter-quartile range of the scores (the box) and median score (the line within the box) which were higher than the other groups. Post hoc analysis was done to see how significant the differences in the students' performances were and the results are displayed later in this chapter. It was also observed that the score distributions differed between the pre-test and the post-test. For the pre-test, the score distributions were almost symmetrical for the control group, positively skewed for the self-evaluation group but slightly negatively skewed for the peer review and peer evaluation groups. For the post-test however, the score distributions for the control and the peer review groups were negatively skewed in contrast with the self-evaluation group which was positively skewed. For the peer evaluation group, the distribution of scores was almost symmetrical. There were several

‘outside’ values. However, they were not considered to be extreme outliers that might affect the results. Therefore, they were included in the rest of the analysis.

The detailed distributions of the students’ performance in each of the four skills namely *inductive reasoning*, *credibility of assertions*, *deductive reasoning* and *identification of assumptions* in the Cornell Critical Thinking Test are as shown in Figures 8 to 15.

Figure 8. The Distribution of Scores on the CCTT-X *Inductive Reasoning* (N=23) Pre-test by Group of Students

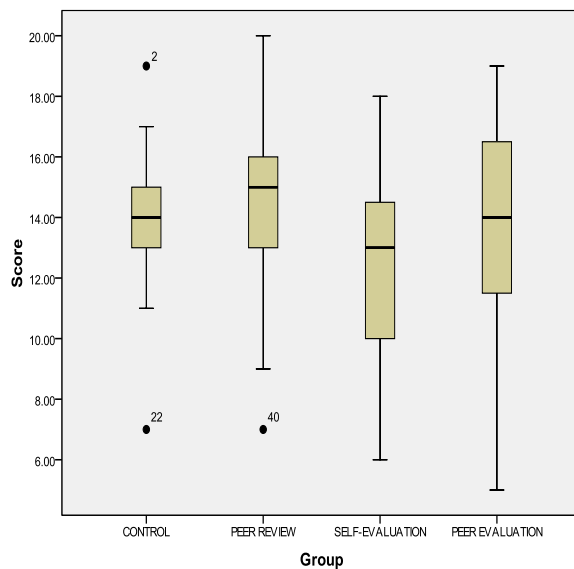


Figure 9. The Distribution of Scores on the CCTT-X *Inductive Reasoning* (N=23) Post-test by Group of Students

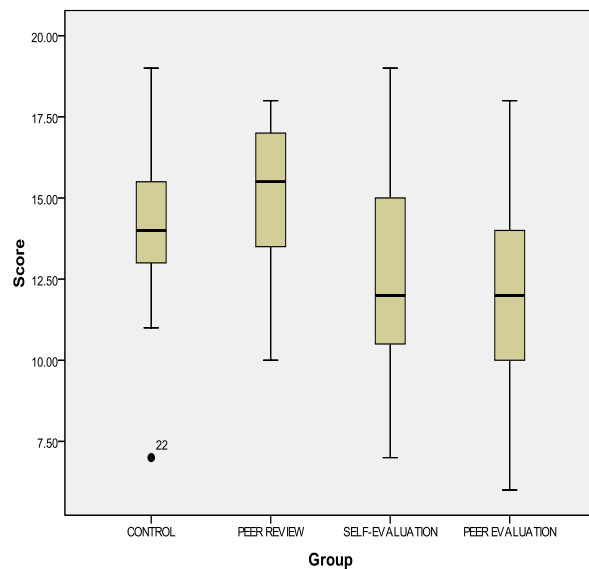




Figure 10. The Distribution of Scores on the CCTT-X *Credibility of Assertions* (N=24) Pre-test by Group of Students

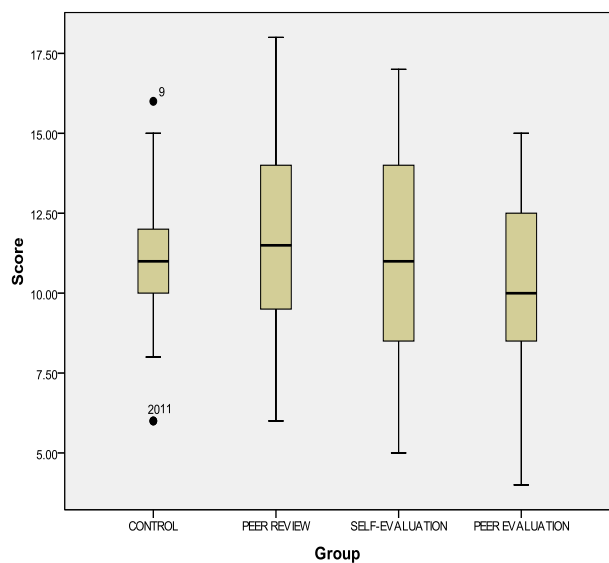


Figure 11. The Distribution of Scores on the CCTT-X *Credibility of Assertions* (N=24) Post-test by Group of Students

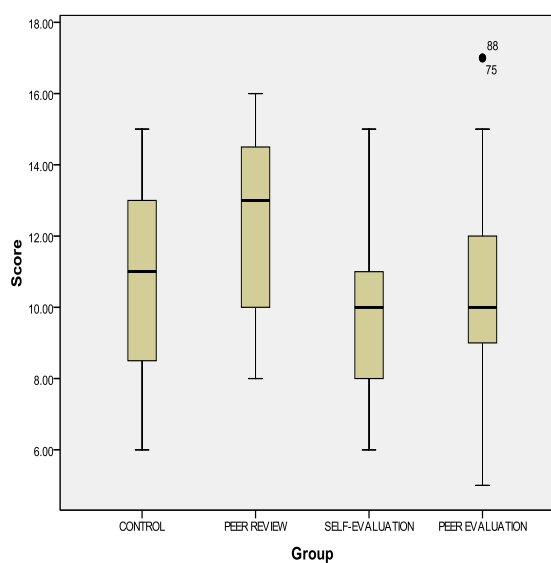


Figure 12. The Distribution of Scores on the CCTT-X *Deductive Reasoning* (N=14) Pre-test by Group of Students

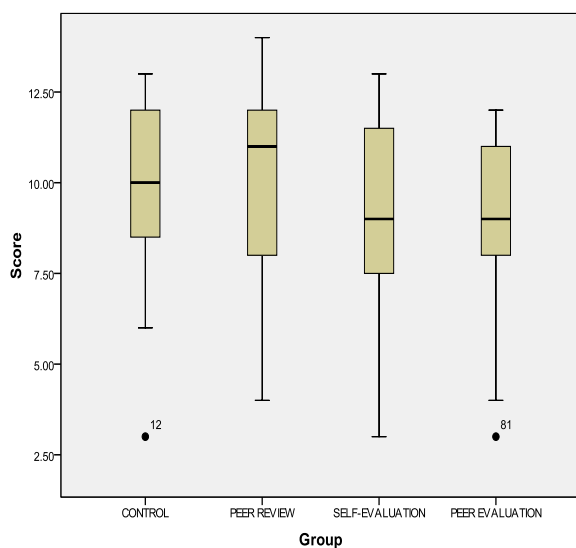


Figure 13. The Distribution of Scores on the CCTT-X *Deductive Reasoning* (N=14) Post-test by Group of Students

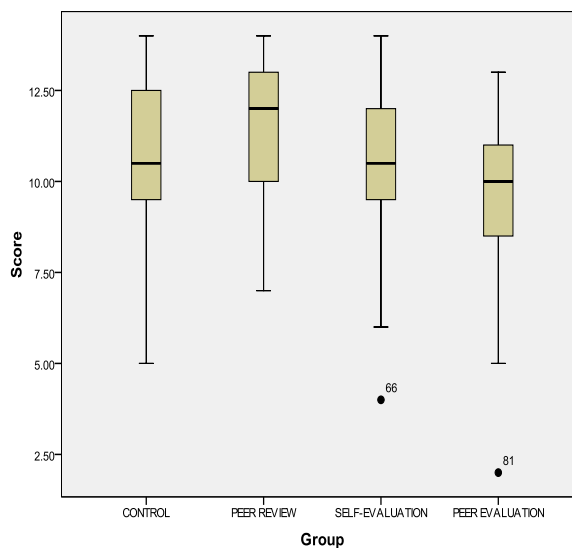


Figure 14. The Distribution of Scores on  
the CCTT-X *Identifying Assumptions* (N=10)  
Pre-test by Group of Students

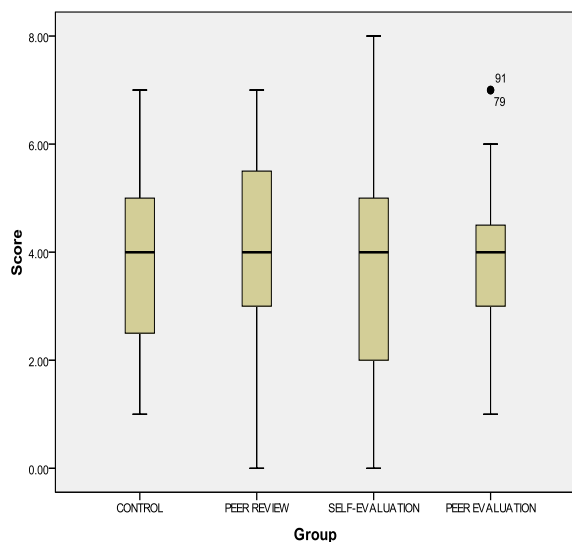
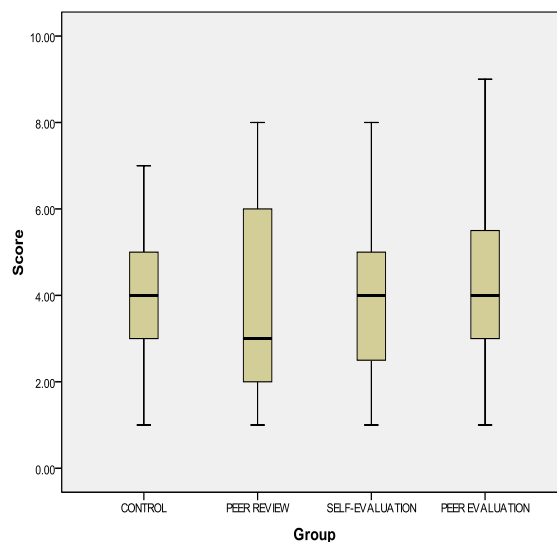


Figure 15. The Distribution of Scores on  
the CCTT-X *Identifying Assumptions* (N=10)  
Post-test by Group of Students



When the distributions of the scores for the individual skills for the pre- and post-test were compared and contrasted within and among the groups, it was apparent that the spreads of scores were more varied. To be specific, compared with the total pre- and post-test scores, there was more variation in the spread of the individual skill pre- and post-test minimum to maximum scores as well as the inter-quartile ranges for each group. For instance, for the *identification of assumptions* pre-test scores, the spread of scores by the self-evaluation group was noticeably the widest while for the same skill post-test scores, the peer review group clearly exhibited the biggest inter-quartile range compared to the other groups. In addition, the distributions of the scores were generally symmetrical. Only in certain instances they are slightly skewed. For example, for *inductive reasoning*, the self-evaluation group pre-test scores were marginally negatively skewed but they were slightly positively skewed in the post-test.

Although some skewness was present in the scores, it was not sufficient to warrant any transformation of scores when later analyses requiring assumptions of normal distributions were conducted (such as, ANCOVA). The box plots also indicated that except for the last skill (*identification of assumptions*), for the first three skills, the peer review group seemed to outperform the others. Analyses were conducted to see whether the differences were significant, and the results are presented later in this chapter. It can also be seen from the box plots that there were some outliers but none were extreme values that might affect the results.

## **5.2 Checking the Possible Effects of Extraneous Variables**

The quasi-experimental design of the study used four intact groups. The possible effects of having four faculties i.e. Law, Human Sciences, Accountancy and Economics, five levels of English proficiency i.e. Grade 1 to 5 on post-test scores and two genders were of interest. Investigations on the influence of the different faculties could not be performed on the peer review and self-evaluation groups due to students' coming from a mix of faculties. Only the control group and the peer evaluation group consisted of students from the same faculties i.e. Economics and Accountancy. ANCOVA with the CCTT-X post-test as the dependent variable was used to control for any undesirable pre-existing extraneous variables (Rutherford, 2001; Tabachnick & Fidell, 2001). The results of the analysis are as shown in Table 18.

Table 18. ANCOVA- Controlling for the Effects of Faculty and English proficiency Level on the CCTT-X Post-test Scores of the Control and Peer Evaluation Groups

Source	Sum of Squares	df	Mean Square	F	Sig.
Pre-test	426.12	1	426.12	13.29	0.00
Group	31.05	1	31.05	0.97	0.33
Faculty	5.55	1	5.55	0.17	0.68
English proficiency	43.16	4	10.79	0.34	0.85
Group * Faculty	3.92	1	3.92	0.12	0.73
Group * English proficiency	102.19	3	34.06	1.06	0.38

Based on the above table, it was found that for the two groups, faculty and English proficiency were not significant. There was no statistically significant interaction between faculty and English proficiency nor between the two treatments.

A separate analysis was also executed to see the possible effect of the English proficiency level on the CCTT-X post-test of all groups combined. The results are as in Table 19.

Table 19. ANCOVA- Controlling for the Effects of English Proficiency Level on the CCTT-X Post-test Scores of All Groups

Source	Sum of Squares	df	Mean Square	F	Sig.
Pre-test	966.12	1	966.12	38.35	0.00
Group	131.84	3	43.95	1.75	0.17
English proficiency	33.22	4	8.30	0.33	0.86
Group * English proficiency	196.95	9	21.89	0.87	0.56

Table 19 reveals that there was no significant interaction between the English proficiency and the treatment.

As for the effect of gender on the CCTT-X post-test of the four groups, the results are displayed in Table 20.

Table 20. ANCOVA- Controlling for the Effects of Gender  
on the CCTT-X Post-test Scores of All Groups

Source	Sum of Squares	df	Mean Square	F	Sig.
Pre-test	1446.04	1	1446.04	58.79	0.00
Group	133.54	3	44.51	1.81	0.15
Gender	6.00	1	6.00	0.24	0.62
Group * Gender	29.01	3	9.67	0.39	0.76

As seen in Table 20, gender did not have any significant effect on the CCTT-X post-test scores. No significant interaction between the gender and the treatment in the prediction of the post-test scores was observed for all the groups.

Since the three analyses above indicated that the differences were not important, later analyses with just the pre-test as the covariate were sufficient.

### **5.3 Developing Critical Thinking Skills Using Peer Evaluation vs. Self-Evaluation vs. Peer Review**

To answer the first question about whether students develop greater critical thinking skills when they use the *CAWAR* in peer evaluation exercises to assess academic writing than in self-evaluation activities, or when using a checklist in peer review activities, analyses on the CCTT-X and on the questionnaire were conducted.

### 5.3.1 Performance on the Cornell Critical Thinking Test

The mean scores of the four groups of students on the pre- and post-tests, and the mean of gain scores (post-test score minus pre-test score) were compared. See Table 21.

Table 21. CCTT-X Pre-test and Post-test Means and Standard Deviation of Scores by Group of Students

No. of Items	Group	N	Pre-test		Post-test		Gain (Post-test - Pre-test Scores)			
			Mean	SD	Mean	SD	Mean	SD	Min, Max	Range (Max-Min)
71	C	24	38.33	5.49	39.96	5.88	1.63	5.40	-7, 12	19
	PR	24	40.58	6.45	42.71	5.77	2.13	5.60	-6,17	23
	SE	24	36.25	6.26	36.83	5.55	0.58	4.75	-8,12	20
	PE	27	36.63	6.93	36.52	7.59	-0.11	5.89	-18,10	28

C = Control

PR = Peer Review

SE = Self-evaluation

PE = Peer Evaluation

As seen in Table 21, the pre-test mean scores indicated that the peer review group outperformed the other groups with  $\bar{x} = 40.58$  followed by the control group ( $\bar{x} = 38.33$ ). The peer evaluation and the self-evaluation groups had similar mean scores ( $\bar{x} = 36.63$  and  $36.25$  respectively). For the post-test, both the peer review and control group had the two highest means ( $\bar{x} = 42.71$  and  $39.96$  respectively). The self-evaluation group and peer evaluation group again had very similar means ( $\bar{x} = 36.83$  and  $36.52$  respectively). To see how different the post-test means were from the pre-test means scored by each group, the gain scores were calculated. It was evident from the result that the peer review and the control groups still had the two highest means ( $\bar{x} = 2.13$  and

1.63 respectively) followed by the self-evaluation group with  $\bar{x} = 0.58$ . The peer evaluation performance regressed with a mean  $\bar{x} = -0.11$ . Figure 16 depicts the overall performance of the four groups in the pre- and post-test.

Figure 16. CCTT-X Pre-test and Post-test Overall Means of Scores

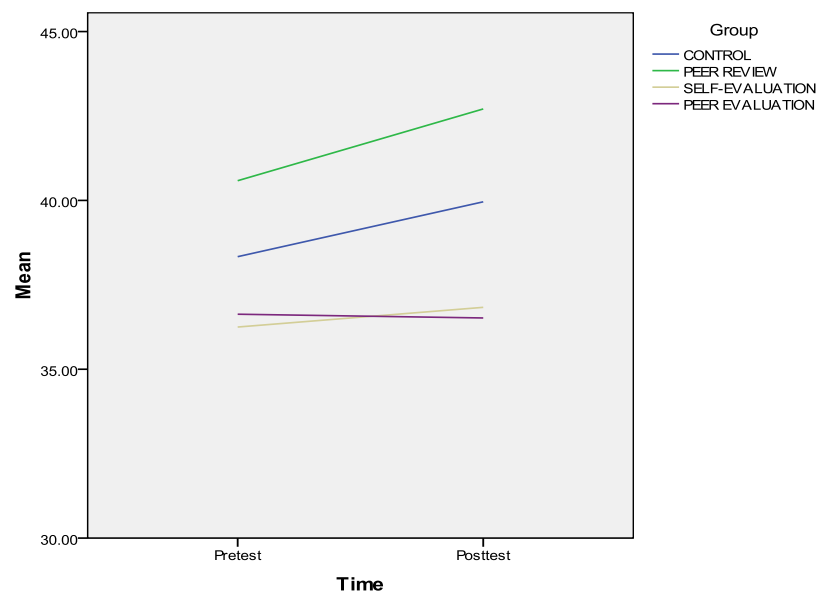
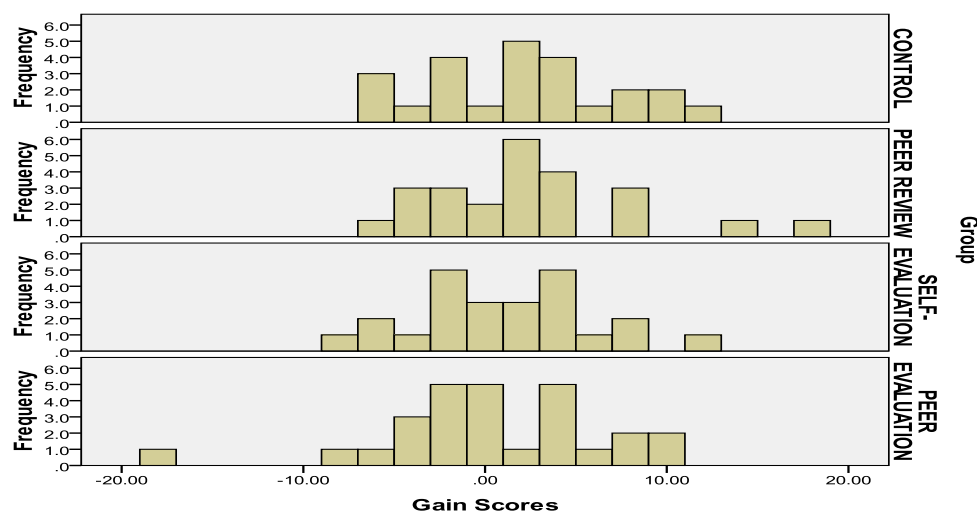


Figure 17 shows a further analysis of the individual students' gain scores to identify any outlying gain scores that might have affected the above results.

Figure 17. Gain Scores by Group



The histogram reveals that the student with a gain score of -18 was the strongest outlier. This student from the peer evaluation group, had a pre-test score of 44 and post-test score of 26. To check the influence on the study, some key analyses were run with and without the student to see how different the results were. These were the one-way analyses of variance (ANOVA) on the pre-test and post-test and the analysis of covariance (ANCOVA) on the post-test with the pre-test as a covariate together with a post hoc test. The results before exclusion of the -18 score and after the exclusion for the ANOVA on pre-test scores were  $p=0.07$  vs.  $0.06$  and on post-test scores, the results were the same ( $p=0.01$ ). As for the ANCOVA on the post-test with the pre-test as a covariate, the results were  $0.05$  vs.  $0.09$  but the post hoc test results comparing differences among the groups did not show any significant results both before or after the -18 score was removed. All these results indicated that the -18 gain score by the student in the peer evaluation group was not an extreme case that would change the significance values in any of the above analyses. In other words, excluding the student did not make much difference in the results. Therefore, data from the student remained in all analyses.

The ANCOVA test was then used to determine the significance of differences between the scores. The results of the ANCOVA for post-test with pre-test as covariate showed that the different treatments received by the groups had a statistically significant effect on their performance in the CCTT-X ( $F(3,94)=2.79$ ,  $p=0.05$ ). To further investigate how the groups differed, Bonferroni post hoc tests were run. Bonferroni tests adjust the significance level to allow for the multiple testing made when several group comparisons are made using the same data (Hochberg, 1988).



Prior to this, the group post-test mean scores were adjusted to what they would be if all the groups scored identically in the pre-test using ANCOVA (Tabachnick & Fidell, 2001). The results of the Bonferroni post hoc test based on adjusted post-test means which were calculated at the overall pre-test mean value of 37.91 are as in Table 22.

Table 22. Post hoc Bonferroni test for pairwise comparisons for CCTT-X

Group vs. Group	Adjusted Group Means	Mean Difference	Sig.
C vs. PR	39.69 vs. 41.04	-1.34	1.00
C vs. SE	39.69 vs. 37.87	1.82	1.00
C vs. PE	39.69 vs. 37.32	2.37	0.54
PR vs. SE	41.04 vs. 37.87	3.17	0.20
PR vs. PE	41.04 vs. 37.32	3.72	0.06
SE vs. PE	37.87 vs. 37.32	0.55	1.00

From the table, the largest difference was between the peer review and the peer evaluation group (adjusted mean difference of 3.72) but even this was not significant. This means that, in this study although the students scored differently in the post-test, the difference in performance level was not statistically significant. Thus, it could not be concluded that one group outperformed others in their general critical thinking ability due to the different treatment received.

A further analysis was then conducted to investigate if there were differences when the four skills in the CCTT-X were analysed individually. The same analysis procedure was operated for this. The raw pre-test and post-test means were first compared and the gain scores were calculated. The detailed scores are as presented in Table 23.

Table 23. Pre-test and Post-test Means of Scores by Skill and Group of Students

Skill	No. of Items	Group	N	Pre-test		Post-test		Gain (Post-test - Pre-test Scores)			
				Mean	SD	Mean	SD	Mean	SD	Min, Max	Range (Max- Min)
Inductive reasoning	23	C	24	13.88	2.35	14.29	2.79	0.42	2.52	-4,6	10
		PR	24	14.67	3.12	15.08	2.43	0.42	3.17	-6,9	15
		SE	24	12.63	3.00	12.71	3.20	0.08	2.84	-8,7	15
		PE	27	13.56	3.60	12.26	3.18	-1.30	3.68	-8,7	15
Credibility of assertions	24	C	24	11.04	2.46	10.67	2.65	-0.38	3.16	-6,5	11
		PR	24	11.50	3.22	12.33	2.65	0.83	2.97	-4,9	13
		SE	24	11.04	3.43	9.80	2.25	-1.25	3.60	-8,7	15
		PE	27	10.60	2.82	10.63	3.14	0.04	3.22	-7,6	13
Deductive reasoning	14	C	24	9.58	2.45	10.67	2.65	1.08	1.79	-2,4	6
		PR	24	10.13	2.83	11.42	2.00	1.29	2.48	-3,8	11
		SE	24	9.00	2.52	10.29	2.42	1.29	1.68	-2,5	7
		PE	27	8.74	2.40	9.44	2.78	0.70	2.76	-5,8	13
Identifying assumptions	10	C	24	3.83	1.81	4.33	1.55	0.50	1.93	-4,6	10
		PR	24	4.30	1.65	3.88	2.30	-0.42	2.57	-5,4	9
		SE	24	3.58	1.95	4.04	1.81	0.46	2.17	-4,5	9
		PE	27	3.74	1.56	4.19	1.88	0.44	2.06	-3,5	8

Table 23 reveals that from the pre-test means, the peer review group scored the highest in all the four skills and these were followed closely by the control group ( $\bar{x}$  =14.67, 11.50, 10.13 and 4.30 vs. 13.88, 11.04, 9.58 and 3.83 for *inductive reasoning*, *credibility of assertions*, *deductive reasoning* and *identifying assumptions* skills respectively). The peer evaluation was third, outperforming the self-evaluation group for *inductive reasoning* and *identifying assumptions* skills ( $\bar{x}$  =13.56 and 3.74 vs. 12.63 and 3.58 respectively) but the self-evaluation group was better than the peer evaluation group in the other two skills namely *credibility of assertions* ( $\bar{x}$  =

11.04 vs. 10.60) and *deductive reasoning* ( $\bar{x}$  =9.00 vs. 8.74). In brief, the peer review and control groups performed the best on all the skills while the peer evaluation and self-evaluation groups outperformed each other for two out of the four skills. These results from the descriptive analysis were then analysed using a one-way ANOVA to check if the differences were significant. The results are as indicated in Table 24.

Table 24. ANOVA for Individual Skill Pre-test Scores (N=99)

		Sum of Squares	df	Mean Square	F	Sig.
Inductive reasoning	Between Groups	51.41	3	17.14	1.82	0.15
	Within Groups	894.25	95	9.41		
	Total	945.66	98			
Credibility of assertions	Between Groups	10.47	3	3.49	0.39	0.76
	Within Groups	854.44	95	8.99		
	Total	864.91	98			
Deductive reasoning	Between Groups	28.68	3	9.56	1.47	0.23
	Within Groups	617.64	95	6.50		
	Total	646.32	98			
Identification of assumptions	Between Groups	6.71	3	2.24	0.73	0.53
	Within Groups	289.31	95	3.05		
	Total	296.02	98			

Despite the different pre-test means scored by the four groups, Table 24 shows that no significant differences between the groups were found.

The post-test results however, showed different findings. The peer review group scored the highest followed by the control group in the first three skills ( $\bar{x}$  =15.08 vs. 14.29, 12.33 vs. 10.67 and 11.42 vs. 10.67 for *inductive reasoning*, *credibility of assertions* and *deductive reasoning* respectively). The peer review group, however scored the lowest in the last skill, *identifying*

*assumptions* in contrast to the control group who scored the highest ( $\bar{x} = 3.88$  vs. 4.33 respectively). The peer evaluation group outscored the self-evaluation group for *credibility of assertions* ( $\bar{x} = 10.63$  vs. 9.80) and *identifying assumptions* ( $\bar{x} = 4.19$  vs. 4.04) skills but for *inductive reasoning* and *deductive reasoning* skills the self-evaluation outdid the peer evaluation group with means of 12.71 vs. 12.26 and 10.29 vs. 9.44 respectively.

The following Figures 18 to 21 help to illustrate the groups' performance on the individual skills in the CCTT-X pre-test and post-test.

Figure 18. CCTT-X Pre- and Post-test  
Means of Scores for *Inductive Reasoning*

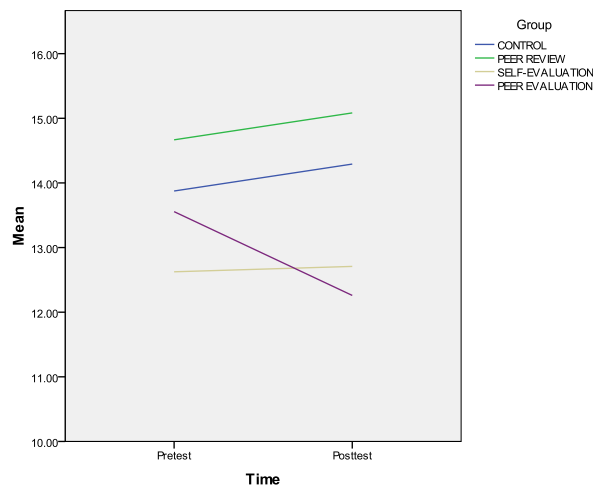


Figure 19. CCT-X T Pre- and Post-test  
Means of Scores for *Credibility of Assertions*

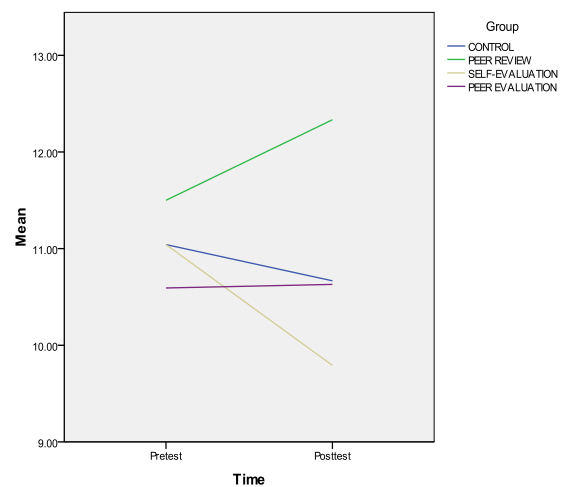


Figure 20. CCTT-X Pre- and Post-test  
Means of Scores for *Deductive Reasoning*

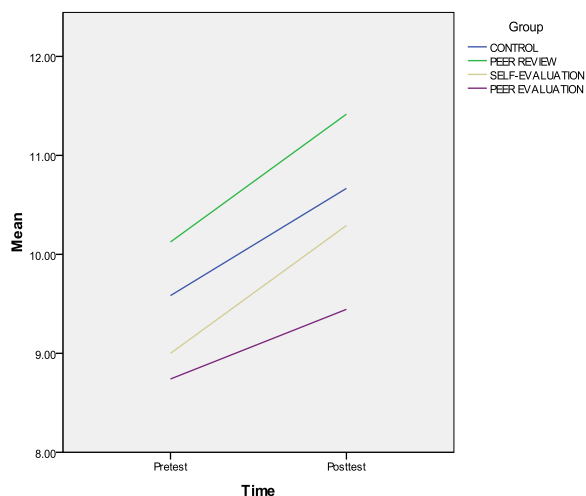
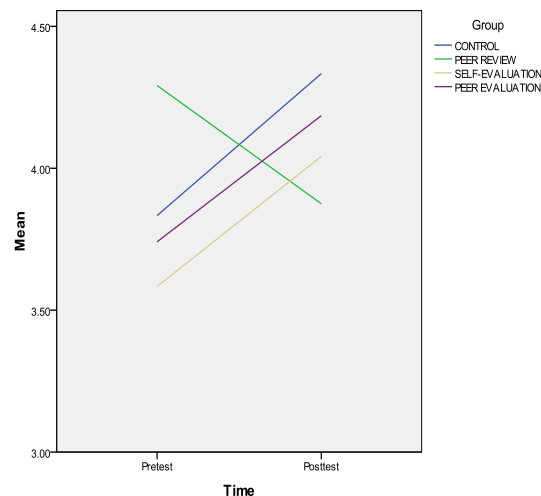


Figure 21. CCTT-X Pre- and Post-test  
Means of Scores for *Identification of Assumptions*



The gain scores were then studied to see how different the post-test scores were compared to the pre-test scores scored by each group. It was revealed that the peer review group shared the best gain mean with the control group and self-evaluation group for *inductive reasoning* (gain  $\bar{x}$  =0.42) and *deductive reasoning* skills (gain  $\bar{x}$  =1.29) respectively. It was also the best group for *credibility of assertion* skill (gain  $\bar{x}$  =0.83). However, surprisingly, the group scored the least for the *identifying assumptions* skill (gain  $\bar{x}$  = -0.42). As for the control group, other than sharing the best gain scores with the peer review group for *inductive reasoning* skill, the group was also the best group for the *identifying assumptions* skill (gain  $\bar{x}$  =0.50), the second for *deductive reasoning* skill (gain  $\bar{x}$  =1.08) but the worst for the *credibility of assertion* skill (gain  $\bar{x}$  = -0.38). The self-evaluation group outscored the peer evaluation group for *inductive reasoning* skill and *deductive reasoning* skill with gain  $\bar{x}$  =0.08 vs. -1.30 and 1.29 vs. 0.70 respectively but both groups shared similar gain mean for *identifying assumptions* skill i.e. 0.46 vs. 0.44. The peer evaluation only outperformed the self-evaluation group for the *credibility of assertion* skill (gain

$\bar{x} = 0.04$  vs.  $-1.25$ ). Table 25 shows the average gain scores for each group in each skill in the Cornell Critical Thinking Test.

Table 25. A Summary of the Groups' Performances on the CCTT-X

Group \ Skill	Mean Gains			
	Inductive reasoning	Credibility of assertions	Deductive reasoning	Identification of assumptions
Control	0.42	-0.38	1.08	0.50
Peer Review	0.42	0.83	1.29	-0.42
Self-evaluation	0.08	-1.25	1.29	0.46
Peer Evaluation	-1.30	0.04	0.70	0.44

Note: +ve gain = improved

-ve gain = declined

In general, although the groups performed better in certain skills than others, each group experienced some progress in their performance on three skills but the performance declined on one skill. To be specific, the peer evaluation group demonstrated improvement in *credibility of assertions*, *deductive reasoning* and *identification of assumptions* skills but regressed in *inductive reasoning* skill. The peer review group displayed better development of *inductive reasoning*, *credibility of assertions*, and *deductive reasoning* skills but not *identification of assumptions* skill. The other group, self-evaluation, exhibited progress in *inductive reasoning*, *deductive reasoning* and *identification of assumptions* skills but regressed in *credibility of assertions* skill.

To determine whether the differences in the scores were statistically significant, the ANCOVA test was used. The results are as displayed in Table 26.

Table 26. ANCOVA for Treatment and Post-test with Pre-test as Covariate (N=99)

Skill	Source of variation	Sum of squares	df	Mean square	F	Sig.
Inductive reasoning	TREATMENT	81.82	3	27.27	4.02	0.01**
	PRE-TEST	175.39	1	175.39	25.86	0.00
Credibility of assertions	TREATMENT	70.41	3	23.47	3.64	0.02*
	PRE-TEST	88.17	1	88.17	13.66	0.00
Deductive reasoning	TREATMENT	17.02	3	5.67	1.43	0.24
	PRE-TEST	215.14	1	215.14	54.12	0.00
Identifying assumptions	TREATMENT	4.78	3	1.59	0.47	0.70
	PRE-TEST	26.72	1	26.72	7.94	0.01

\* $p < 0.05$ \*\* $p \leq 0.01$ 

From Table 26, it was revealed that there was a significant effect of treatment for *inductive reasoning* and *credibility of assertion* ( $p \leq 0.01$  and  $p < 0.05$  respectively). Post hoc tests were then conducted to evaluate pairwise differences among the adjusted post-test means for the *inductive reasoning* and *credibility of assertion* skills scored by each group. The analyses for each skill were done by first calculating the adjusted post-test mean scores and then assessing the differences using Bonferroni adjustments.

For *inductive reasoning* skill, the results yielded from the test based on the overall pre-test mean value of 13.68 are simplified in the following table.

Table 27. Post Hoc Bonferroni Test for Pairwise Comparisons for *Inductive Reasoning*

Group vs. Group	Adjusted Group Means	Mean Difference	Sig.
C vs. PR	14.20 vs. 14.65	-0.45	1.00
C vs. SE	14.20 vs. 13.17	1.03	1.00
C vs. PE	14.20 vs. 12.31	1.89	0.07
PR vs. SE	14.65 vs. 13.17	1.48	0.36
PR vs. PE	14.65 vs. 12.31	2.34	0.01**
SE vs. PE	13.17 vs. 12.31	0.86	1.00

\*\* $p \leq 0.01$

The only significant difference in the groups' performances for the ability to apply *inductive reasoning* was between the peer review and the peer evaluation groups as a significant mean difference was observed between the two groups ( $p \leq 0.01$ ). To be specific, the participants in the peer review group significantly outperformed the participants in the peer evaluation group in their ability to apply *inductive reasoning* with a mean difference of 2.34. There were no other pairs with significant performance differences indicated.

Table 28 shows the results of the Bonferroni post hoc test applied to the *credibility of assertion* skill. The adjusted post-test mean scores had an overall pre-test mean value of 11.03.



Table 28. Post hoc Bonferroni test for Pairwise Comparisons for *Credibility of Assertions*

Group vs. Group	Adjusted Group Means	Mean Difference	Sig.
C vs. PR	10.66 vs. 12.18	-1.52	0.25
C vs. SE	10.66 vs. 9.79	0.87	1.00
C vs. PE	10.66 vs. 10.77	-0.11	1.00
PR vs. SE	12.18 vs. 9.79	2.39	< 0.01**
PR vs. PE	12.18 vs. 10.77	1.41	0.31
SE vs. PE	9.79 vs. 10.77	-0.98	1.00

\*\* $p < 0.01$

Table 28 reveals that for the skill of judging the *credibility of assertions*, only the peer review group and self-evaluation group showed a highly significant mean difference ( $p < 0.01$ ).

It is apparent that the peer review group scored significantly higher than the peer evaluation group for *inductive reasoning* skill and also significantly higher than the self-evaluation group for the *credibility of assertions* skill. This suggests that engaging the students in revising and editing their peers' work mainly through discussion (rather than through grading) promotes higher critical thinking skills among the students. Although asking the students to first grade their peers' work and later discuss their ideas, or even evaluate their own work helped improve certain critical thinking skills tested in the CCTT-X, the peer evaluation and self-evaluation activities had less of an impact on their critical thinking skills development than the peer review activity. This finding is discussed in detail in section 6.1.2 where this finding was contrasted with findings from qualitative analyses.

### 5.3.2 Performance on the Term Paper

The second measure of students' critical thinking development was the term paper. The groups' performances rated by the two instructors were analysed by comparing the means and standard deviations. Table 29 displays the results.

Table 29. Means and Standard Deviation of Scores by  
Group of Students for the Term Paper

Rater	Group	N	Min, Max	Range (Max- Min)	Mean	Std. Deviation
1	CG	24	15.5, 27	11.5	21.94	2.66
	PR	24	19, 27	8	22.98	2.20
2	SE	24	20.5, 27.5	7	23.73	1.69
	PE	27	19.5, 27	7.5	22.80	1.69

The self-evaluation group performed the best in the term paper followed by the peer review, peer evaluation and control groups ( $\bar{x}$  =23.73, 22.98, 22.80 and 21.94 respectively). This order was slightly different from results from the CCTT-X where peer review was the best group. The leniency of the second instructor in scoring the term paper could be a factor. To see if there was any significant difference between the two raters' scoring and the four groups' performance, a two-way nested ANOVA test was conducted. The test was used because each of the two groups' term paper was scored by a different rater. According to Stagliano (2004), "In nested designs, we designate main factors and corresponding subgroups factors. Subgroup factors that are nested within one hierarchy cannot be compared to other factors outside of their hierarchy" (p. 200). Table 30 shows the results.

Table 30. Two-Way Nested ANOVA for Groups' Performances on Term Paper by Two Raters

Source		Type III Sum	df	Mean		
		of Squares		Square	F	Sig.
Intercept	Hypothesis	51603.41	1	51603.41	4288.57	0.00
	Error	24.08	2	12.03		
Rater	Hypothesis	15.97	1	15.97	1.33	0.37
	Error	24.08	2	12.03		
Group(Rater)	Hypothesis	24.08	2	12.04	2.77	0.07
	Error	413.52	95	4.35		

The results revealed that there was no significant difference between the raters and group performance.

### 5.3.3 Students' Perceptions about the Extent to Which They Had Developed Their Critical Thinking Skills in Academic Writing

The question of whether the students' critical thinking skills in academic writing improved with the peer evaluation activity was investigated through analyses of questionnaires distributed to the students prior and subsequent to the interventions. The same repeated measure of students' perceived academic writing ability on the 12 critical thinking skills for academic writing listed in the *CAWAR* was taken on a scale of 1 = Poor, 2 = Average, 3 = Good and 4 = Excellent. A paired sample t-test was conducted on individual students' average scores of the 12 items before (pre-test) and after (post-test) the interventions. The results are as in Table 31.

Table 31. Paired Samples Test for Academic Writing Skill Survey

Group	Pre-test Mean	Post-test Mean	Gain Mean	Std. Deviation	t	df	Sig. (2- tailed)
C	2.57	2.51	0.06	0.44	0.64	23	0.53
PR	2.97	2.60	0.37	0.46	3.90	23	0.001***
SE	2.91	2.42	0.50	0.37	6.46	23	0.001***
PE	2.75	2.41	0.34	0.40	4.40	26	0.001***

\*\*\*p≤0.001

All three experimental groups perceived that their academic writing ability improved with the interventions. The self-evaluation group shows the greatest improvement (gain  $\bar{x}$  = 0.50) followed by peer review (gain  $\bar{x}$  =0.37) and peer evaluation (gain  $\bar{x}$  =0.34). These are all statistically significant based on the paired sample t-test. This contradicts the findings from the analyses on the CCTT-X which revealed the peer review group benefitting the most from the interventions. In particular, despite the actual achievement in the CCTT-X, the self-evaluation group felt that they had improved their academic writing skills more than what was felt by the peer review group who relied on a checklist during the interventions. The control group showed no such difference in their perceptions of their writing ability.

To look further into which academic writing skills the students perceived had been improved, the Wilcoxon signed ranks test, a non-parametric alternative to t-test (Larson-Hall, 2010) was carried out on each of the 12 items in the questionnaire. This non-parametric test was appropriate because the individual items had an ordinal scale. The results are summarised below:

Table 32. Wilcoxon Signed Ranks Test for Individual Survey Item

Group		Item 1 (Post-Pre)	Item 2 (Post-Pre)	Item 3 (Post-Pre)	Item 4 (Post-Pre)	Item 5 (Post-Pre)	Item 6 (Post-Pre)	Item 7 (Post-Pre)	Item 8 (Post-Pre)	Item 9 (Post-Pre)	Item 10 (Post-Pre)	Item 11 (Post-Pre)	Item 12 (Post-Pre)
CG	Z	-1.27	-1.51	-1.90	-0.30	-0.26	-1.21	-0.23	-0.17	-0.78	0.00	-0.44	-1.67
	Sig.	0.21	0.13	0.06	0.76	0.80	0.23	0.82	0.87	0.44	1.00	0.66	0.10
PR	Z	-2.50	-2.89	-2.83	-2.23	-1.45	-2.31	-1.80	-2.65	-1.73	-2.35	-1.30	-1.39
	Sig.	<b>0.01**</b>	<b>0.01**</b>	<b>0.001***</b>	<b>0.03*</b>	0.16	<b>0.02*</b>	0.07	<b>0.01**</b>	0.08	<b>0.02*</b>	0.19	0.17
SE	Z	-3.82	-3.12	-3.64	-2.14	-2.36	-1.73	-1.67	-3.49	-2.67	-3.28	-3.40	-1.29
	Sig.	<b>0.001***</b>	<b>0.01**</b>	<b>0.001***</b>	<b>0.03*</b>	<b>0.02*</b>	0.08	0.10	<b>0.001***</b>	<b>0.01**</b>	<b>0.001***</b>	<b>0.001***</b>	0.20
PE	Z	-1.32	-2.99	-1.61	-2.33	-2.57	-1.11	-1.25	-2.06	-1.00	-2.83	-3.23	-1.73
	Sig.	0.19	<b>0.01**</b>	0.11	<b>0.02*</b>	<b>0.01**</b>	0.27	0.21	<b>0.04*</b>	0.32	<b>0.001***</b>	<b>0.001***</b>	0.08

\*p&lt;0.05

\*\*p≤0.01

\*\*\*p&lt;0.001

Items:

1. Clearly states the thesis
2. Explains difficult terms, concepts, facts or/and ideas clearly
3. Properly breaks down the issue into parts for detailed analysis
4. Supports arguments well
5. Uses only reliable literature
6. Has organised ideas and/or information coherently
7. Integrates other people's ideas accurately
8. Demonstrates a clear stand on the issue
9. Concludes the essay strongly
10. Has used the appropriate academic writing style
11. Cites the literature accurately
12. Writing is free from grammatical, spelling and punctuation errors

Table 32 shows that no significant changes were observed in the control group's perceptions of their critical thinking in academic writing. All the experimental groups, on the other hand, showed significant change of perceptions towards some improvement in many of their skills especially skills 2, 4, 8 and 10. Students in each experimental group reported that they were better able to explain difficult terms, concepts, facts and/or ideas clearly, supports arguments well, demonstrate a clear stand on the issue, and use the appropriate academic writing style. None of the groups, however indicated positive change in their skill of integrating other people's ideas accurately or writing in a way that is free from grammatical, spelling and punctuation errors.

This analysis shows that students perceive that the interventions have been more effective in developing some skills than others. This finding opens up questions about how the rubric might be further developed.

#### **5.3.4 Relationship between Students' Performance in the CCTT-X and Academic Writing**

An investigation of the relationship between the students' critical thinking skills and academic writing ability at the end of the interventions was crucial to see the impact of the different treatments for each group. To measure this, the CCTT-X post-test scores were correlated with the students' final term paper scores. Table 33 below displays the results of Pearson's Correlation Coefficients for each group.

Table 33. Pearson's Correlation Coefficients for CCTT-X Performance and Academic Writing Ability by Group of Students

Group	N	Inductive Reasoning		Credibility of Assertions		Deductive Reasoning		Identifying Assumptions		CCTT -X Overall	
		<i>r</i>	Sig.	<i>r</i>	Sig.	<i>r</i>	Sig.	<i>r</i>	Sig.	<i>r</i>	Sig.
CG	24	0.21	0.34	0.00	1.00	0.10	0.65	0.32	0.13	0.23	0.29
PR	24	0.34	0.11	0.45*	0.03	0.14	0.51	0.16	0.47	0.46*	0.03
SE	24	0.32	0.12	0.35	0.09	0.02	0.93	0.13	0.56	0.38	0.07
PE	27	0.23	0.26	0.35	0.08	0.35	0.08	0.24	0.24	0.42*	0.03

\* $p < 0.05$

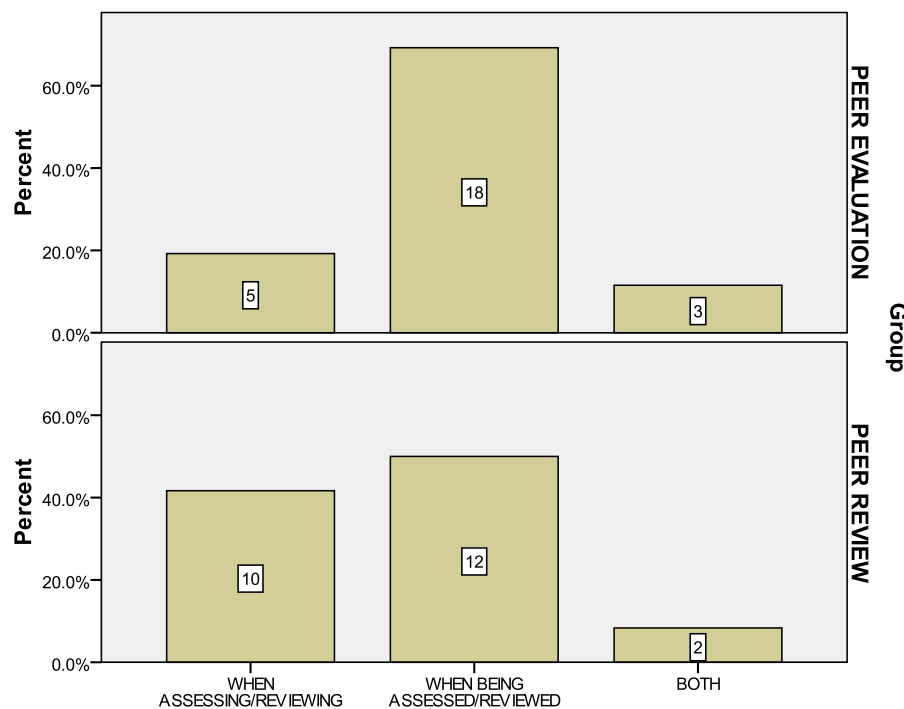
Overall, significant correlations between critical thinking skills and academic writing ability were observed on the peer review group and peer evaluation group (both at  $p < 0.05$ ). For the peer review group, the relationship was strongest for the *credibility of assertions* skill i.e.  $r = 0.45$  ( $p < 0.05$ ). This could be linked to the extra time the group had for discussion that enabled the skill to be developed better than the other groups. Though not significant, higher correlation between critical thinking skills and academic writing ability were observed on the self-evaluation group compared with the control group indicating that all the three groups were different from the control group in their development of critical thinking skills after the learning interventions.

In general, the findings from the students' performance in CCTT-X, the term paper, correlations between the students' performance in CCTT-X and the term paper, and ratings of their perceived critical thinking in academic writing development before and after the interventions suggest that the peer review group promoted critical thinking in academic writing better than the self-evaluation and peer evaluation groups.

### 5.4 Students as Assessors vs. Assesseees

To examine the second research question (i.e. whether students develop their critical thinking skills better when they use the rubric to assess their peers, or when being assessed by their peers), the perceptions of the two groups of students who worked with their peers (i.e. peer evaluation and peer review groups) who experienced being the assessor and assessee using the same assessment criteria were analysed. The results are displayed in the following figure comparing the two groups.

Figure 22. Students' Perception on Whether They Benefitted More as Assessors/Reviewers or Assesseees/Reviewees



It is clear that majority of the students in the peer evaluation group believed that they benefitted more from being assessed with the rubric (69.2%) and only 19.2% of them reported gaining more when assessing others. Very few indicated they benefitted equally as assessor or assessee (11.5%). The relative benefits were less distinctive for the peer review group. Half of the peer



review group believed that they developed their critical thinking more when they were being reviewed. Quite a large proportion also reported they gained more when reviewing others' work (41.7%).

A further analysis of the students' responses to find out whether their perceptions were influenced by (1) their academic writing ability and (2) their critical thinking ability was conducted. The answers to these questions were investigated through quantitative analyses discussed below and also qualitative analyses which are presented in the following chapter.

Independent-sample t-tests were conducted on the students' perceptions of whether they benefitted more from being the assessor/reviewer or the assessee/reviewee and

1. the measure of their academic writing ability i.e. the term paper scores
2. the two measures of their critical thinking ability i.e. the gain scores (the post-test scores minus pre-test scores) and the pre-test scores of the CCTT-X.

The rationale for choosing the CCTT-X pre-test and gain scores were that the gain scores showed how much the students had improved in the test while the pre-test scores indicated where they were to start with prior to the treatment. Table 34 presents the results of the t-tests.

Table 34. T-Tests on Students' Perceptions of Whether They Thought They Benefitted More from Being the Assessor/Reviewer or the Assessee/Reviewee and Their Academic Writing and Critical Thinking Ability

		Perceptions	N	Mean	Std. Deviation	t	Sig.
ACADEMIC WRITING ABILITY	Term Paper	As assessor	15	23.20	1.40	0.79	0.43
		As assessee	30	22.72	2.15		
CRITICAL THINKING SKILL	Pre-test	As assessor	15	40.13	7.62	1.18	0.24
		As assessee	30	37.40	7.15		
	Gain Scores	As assessor	15	2.60	5.72	1.17	0.25
		As assessee	30	0.47	5.76		

As indicated in Table 34, there was no significant mean difference in the perceptions of benefitting more from being the assessor/reviewer or the assessee/reviewee among students with different academic writing ability ( $t=0.79$ ;  $df=43$ ;  $p=0.43$ ), critical thinking ability based on the pre-test score ( $t=1.18$ ;  $df=43$ ;  $p=0.24$ ) and critical thinking ability based on the gain scores ( $t=1.17$ ;  $df=43$ ;  $p=0.25$ ). Thus, for this study with regard to the academic writing and critical thinking abilities, it can be concluded that there was no difference between students who thought that they benefitted more from being the assessor/reviewer and those who thought they benefitted more from being the assessee/reviewee. In other words, the students' perceptions were not related to their academic writing and critical thinking skills.

## **5.5 Students' Perceptions of Peer Evaluation to Foster Critical Thinking Skills in Academic Writing**

Knowing students' perceptions on a learning activity which they have been involved, provides vital feedback on how much the activity was helpful for their learning. In other words, feedback is important for instructional improvement (McKeachie, 1987), thus facilitating learning. In this study, it was important to investigate the students' perceptions on the activities to see the possibility of adopting and developing the activities to help enhance the students' critical thinking while learning academic writing. For this, a survey was administered to all the groups before and after the intervention sessions. In this analysis, the perceptions of each group on each activity were compared and contrasted. Changes in students' perceptions before and after the treatment session were also analysed to determine if their perceptions were influenced by how much the students had benefitted from the different activities. Benefit was exhibited in the gain scores on the CCTT-X.

### **5.5.1 The Potential of Activities to Foster Critical Thinking Ability in Academic Writing as Perceived by the Groups**

Figures 23, 24, and 25 show the perceptions of each group of students on the extent to which their respective intervention namely peer evaluation, peer review and self-evaluation would help to improve their critical thinking skills in academic writing.

Figure 23. Peer Evaluation Group's Perceptions of the Peer Evaluation Activity Potential to Foster Critical Thinking Ability in Academic Writing: Before and After Interventions

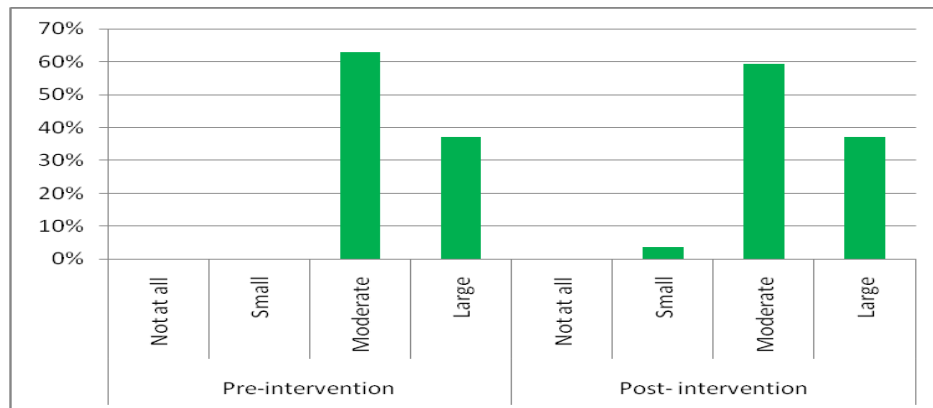


Figure 24. Peer Review Group's Perceptions of the Peer Review Activity Potential to Foster Critical Thinking ability in Academic Writing: Before and After Interventions

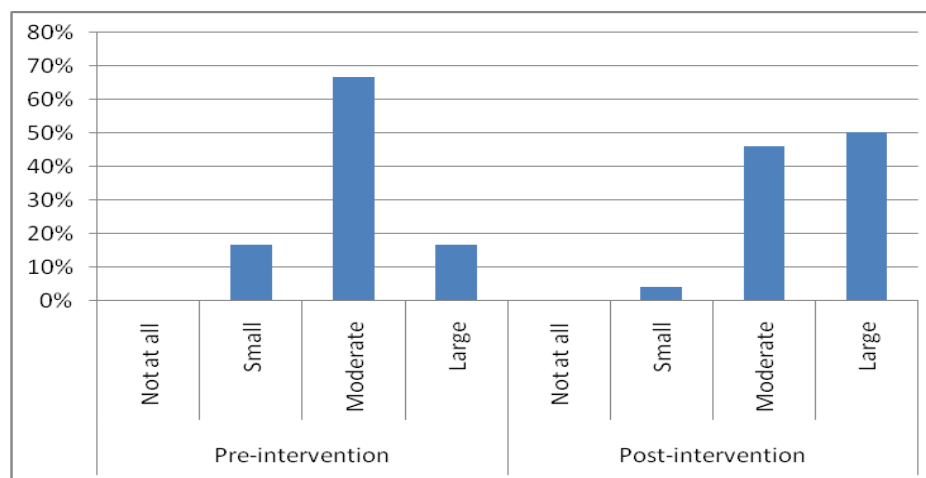
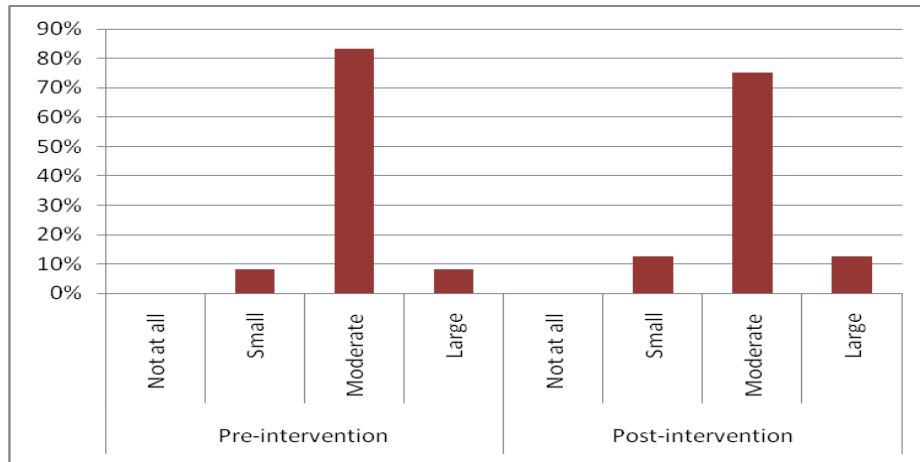


Figure 25. Self-Evaluation Group's Perceptions of the Self-Evaluation Activity Potential to Foster Critical Thinking ability in Academic Writing: Before and After Interventions



Comparing the individual groups' perceptions towards the different learning activities before and after the interventions, the peer review group's perception of the effectiveness of their own strategy was the greatest compared to the peer evaluation and then self-evaluation groups. This could be due to the sufficient experience the peer review had to actually get to realise the value of the activity.

Prior to the interventions, the peer evaluation group was more positive about the potential of the learning activities compared to the peer review and self-evaluation groups that had similar perceptions towards their respective intervention. However, after the intervention, it was the peer review group that was most positive followed by the peer evaluation and self-evaluation groups. The peer review group had a mean change of 0.46 with the majority of the students rating the activity as moderately or largely capable of promoting critical thinking. For the peer evaluation group, the students also perceived the activity as capable of fostering critical thinking in their academic writing to a moderate or large extent. However, they had a slight drop of -0.04 mean

change in their perceptions towards the peer evaluation activities. The marginal drop was by one student who believed that the activity was helpful to a small extent compared to moderately helpful initially after experiencing the activity. The reason for the decline in the student's perception towards the activity as revealed in the questionnaire was due to the student's low confidence to assess others. The students in the self-evaluation group did not indicate any mean change in their perception. Most students perceived the activity as moderately helpful in fostering critical thinking in academic writing.

### **5.5.2 Students' Perceptions of the Potential of the Activities to Develop their Critical Thinking Ability and Their Actual Performance in the Cornell Critical Thinking Test (CCTT-X)**

To see if the above patterns of perceptions were influenced by how much the students gained from the learning activities, Spearman's correlation coefficients were used. Spearman's correlation coefficient, a non-parametric measure was used following Nana and Sawilowsky's (1998) suggestion to use non-parametric test for analyses involving Likert scales.

Table 35. Spearman's Correlation Coefficients for Individual Group of Students' Perceptions towards the Learning Activity and CCTT-X Performance

		Gain Scores
Peer Review	Correlation Coefficient	-0.13
	Sig. (2-tailed)	0.51
Self-evaluation	Correlation Coefficient	0.08
	Sig. (2-tailed)	0.70
Peer Evaluation	Correlation Coefficient	-0.04
	Sig. (2-tailed)	0.85

Table 35 reveals very weak (and in two cases, negative) correlations between the students' perceptions of the activities they experienced and their gain scores of the CCTT-X. This suggests that the students' perception towards the activities were unrelated to how much they gained from the learning experiences.

## 5.6 Key Findings

Analyses of the quantitative data yielded five key findings:

1. Quantitative analyses of the effects of different interventions (i.e. peer review, self-evaluation and peer evaluation) on students' critical thinking revealed that all three groups showed some improvement in their development of critical thinking skills after the interventions based on their performances on the CCTT-X. The students' critical thinking skills were accelerated the more via the peer review activity than the self-evaluation and peer evaluation activities. In particular, a significant difference was observed between the peer review group and the self-evaluation group for the *credibility of assertions* skill and between the peer review group and peer evaluation group for the *inductive reasoning* skill.
2. Different results, however were observed in the students' performance of the LE 4000 term paper. The self-evaluation group had the highest mean score but it was not statistically significant in comparison to the other two treatment groups.
3. The analysis of the groups' perceptions of the extent to which they had developed the 12 critical thinking skills in academic writing listed in the CAWAR revealed that the self-evaluation group showed the greatest improvement after the intervention. This was followed by the peer review group and then the peer evaluation group. However, a check

on whether the effect of the interventions on students' critical thinking had any significant relationship with the students' academic writing ability revealed a significant relationship only for the peer review and peer evaluation groups.

4. The students believed they benefitted more from being the assessee than the assessor. A further check on whether this was influenced by their academic writing ability and their prior and current levels of critical thinking skills indicated no significant relationship.
5. Finally, peer review was perceived as the most likely activity to promote critical thinking in academic writing followed by peer evaluation and then self-evaluation activities. This however did not have any significant relationship with how much they gained from the learning activities.



## **CHAPTER SIX: QUALITATIVE ANALYSES: A CROSS- EXAMINATION OF THE QUANTITATIVE FINDINGS**

### **6.0 Introduction**

This chapter presents the qualitative analyses of the data gathered from the classroom observations, interviews and questionnaires. The three research questions are addressed in order. The chapter ends with a discussion of the key findings from the qualitative investigation.

### **6.1 The Development of Critical Thinking Skills Levels within and among the Groups**

This section addresses the first research question by, firstly, providing an account of the qualitative evidence of critical thinking skills when students peer evaluated, self evaluated and peer reviewed their academic writing, and secondly, discussing possible explanations for differences in the development of critical thinking skills among the experimental groups. Claims for the effective promotion of critical thinking particularly for academic writing are made based on the findings.

#### **6.1.1 The Evidence of Critical Thinking from the Questionnaires and Interviews**

Interviews with the students and their responses of the questionnaire indicated that critical thinking was provoked during the interventions even though different amounts of time were allocated for each group to complete the assessment work. The act of assessing or reviewing and judging the feedback provided by peers required students to think critically. The evidence obtained from the students is presented for each group so that what was experienced by each group in relation to the development of their critical thinking is clearly manifested.

### 6.1.1.1 The Peer Review Group

When asked to reflect on what they gained from the learning experiences, the students in the peer review group indicated some critical thinking had been triggered. Student PR-23 for example reported:

*When reviewing, we have to review other people's work and we have to think what the best is for that person. We may come with the ideas to help the writer. (PR-23, Survey response)*

This response suggests that reviewing required the student to think critically in order to give ideas on how the peer could improve his or her drafts.

Besides this, looking at others' work also benefitted the reviewers' critical thinking as they would be made to reflect on their own errors when judging others' errors. Student PR-6 noted:

*When you look at something like that (the errors made by peers), you tend to think. "Oh, maybe I am doing the same thing, I might be doing the same mistakes." I guess it helps my work as well at the same time. (PR-6, Interview)*

Being reviewed on the other hand, also invoked thinking as students worked on ways to refine their work based on the comments made by peers. This was asserted by student PR-23 above. She claimed that

*When being reviewed, we know our mistakes. This also makes us think more to correct our mistakes. (PR-23, Survey response)*

and this was supported by student PR-10 who wrote:

*Peer review also helps me in writing and also improves my critical thinking as when they comment, it will make me think again and again. (PR-10, Survey response)*

In addition, reviewing against a checklist enabled the writers to practise defending themselves by explaining and giving reasons to their choices to their peers; something they did not normally do with their course instructor. Students PR-17 and PR-6 said:

*Sometimes they said, “Hey, I already put this stand”. So, I’ll ask back, “So, where is the stand?” So, and then after that, it means it’s not clear. (PR-17, Interview)*

*You have to explain why this and that (when peer checking) - sometimes you can’t go against your lecturer - especially when your lecturer is the one who marks your final term paper and everything else, so you cannot allow things when you’re dealing with your lecturers - at times; not all lecturers of course. (PR-6, Interview)*

The peer review activity encouraged the students to think critically both when they were reviewing the work of others and when judging the comments given by their peers on their own work.

#### **6.1.1.2 The Self-Evaluation Group**

For the self-evaluation group, the task of evaluating their own work against the rubric encouraged them to think about their own performance. Students SE-14 and SE-9 commented thus:

*It’s just like - I’m talking to myself. (SE-14, Interview)*

*Self-evaluation really assists me in developing my writing skills based on my own observation. (SE-9, Survey response)*

The experience of evaluating themselves, taught the students to judge their own performance especially by checking it against the performance criteria listed in the rubric.

The evaluation task also allowed comparisons between students' and the instructor's expectations. Students SE-26 and SE-29 wrote:

*After I have done my term paper, I can see clearly my work and after my instructor gives some corrections and comments, I know where my weaknesses and strengths are. (SE-26, Survey response)*

*I think the self-evaluation form is really helpful because I can see the differences between my expectations and madam's (the instructor's) expectations. When there are some mismatches, I will improve my term paper. (SE-29, Survey response)*

The comparison was made possible when the teacher also used the CAWAR to evaluate the students' writing. Seeing the different ratings especially helped to trigger the students' thinking on why the ratings did not match leading to further refinement on the students' work.

#### **6.1.1.3 The Peer Evaluation Group**

The peer evaluation group also reported that the peer evaluation activity fostered aspects of their critical thinking. The students had to judge what was acceptable and what was not acceptable in academic writing, giving more thought to the writing they had produced. Student PE-21 reported:

*I have never done this (peer evaluation) before, so I had to think more, "Is it appropriate to use this source?" Then, I looked at the explanation, "Does it match - with the sources she is using?" (PE-21, Interview)*

Student PE-21 reported evaluating others' work required thoughtful considerations on the work quality, comparing and contrasting within and between texts before a decision is made.

Furthermore, during the evaluation, the students had to seek clarifications and confirmations from their peers. This encouraged questioning between them which required active thinking. The following response suggests this:

*I asked them back, what they meant. If I made corrections, I asked them again, “Is this what you mean? Like this?” (PE-17, Interview)*

Not only did the students question their peers, they actively questioned themselves as they evaluated a peer’s work as reflected by student PE-17:

*When evaluating, I didn’t know whether to put 4, 5. I reread it, “Is she good or not?” I seemed to think. (PE-17, Interview)*

In this instance, student PE-17 was made to think critically as she was trying to decide on the appropriate score to give the peer’s work by reflecting on and matching the score with the quality of the peer’s writing.

In general, for the peer evaluation group, the rubric had given these students a tool that enabled them to think critically when evaluating others’ work.

### **6.1.2 Explanations for the Promotion of Critical Thinking Skills through the Learning Activities**

The findings from the quantitative analyses on students’ performances on the CCTT-X suggested that peer reviewing promoted critical thinking skills more than peer evaluation and self-evaluation activities. In particular, the *inductive reasoning* skill was found to be promoted more through peer reviewing than peer evaluation while *credibility of assumptions* skill was also triggered more by the peer reviewing activity compared to the self-evaluation activity (Section

5.3.1.1). The students' perceptions of the extent to which they had developed their critical thinking in academic writing differed from the evidence from the quantitative study. This section focuses on exploring the possible reasons for these differences.

To unravel the possible explanations for these findings four aspects were explored: the collaborative nature of the learning activities, the use of a rubric as opposed to a checklist, the implementation of the learning activities, and students' previous experience with the learning activities and motivation to learn. Insights on these were drawn from the students' oral and written responses from the interviews and questionnaires. Data from the interviews with the course instructors, classroom observations and the students' use of the commentary space on the *CAWAR* and checklist were also referenced. These data provided evidence about the instructors' teaching practices and the students' learning activities, particularly those aspects which were absent from or could not be collected through interviews with the students and the experimental procedures. Some considerations for discussion in the final chapter were drawn out of the analyses.

#### **6.1.2.1 Learning Environment: Collaboration or Individual Activity?**

Collaborative learning opportunities experienced by the peer evaluation and peer review groups enabled the students to improve their drafts by learning from each others' weaknesses and strengths. Both the instructors (*Instructor 1* who was in charge of the self-evaluation and the peer evaluation groups and *Instructor 2* who taught the control and the peer review groups) agreed with this. They commented:

*I think students always learn a lot better by looking at each other's work, by looking at each other's mistake, by learning from other people's good sentences or bad sentences. I think it makes a huge impact in terms of improving their work. (Instructor-1, Interview)*

*The presence of an audience I think is important for every writer. I think when the students know that their work will be peer reviewed, they're more serious in actually finishing their work in the first place. And another thing is with the activities that we did, I think the students can see that writing is a process. Something that is sometimes missing in our writing classes because they just write and submit, but when they know that, it can be revised and so on it is a good way for the students to see that process. (Instructor-2, Interview)*

Some students referred to different abilities in the group helped the members to learn from each other. This was an advantage gained by the peer evaluation and peer review groups.

*We could also learn from their writing. (PE-21, Interview)*

*I get to see my weaknesses from the eyes of others. (PR-6, Interview)*

Those who were weak in English, for example, commented that their peers pointed out their language errors along with other weaknesses as stated by students PR-17 and PE-21:

*Most of them gave their comments on grammar and content, okay – structural – okay. (PR-17, Interview)*

*My peer explained my errors. She told me, I had to change this and that including my grammar. (PE-21, Interview)*

Others who were better in their English language proficiency also learned from their peers' work.

*I found her conclusion really good you know because she just make an impact if any reader were to read it. Myself, I was reading it, "Wow! That was a very nice conclusion."*  
(PR-23, Interview)

*It's like Nur [a pseudonym], she could simplify things very clearly but she has problems with grammar. Obviously, I don't have problems with grammar but I have problem with simplifying things.* (PR- 29, Interview)

Students PR-23 and PR- 29 admitted benefitting from reading the work of peers' who were not as good in the language but had better skills in writing conclusion and presenting ideas.

Skilled students benefitted because they articulated their thinking and weaker students learned from the work of others.

Shared learning was anticipated by the control group who had heard about peer review and peer evaluation through an explanation during the interview. Student CG-12 for example claimed that:

*I can experience many level[s] of writing. If let's say I need to assess three papers - I need to differentiate them - which is good, which is not good [and] which is average. Because - because different people will write [differently and] come out with different ideas. So, from that assessment I can identify which one - is the best and which one is not the best.* (CG-12, Interview)

Student CG-12 asserted that, if he was given the opportunity to check others' drafts, he would be able to see how they were different or similar from each other which would then lead him to learn what made one better or worse than others.



Working collaboratively with peers to check on each others' term papers also helped provide multiple comments to guide the students in improving their drafts. Again, the peer evaluation and the peer review groups particularly had this advantage. The students were glad that they could have comments from peers and not just from the course instructor that permitted them to learn more and helped them improve on the drafts.

*It's important to get comments from peers so that I can work on my shortcomings. (PE-8, Survey response)*

*If my friends check and then madam (the instructor) also checks [the draft], I can compare them. (PE-17, Interview)*

This is also shared by student PR-10 from the peer review group who reported:

*Peer review helps me in writing and also improves my critical thinking as when they comment, it will make me think again and again. (PR-10, Survey response)*

Students PE-8, PE-17 and PR-10 appreciated the opportunity they had during the interventions i.e. receiving views from peers. They regarded the peers' opinions on their work as equally as significant as the instructor's comments. The multiple comments received allowed them to see more areas in their drafts that could be improved which later led to a better final piece of the term paper as explained by students PE-24 and PE-20:

*With my essay being assessed by my friends and my lecturer, I know where the mistakes that I cannot detect by my own are. It helps me to improve my writing skills. (PE-24, Survey response)*

*After doing the activities, peer evaluation helps a lot in improving my term paper rather than individual grading (by the teacher). (PE-20, Survey response)*

When evaluating the comments on their work, students' critical thinking was invoked. They themselves had to decide what was best to do based on the feedback gathered in order to improve their work.

*The peers only provided us with ideas. They could not change our work. They just threw some ideas to us and we had to compare. (PE-21, Interview)*

This advantage of sharing ideas and comments was absent in the self-evaluation activity. Some students in the self-evaluation group wished that they had had someone to look at their work and comment on it. Students SE-4 and SE-15 stated:

*It is very helpful if there are other people to evaluate our work. We can not only look at our work, but also compare it with others', so we can know which one is better, and then we can improve it. (SE-4, Interview)*

*Others can give ideas and criticise, and tell us our weaknesses and strengths. (SE-15, Survey response)*

The self-evaluation group received feedback only from the instructor. However, the CAWAR provided space for comments. Therefore, students in this group were able compare their expectations with the instructor's. Any observed discrepancies among the sets of comments signaled the need for further reflection which would lead to some improvement on the paper. Students SE-29 and SE-26 stated:

*I think the self-evaluation form (the rubric) is really helpful because I can see the differences [between] my expectation [and] madam['s] (the instructor's) expectation. When there are some mistakes corrected, I was able to see them and improve my term paper. (SE-29, Survey response)*

*After I have done my term paper and self evaluated it, I can see my work clearly and after my instructor has given some corrections and comments, I know where my weaknesses are and which parts I am good at. (SE-26, Survey Response)*

Nevertheless, the need for others to check their work before submitting it to the instructor was perceived helpful especially when students were unable to detect their own weaknesses. Students SE-14 and SE-11 are among those students who identified this problem.

*I don't know what is wrong or right because I write for myself. (SE-14, Interview)*

*When me, myself, evaluate my own paper, many things were overlooked. I think I have done it well but when madam (the instructor) has corrected it, [there are] many mistake[s] here and there. I think if I had time for peer evaluation or peer review, I can improve a lot. (SE-11, Survey Response)*

These students from the self-evaluation group pointed out that the difficulty faced was due to having no one else to evaluate their work.

Similar views on the peers' role to enhance learning were shared by the control group. Students CG-28, CG-7 and CG-30 commented:

*I think it would be better when they (peers) can read my work because each one of us can comment on each other's work. We can exchange our opinions. (CG-28, Interview)*

*I take it better to write with my friend's judgment on me. (CG-7, Survey response)*

*Other[s'] opinions will let my mind broader. (CG-30, Survey response)*

Another student from the control group, student CG-32, felt the need to have someone other than the course instructor check her work. She reported that asking friends to help look at her work outside the classroom was not a great help as they were not willing to devote the time they had to her.

*I feel like asking others' help like other teachers, or friends whose English is good. I did refer to a few friends but they did not really help much because they also have other things to do. (CG-32, Interview)*

Students who had not experienced peer evaluation or peer review activities in the classroom but had heard about them reported working collaboratively in the classroom would force them to work together within class time.

The comfortable learning setting developed through collaborative work promotes autonomous learning. This means that the students had better control of their learning and therefore developed greater confidence and later competence in learning. This is reflected in the student PR-6's report:

*I find that it's a lot of fun doing it. We feel like a lecturer sometimes. (PR-6, Interview)*

Student PR-6 clearly appreciated the potential of the learning activity to foster confidence when peer checking. Students in Malaysia commonly lack the experience of reviewing or evaluating their peers' work and for this particular student, he cherished the given opportunity. He further added:

*I think it's really good, because sometimes it's easier for you to interact with your peers rather than to interact with your lecturers, you know. Sometimes there's this bond between - not the bond like this. There's a wall, you know that separates students from*

*the lecturer. Sometimes there was a tendency not to see the lecturer when we have some problems - probably some petty problems. So, by having this kind of peer reviewing sessions, you can actually feel much more comfortable, because they are just like you, you know. Even if they make mistakes, you're making them as well, so you interact better. It's more of an open session. You can say anything you want. There're no hard feelings about it. (PR-6, Interview)*

Student PR-6 explained that at times, he felt more comfortable discussing his drafts with peers than with the instructor because he thought that the gap that existed between a student and an instructor discouraged a student from consulting the instructor and discussing matters openly together. Peers, on the other hand, being in the same position, could better understand him as a learner, thus they were able to give good support to enhance learning. Student PR-29 corroborated this.

*Basically it's helping a lot since everyone is reviewing everyone - because sometimes the lecturer oh, we have to do it in very high expectation, so it's kind of very pressured but then with friends we know we are of the same standard and then after that we know that madam would review it but then it helps better because we could, advise straight in the way that this is wrong this is not okay or not - and especially - good explanation, punctuation and some of my peers they have lack but at the same time they have something more than me that I really want to have it. (PR-29, Interview)*

Student PR-29 added that an instructor normally has high expectations of his or her students' performance which creates some degree of anxiety for the students. Students felt obliged to follow instructions without much questioning. However, through peer checking activities, either reviewing or evaluating, they could exchange opinions unreservedly and make the necessary corrections before submitting their work to the instructor.

Furthermore, a student in the peer evaluation group, student PE-13, claimed that the learning activity helped build up his academic writing ability. He wrote:

*Those three evaluations and review gave good results in performance. I think that peer advice is also important instead of guidelines by lecturers. At the beginning, I felt that I was weak writing an argumentative essay, but by comparing with others', I know what I also can do. Thus, peer evaluations are important.* (PE-13, Survey response)

To student PE-13, the chance to examine others' writing brought about a positive effect on him. The belief that he could not produce a good piece of academic writing changed when he could see how his peers worked on their papers.

A cooperative and collaborative learning environment through peer review and peer evaluation enabled shared and autonomous learning which enhanced self-confidence and critical thinking in contrast to individual learning experienced by the self-evaluation group. This corroborates what has been claimed in the literature, for example, by Vygotsky (1978), Johnson and Johnson (1986), Totten et al. (1991), Nelson, 1994 and Gokhale (1995) that a higher level of thought was fostered when students worked collaboratively than individually. The diversity of knowledge, experience and skills benefitted their learning. This explained the above finding of the influence of learning environment to promote critical thinking which favoured peer review and peer evaluation. Hence, this could be one explanation for why the self-evaluation group did not perform better than the peer review group. The self-evaluation group was especially deprived of a shared learning experience and only experienced some degree of autonomous learning by the use of the CAWAR.

Based on the above, both the instructors and the students indicated that they valued the learning that arose through collaborative activities where students discussed the comments of their peers with each other. This applies both to students who had experienced peer review or evaluation and those from the control group and the self-evaluation group who heard about it from others. This analysis leads to the claim that:

**Claim 1: Working collaboratively to evaluate or review the work of peers enhances critical thinking.**

#### **6.1.2.2 Learning Instruments: The CAWAR or Checklist?**

Explanations for the greater promotion of the critical thinking skills among the peer review group were also sought by studying the influence of the CAWAR used by the peer evaluation and self-evaluation groups and the checklist used by the peer review group.

The CAWAR evaluation form, as explained in the third chapter consisted of a list of assessment criteria, a rating scale of performance and a commentary space while the checklist only had the assessment criteria and the commentary space. Therefore, both instruments showed the students the criteria for quality academic writing and enabled the specific details that needed improvement to be identified. Students PR-17, SE-4, SE-20 and PE-12 said:

*If you want to comment just briefly, it's really hard. So, commenting by using those points is much easier because based on the points, I know what to check. (PR-17, Interview)*

*I prefer the evaluation form because that one - the rank and also the points of our evaluation of the project is very detailed relating to what I have to do for the project, so it's very helpful if we have the evaluation form. (SE-4, Interview)*

*Listed (in the rubric) were a lot of items - which we needed to have in the writing, didn't we? So, we studied them and we knew that we, "Oh, I don't do well in this," ... like the citation - something was wrong there, so later we corrected it. We tried to get the excellent ranking in the list. (SE-20, Interview)*

*For my paper, some of my paragraphs are very long, so my friends commented and so I have to break down those things in shorter paragraphs. (PE-12, Interview)*

These comments indicated that by knowing the criteria the students were well guided in their task. They knew what to fulfill in order to achieve a good grade. The long paragraphs of the third student, for example, was commented on by his peer who was guided by the scoring guide which included appropriate breaking down of issues into parts for detailed analysis as a criterion of a quality academic writing.

The control group's opinion was also sought. Student CG-32 welcomed the use of a guide that would indicate the skills needed to be displayed in her academic writing that would be evaluated by the instructors.

*I felt I need more time (to get feedback) from her. If she could have a scoring guide, we can know where we are weak in and good at. That's what I feel. (CG-32, Interview)*

However, indications of the quality of performance were expressed differently by the CAWAR and the checklist. The CAWAR was capable of giving the students a measure of the quality of their work. The students using the CAWAR i.e. self- and peer evaluation groups claimed:

*We know actually which level we are at. For example, if we got 3 over 6 so we know we are weak in certain areas. (PE-12, Interview)*



*Worth noting, the particularity of the skills listed [in the scoring rubric] also makes it easier to know what one's standard of writing is. (SE-9, Survey response)*

*Before the peer evaluation with scoring mark takes place, the way of doing an essay is a bit tough as there was no direction and I've no idea what to write about. But with peer evaluation, I'll be more confident as I know what direction to be used as it is discussed by the peer and it helps me out to come with good essays. (PE-3, Survey response)*

To students PE-12, SE-9 and PE-3, having the numbers to quantify their performance was important as it informed fairly accurately the quality of their work and how much improvement was required of their papers. As indicated by the third comment, the grades also assisted the student in planning what to do next in order to improve the paper. This was a new experience for the student who previously had limited knowledge on how to go about composing his term paper.

Even a student in the peer review group, student PR- 6 agreed with this. When asked about peer evaluation, he commented:

*It's good, because at least you can tell whether this person is in a good level, or very good, excellent. Right now we just say okay, this is correct, this not right. This is wrong, but you don't know exactly what, how good I am. So, if you were to give some grades to it for example 3 marks, 5 marks, and then you know, it helps. (PR- 6, Interview)*

Nevertheless, this advantage was not seen positively by other peer review students who argued:

*I think peer review can benefit the students more and develop their critical thinking skills because without a scoring guide, they need to criticise it properly and give their reasons, not just marks. (PR-17, Survey response)*

*A scoring guide limits the critical abilities to that of the scoring rubric. Students might not be as analytical when guided by a scoring guide. (PR-1, Survey response)*

Students PR-17 and PR-1 claimed that using a rubric restricted one's critical thinking as the focus was on rating. However, as hinted by the peer evaluation students earlier, evaluating and being evaluated did foster their critical thinking as the students had to judge the extent to which each criterion in the rubric matched the quality of their academic writing and this task required them to be analytical. The students in the peer evaluation group in particular were even asked to discuss their evaluation orally, especially the vague or weak points by questioning, clarifying and reasoning.

In the next section, the ways in which the CAWAR and the checklist were adopted are investigated to find possible reasons for the lower performance of the peer evaluation group compared to the peer review group. It had been hypothesised that the peer evaluation activity would have been able to offer more learning advantages than the peer review activity by including the evaluation feature in order to encourage a more active judgment of the quality of the drafts. This lack of the grading feature in the peer reviewing activity was reflected upon by student SE-4 in the self-evaluation group. The student wrote:

*The peer review, I think, is not really helpful because it does not include the evaluation form for others to grade our papers. (SE-4, Survey response)*

The use of a rubric such as the CAWAR, and a checklist, helps to indicate the quality of students' work. This advantage was realised not only by the peer evaluation and self-evaluation groups who had experienced using a rubric, but also the peer review and the control groups. Therefore, a second claim is that:

**Claim 2: The use of a rubric or a checklist enhances students' awareness of the critical thinking skills listed.**

Evidence from the study suggests that a rubric offers greater potential for promoting critical thinking than a checklist. A rubric, such as the *CAWAR*, enables students to practise more of their critical judgment as they grade the work in addition to commenting on the assessment criteria. Another claim emerged from this is:

**Claim 3: A rubric has the potential to involve more detailed critical thinking than the checklist because it requires students to make judgments about achievement.**

Although the self-evaluation group also used the *CAWAR*, they were disadvantaged by not having the opportunity to discuss the ratings with someone/a peer. One possible reason for the significantly lower performance on the *credibility of assertions* skill among the self-evaluation group was that the students lacked the opportunity to practise assertiveness through interactions with others in discussion activities. This led to the fourth claim that:

**Claim 4: Both a rubric and a checklist are more valuable when followed by discussion.**

Hitherto, only explanations of the lower performance of the self-evaluation group compared to the peer review group have been exposed i.e. the absence of a collaborative learning environment and the opportunity for peer discussion. To better understand the differences in performance between the peer review and peer evaluation groups the ways the respective learning activities were implemented were investigated.

### 6.1.2.3 Implementation of Learning Activities

This section considers the possibility of differences between the groups based on how the learning activities were implemented.

The instructors' lesson delivery did not account for the gap in the peer review and peer evaluation groups' performances in the CCTT-X. Through classroom observations, it was clear that both instructors went through the reading materials assigned via the Learning Management System (LMS) and attended to questions regarding the topics. Both instructors also pointed out to the students the value of the learning interventions for the students' learning. From the start of the interventions, both instructors showed their interest in seeing how the learning activities would help improve students' learning. This was evident as they gave way for the interventions to take place. Although there is no direct evidence, and there is no evidence to the contrary, it is likely that:

**Claim 5: A teacher's belief in the importance of a learning activity and the learning instrument and the teacher being able to adopt the activity and the instrument are important.**

The ways students were grouped and the time allocated to the activities however were seen as the possible reasons for the difference in the performance of peer evaluation and peer review groups.

As reported earlier in the methodology chapter, unlike the students in the self-evaluation and peer review groups, the peer evaluation group consisted of students from different faculties (self-evaluation and peer review groups were all human science and law students respectively while

the peer evaluation group was made up of 18 economics and nine accounting students). The mix of faculties in the class posed some difficulties for the peer evaluation students when they worked together checking the drafts, particularly as the groupings of students was not arranged by faculty within the class. Therefore, groups in the peer evaluation class included members from both faculties. When asked about the problems faced during the intervention two of the five students interviewed commented:

*Our friends, most of them cannot comment for us because in my group I have BBA students but actually I prefer to work with the accounting students. I think, then only I can comment more on their paper. When I'm with the BBA students, I don't have much knowledge on their paper. (PE-5, Interview)*

*Maybe because EAW mixes the students. I had BBA students in my group, so when I used BACC terms like MBSB they did not know it, so they just read it quickly and didn't know what to say. Their understanding was different from us the BACC students. (PE-13, Interview)*

Due to the imbalance of the number of students from the two faculties in the peer evaluation group and late attendance by some students, a few groups were formed by students from different faculties. To students PE-5 and PE-13, this did not allow for effective peer work among them as the different backgrounds inhibited them from contributing to each other's learning. The students preferred to work with those from the same faculty with whom they could share their ideas. In academic writing lessons, there is a great tendency for the students to write on a topic pertaining to their own field which might not be easily comprehended by students from another field. Thus, for peer evaluation to be more effective and meaningful to the students of academic writing, the group formation should take into account the students' shared subject knowledge. The next claim is thus,

**Claim 6: Sufficient knowledge of a particular topic allows more contribution from a peer doing an assessment. Peers who understand the writing they are critiquing are able to comment in greater depth than peers from a different kuliyyah/faculty.**

The time allowed for the interventions was not equal between the two groups. The 14 week/semester (two 1½-hour meetings a week) allocated for the course had to cover the lessons, the mid-term test, oral presentations as well as the interventions. Although the instructors initially agreed that the whole 1½-hour period would be used for the respective interventions, the instructor teaching the peer evaluation group only allowed one hour for each intervention period as she needed to use the other half an hour for teaching. This contrasted with the other instructor who allowed the full class period to be used for the respective intervention. The shorter allocation of time given for peer evaluation had reduced the discussion activity which was supposed to be done after each draft was read and rated. This was especially experienced by those students who spent more time on reading and rating as they did not have the time to sufficiently discuss the drafts. The students were still in the midst of the discussion when the instructor asked them to stop. On average, out of the 20 minutes given to read, rate and discuss each draft, three quarters of the time was spent on reading and rating the draft. The limited time was a setback for the peer evaluation group. A student in the group, student PE-12 indicated this experience.

*The time given to complete the tasks (reading, rating and discussing) was too short - in terms of checking the essay, so we cannot identify more problems.” (PE-12, Interview)*

When the instructor of the peer evaluation group was asked in an interview about what could impede the successful implementation of peer evaluation, she identified adequate time for the multiple tasks of reading, rating and discussion would be a prerequisite. She said:

*It depends a lot on the time constraint. If we have a big number of students and then the time for the class itself is reasonable enough, we can have peer evaluation. After the students look at each other's work, they can have discussion, but if it is just one hour class, then that's going to be a bit tricky. I think the time factor is something that has to be rectified. (Instructor-1, Interview)*

Sufficient time for the activity nonetheless, was lacking during the interventions. It was beyond the instructor's control to increase the time.

The instructor teaching the peer review group, Instructor-2, however, dedicated the full period for each intervention. She did not mention lack of time as a problem to the success of peer review activity. Having the whole period of 1½ hours was enough to read and discuss the drafts. The allocation of time issue is discussed further in the next chapter. A claim from this analysis is:

**Claim 7: Sufficient time for students to discuss ratings made of their peers' work must be provided.**

One piece of evidence that could indicate time allocation differences between the peer evaluation and the peer review groups was the students' use of the commentary space on the CAWAR and the checklist respectively. The commentary space was included in the rubric and on the checklist so that students could write details of the strengths and weaknesses of the peers' academic writing for discussion. A student in the self-evaluation group, student SE-9 pointed out that the rubric only showed where one was along the scale but it didn't provide a detailed account of the

performance especially what exactly was wrong with the paper. Thus, there was a need for the evaluator to fill the gap. He commented:

*If the mark was low, I could improve certain relevant parts. However, the problem was, I did not know the details like when a problem was identified with the thesis statement, I didn't know which part exactly within the thesis statement needed improvement.*" (SE-9, Interview)

The commentary space was expected to be used to address this problem. Compared to the peer review group, the commentary space was used less by the peer evaluation group. Analyses of the students' use of the commentary space on the rubric showed that the space was either left blank or filled with very few words indicating focus was given more on reading and understanding the drafts, and grading during the limited given time (See Appendix 15 for typical samples of peer evaluation and peer review students' written comments on their peers work). When asked what sorts of written comments he received, a student in the peer evaluation group, student PE-13 stated:

*My friend just put one sentence there, "Improve."* (PE-13, Interview)

Such a comment was too general and did not help the student much to improve his draft. Although having the performance quality indicated on a scale was helpful, the student still needed to be informed what exactly was needed to be corrected or improved.

This suggests that an adjustment needs to be made to the CAWAR. Since the students were deprived of the details of improvements needed to be done, a column has been added to the right of the rubric to give a space for the assessor to include specific details that would help to better



exemplify the rating made on the assessment criteria, such as, the page reference, explanations of errors and suggestions. Please refer to Appendix 16 for the final design of the CAWAR.

Another way to address the above problem is to ensure that students engage in a discussion after they rated their peers' work. The use of the CAWAR alone without any discussion activity following was insufficient for fostering students' critical thinking development. The importance of discussion for fostering critical thinking skills development has been reported in the literature (Totten et al. 1991; Tsui, 2003). Through discussion, students heighten their thinking as they question, negotiate, defend, clarify and justify ideas. The peer review group had the privilege to sufficiently engage themselves in discussion which had especially significantly improved their assertiveness skill as revealed by the quantitative analysis (section 5.5). The shorter time period however, had indeed hampered discussion opportunity among the peer evaluation group. This helped explain why the group did not perform better than the peer review group especially the inductive reasoning skill (section 5.3.1.1). The students might not have enough time to especially justify their ideas due to the limited time given. Another claim is then:

**Claim 8: Sufficient discussion time is needed to promote the development of critical thinking.**

#### **6.1.2.4 Students' Previous Experience with the Learning Activities and Motivation to Learn**

When questioned about the problems faced during the interventions, a few students stated that they were ill-equipped to undertake the assessment as reflected by the following students PE-27 and SE-14:

*I'm not confident in evaluating my friend's work. (PE-27, Survey)*

*I don't know what - is it wrong or is it right because I write for myself.* (SE-14, Interview)

When asked if the rubric used to guide evaluation contributed to the problem, the students thought, not. In fact they claimed that the wording, content and format of the rubric were all fine as reflected by students PE-5 and SE-9:

*I think that the form (the rubric) has included everything required of the paper.* (PE-5, Interview)

*It is clear. The wordings are easy to understand.* (SE-9, Interview)

These comments helped support the face and content validity of the CAWAR. Unlike the earlier groups of students who were referred to at the earlier stage of the development of the rubric (the four University of Canterbury PhD students and the *Writing the Academic Essay* (ENGL 117) course students of the same university) who suggested some improvement, the students in the peer and self-evaluation groups, being asked whether they thought changes would be useful, did not suggest any changes as being necessary to the rubric. This could be due to the adjustments done earlier that the CAWAR was taken well-formed by these students.

Perhaps the students' unfamiliarity with rubrics to guide evaluation led them to regard the rubric used in the present study as complete and well-designed. Further investigation on the rubric for use by other groups of students who have experienced self- or peer evaluation is deemed necessary to confirm the claims made by the students in the present study.

The peer and self-evaluation groups reported that they lacked the experience to evaluate peers or themselves. The students were used to receiving evaluations and comments from their instructors that, when asked to assess and grade their own and others' work, they found it hard. In the Malaysian educational system, assessment for learning involving students assessing their own performance is not a common practice. Students at primary, secondary or tertiary level are seldom given the chance to self- or peer grade their learning performances. Rather, the most they are allowed to do is just to self- or peer review their work but not grade it. They are more commonly expected to complete a given task to the best of their ability and then the teacher assesses and grades their performance. The following students SE-9 and PE-21 for instance, said:

*We have never done self-evaluation before so, the first time we did it, we felt that our work was already good. The truth is when the instructor checked the work, there were problems here and there.* (SE-9, Interview)

*I have never done this (peer evaluation) before, so I had to think more, "Is it appropriate to use this source?" Then, I looked at the explanation, "Does it match - with the sources she is using?"* (PE-21, Interview)

This unfamiliarity with the evaluation task among the self- and peer evaluation groups might have influenced their performances during the interventions.

Motivation influences the way students learn. It affects the degree to which students immerse themselves in the learning activities which then affects their learning performance. Lack of focus on the learning activities is one manifestation of lack of motivation. In the study, it imposed another problem particularly to the peer evaluation and peer review groups. Students PE-13 and PR-11 said:

*My friend, he did not really read it. Just roughly – a little bit of assessment – okay, and then comment. (PE-13, Interview)*

*If we are sincere in reviewing other's work, then we will likely be focused and we will not do it fast and hurriedly. We will just do it one by one. So for me, I took a long time to review one work but I saw many of my friends, they just did it very fast and they were interested to go out. (PR-11, Interview)*

For the peer evaluation group, the multiple tasks of reading, understanding, rating and discussing the drafts could be an explanation for this problem. There was too much to do thus some students failed to (or resolved not to) become actively involved. However, for some other students, it was an attitude of not willing to do the task seriously as mentioned by the student from the peer review group above. This was supported by a student in the peer evaluation group, student PE-12:

*If we only do our job and the other partners don't do their work, it'll be more difficult to share the problem or to share the ideas.” (PE-12, Interview)*

This scenario was reported only by the peer evaluation and peer review groups as the students depended on each other to improve their writing. Observations of the students' activities during interventions, however indicated that there were some students who reflected deeply on their drafts as they graded and wrote comments about their own performances but some others only gave a quick assessment of their work. This stemmed from a lack of understanding about the value of the activity. Ignorance about the benefit that could be gained restrained them from contributing to the learning process. Claim 9 is:

**Claim 9: Students who understand the value of a learning activity are better engaged in the activity.**

In general, analyses of the likelihood that the students' previous experience with the learning activities and motivation to learn influencing their learning performances revealed that the deficiency of experience with self- and peer evaluation could have distressed the self- and peer evaluation group's performances during the interventions. In terms of motivation to engage in the learning activities, all the three groups were observed having some degree of attentiveness to the activities, but their level of motivation was not systematically measured to help determine if it influenced the varying CCTT-X performances. The issue, however is worth a more detailed discussion as it raises the question of how to promote the value of a learning activity to students which, given due attention helps to further enhance the potency of a learning activity to promote critical thinking. This is addressed in Chapter Seven under the discussion of Claim 7 (section 7.3.1).

## **6.2 Perceptions of Benefitting More as an Assessor/Reviewer or Assessee/Reviewee**

After the interventions, the students in the peer evaluation and peer review groups were asked to indicate whether they thought they developed their critical thinking more, when assessing/reviewing a peer's writing or when their own writing was being assessed/reviewed. As mentioned in the previous chapter, the peer review group's perceptions were also studied to allow comparisons on the students' answers especially due to the similar roles they played in the activities i.e. checking others' work or having their work checked. Table 36 displays the students' replies to this question.

Table 36. The Number of Students Who Believed They Benefitted  
More as Assessors/Reviewers or as Assesseees/Reviewees

Task student benefitted more ...	Frequency		TOTAL
	Peer Evaluation	Peer Review	
When assessing/reviewing their peers' work	5	10	15
When being assessed/reviewed by their peers	18	12	30
Both	3	2	5
Missing	1	0	1
<b>TOTAL</b>	<b>27</b>	<b>24</b>	<b>51</b>

The majority of the students in the peer evaluation group (18 out of 27) reported they benefitted more when being assessed by their peers. However, for the students in the peer review group their perceptions of benefitting from reviewing and being reviewed were almost equal (a difference by two students i.e. 10 compared to 12 out of 24). Very few students in both groups claimed benefitting equally from both roles.

Below are the justifications given by the students for the choices made and their interpretations. Answers were gathered from the face to face interviews and responses to the questionnaire.

### 6.2.1 The Peer Evaluation Group

Very few students in the peer evaluation group declared that being an assessor invoked critical thinking. Students PE-5, PE-18 and PE-19 wrote:

*I have to think more to find the weaknesses of the paper.* (PE-5, Survey response)

*When assessing, I will find more ideas and information in doing my[own] term paper.*  
(PE-18, Survey response)

*It is because I have to think carefully and critically when evaluating [the drafts].* (PE-19, Survey response)

The first student claimed that searching for the flaws in her friend's writing required deeper thinking. The second student reported reflecting on her own term paper when looking at her peers' work. The peers' drafts became models for her to improve her own work. The third student asserted that she had to read the peers' drafts carefully and critically before she could suggest improvement and grade the peers' work. It was implied that for a fair rating this student made sure that she had good reasons and explanations to support her decision.

The majority of the students in the group perceived that they gained more when being assessed.

Below are their reasons:

*Because when I'm being assessed I know where my mistake is and I will correct it.* (PE-17, Survey response)

*Because I think when being assessed, I got accurate and more guidance to make corrections in order to produce a good research paper.* (PE-21, Survey response)

*Because, the other members can tell me the shortcoming of my paper that sometime I could not realise some of the mistakes such as grammar. By being assessed, actually I've been corrected and criticised in order to develop and improve my paper.* (PE-26, Survey response)

*We will know where our errors are.* (PE-21, Interview)

*Because we know exactly what our weaknesses are and from there we want to take further step to improve so we do not want to do the same mistake in the future. (PE-12, Interview)*

These responses, however did not show strong evidence that the activities benefitted the students' development of critical thinking. The students related it to the immediate advantage they gained to improve the term paper. They valued the comments received and being directed to the errors they made in the drafts. They did not mention the thinking involved to decide on the suggestions given and the process taken to make all the corrections. Very few actually indicated these, for example:

*With the comments, I tried to develop by comparing others' work as markup level. (PE-13, Survey response)*

*I could get many ideas and combined them with my own ideas. (PE-24, Survey response)*

Students PE-13 and PE-24 reported that what they did with the comments and ideas given by their peers when their work was checked was to analyse the different ideas, to use them for making comparisons with others' work, and to synthesise the ideas to produce a better academic writing piece.

Student PE-27 declared that she did not benefit from assessing others' work due to her lack of skill, and therefore lack of confidence to evaluate. She wrote:

*I'm not confident and sure about evaluating others and the criteria that should be possessed. And I don't think I am eligible enough to evaluate others. (PE-27, Survey response)*



It seemed that this student's perceived inability to offer much help to her peers hindered her from benefitting much when assessing her peers particularly on her development of critical thinking. Student PE-27 did not see other ways how she could benefit from assessing a peer's work, for example, learning from a peer's strengths and weaknesses. This could be one explanation why the other students above also did not see they could benefit from assessing others as they were not confident with their ability to suggest improvement to their peers' drafts.

Only three out of 27 students claimed their critical thinking improved equally when assessing and being assessed. Two of them, student PE-2 and PE-20 commented:

*From both ways, we can think more. [When] people assessed me, I thought more from the assessment. When I assessed others, I would think again. So it also helped me to develop my thinking.* (PE-2, Survey response)

*When assessing, I learned mistakes made by others. When being assessed, I could be corrected and develop my skill to research new things.* (PE-20, Survey response)

From these explanations, when assessing others the students acknowledged learning from the errors made by peers especially after identifying the errors and reminding themselves not to make the same errors in their own work. Similarly, when being assessed they were made to give a deeper thought to the errors they made. This was especially experienced by those who received contradictory feedback from different peers and even between that from peers and the instructor. The discrepancies in the comments provoked greater critical thinking before a solution could be made.

### 6.2.2 The Peer Review Group

Unlike the peer evaluation group, approximately the same number of students in the peer review group claimed benefitting from reviewing others and being reviewed by others. When analysing their reasons for proclaiming that they improved their critical thinking when reviewing, two students attributed it to the better contribution they could offer to their peers especially due to their better English language skills compared to other peers in the group. These students, students PR-11 and PR-23 only saw that they could develop their thinking when they could see good examples. They stated:

*Yeah. Not being assessed because it depends - not that I'm excellent in English but certainly they were – I mean a bit lower in the knowledge of English than me that's why I was like at a disadvantage on that point. (PR-11, Interview)*

*In this particular peer review [activity], those reviewing my work are not up to my standard. At least when I review, I know what is good, and can pin-point it and perhaps use it for my own benefit, thus it changes the way I think. (PR-23, Survey response)*

Others argued:

*I can see my friends' writing and compare theirs with mine. I can ask the reason why we differ from each other, and ask for explanations, plus give suggestions. From the discussion, I can apply it to improve my paper. (PR-18, Survey response)*

*Because when I review others' work I will learn something and with that I will see what I have done and what are the mistakes in my writing. (PR-25, Survey response)*

*It makes me think more about what the subject matter of the essay of my friends is. To understand what they wrote, I have to think more. To criticise them, I need to use the knowledge/guidelines given, so that it will help them more in the writing. (PR-21, Survey response)*

Students PR-18 and PR-21 reported reflecting on their own term paper when looking at their peers' work, thinking on how it could be benefitted to improve their own work especially by comparing the papers, asking for explanations and even sharing ideas together. Meanwhile the third student asserted that identifying the subject matter of her peers' essay, understanding the issues discussed and criticising them all occupied her thinking when reviewing.

A similar number of students stated that they developed their critical thinking more when their work was being reviewed. Like the peer evaluation group, some of these students associated it with the immediate improvement of their drafts. For example:

*When someone views your work, he or she may provide some comments for you to improve your work.* (PR-2, Survey response)

*You can only benefit when people find out your mistakes because sometimes you just overlooked your mistakes.* (PR-6, Interview)

*When being reviewed, I can clearly see which part that I made mistakes, and most of the comments are true.* (PR-13, Survey response)

Students PR-2, PR-6 and PR-13 alleged that they benefitted more when they received feedback from peers who helped them identify their weaknesses for them to improve.

Other student comments showed how their critical thinking skills were developing when being reviewed. They stated:

*Because I can gain ideas when people share their views on my work.* (PR-7, Survey response)

*Because when being reviewed, my friends will comment on the mistake I made. So, it will make me think about it again and to consider whether it is right or wrong. (PR-10, Survey response)*

*I can get different views and different opinions from others. It helps make my own be a better one. (PR-13, Survey response)*

*Comments from peers do help me to improve my writing skills and also develop my thinking to think in different perspectives. (PR-15, Survey response)*

The students above claimed that being reviewed encouraged greater thinking when they could compare ideas and consider suggestions given by the peers. In doing so they were encouraged to look at things in different ways.

Two other students indicated learning by both being a reviewer and reviewee. One of them, student PR-29 explained:

*This is because when reviewing, it helps me to be alert with errors such as literature, punctuations and organisation of ideas. As for when being reviewed, it helps me to realise my own mistakes such as wrongly citing the literature, unorganised ideas and also I have problem with simplifying ideas. Peer review really helps me to improve my academic writing. (PR-29, Survey response)*

Student PR-29 clarified that when reviewing she was made more vigilant to not make the same errors as her peers did. When being reviewed, she came to realise the different kinds of errors she had produced in her own draft.

### 6.2.3 Conclusion

Three key points emerged from the analysis of the peer evaluation and peer review groups' perceptions of benefitting either from assessing/reviewing, being assessed/reviewed or both. These points are presented and discussed below.

Point 1: Those who benefitted from assessing/reviewing (several from the peer evaluation group and half of the peer review group) claimed that it was the act of trying to understand and find the strengths and weaknesses of peers' drafts to comment on that had helped foster critical thinking thus improving their own drafts. On the other hand, those who declared benefitting more from being assessed/reviewed (the majority of the peer evaluation group and half of the peer review group) connected it with the immediate advantage of being able to refine their drafts. The feedback received enabled them to make adjustments and corrections to their work so that a better paper was produced. To do this involved giving some conscious thought to the suggestions given by the peers before deciding what was best for their paper.

Point 2: The students' perceived competence to assess/review others' work influenced their claim of benefitting either from assessing/reviewing or being assessed/reviewed. Some of the students who argued gaining more from being assessed/reviewed were concerned with the limited contribution they could offer to the peers due to their incompetence and credibility to assess/review. However, some others who regarded themselves as good writers claimed that they could only benefit if they reviewed good drafts that could offer them something new to learn. These students also could have doubts about the feedback they received from peers who were less able.

Point 3: Different feedback and ratings received from peers and the course instructor could trigger deeper thinking and problem-solving. One challenge of collaborative learning is the caution of ‘the blind leading the blind’. There is a danger of students misleading each other in the process. However, the findings of this study suggest that the fact that the students were talking and thinking helped provoke critical thinking. In other words, although the accuracy of the peers’ feedback could be questioned or differed from the instructor’s feedback, it was compensated for by the potential to produce critical assessors/reviewers and writers. In this respect, the incongruent comments received actually benefitted the students’ critical thinking development.

The three main points above led to a conclusion that both being the assessor/reviewer and being the assessee/reviewee helped promote the students’ critical thinking. Although some students argued that they could only benefit from checking good drafts (i.e. those who claimed they were good writers) or by being reviewed (i.e. those who were not confident with their competence to assess others’ work), this study showed that their critical thinking was actually triggered as soon as they started to read and analyse the work looking for the strengths and flaws, compare and contrast the drafts, consider the suggestions given and decide what needed to be done in order to improve their work and others’ work.

### **6.3 Students’ and the Instructors’ Perceptions of the Peer Evaluation Activity Using the CAWAR to Foster Students’ Critical Thinking Skills in Academic Writing**

Despite some problems faced, generally peer evaluation activity was well accepted by the students and the course instructors. Students PE-17 and PE-13 said:

*It’s really helpful.* (PE-17, Interview)

*Peer evaluation, I think that it is good. Peer evaluation means that we can evaluate our friends' work in terms of the language and the entire essay. (PE-13, Interview)*

The students then clarified why they thought the peer evaluation activity was helpful. Students PE-12 and PE-17 explained:

*It's something new. Firstly, it helps us learn how to develop an essay and we can share knowledge. We can also correct the errors. (PE-12, Interview)*

*I got to know my weaknesses. For example, my sentences were fine but my peers refined them and then I also corrected theirs. It's giving and taking. So, peer evaluation is really good. We can get our errors corrected. (PE-17, Interview)*

These students valued the opportunities they got assessing and being assessed as they could contribute to each other's effort of producing a quality term paper. They were able to improve the drafts by sharing ideas and by pointing out the weaknesses of their peers' work and by having these pointed out in their own writing. The students continued:

*If I were to do it by myself my paper won't be the same. When the lecturer asked to write the first draft and then on a specific day submit it, it's normal right that students would do it at the last minute and wouldn't have the time to edit it, but with peer evaluation friends would help and half of the work was done. (PE-17, Interview)*

*So I can improve mind. If there's any mistake or if their paragraph, I mean statement or even the citation is wrong, I am able to do much of the correction before Madam (the instructor). (PE-5, Interview)*

Without peer evaluation or peer review, students had to work on their term paper drafts by themselves individually and this was not as effective as if it was done with peers' help. There was a great tendency that they would only do it at the last minute hence without much time left

to revise and edit the work. Having the peers to assess the work, on the contrary, in a way forced the students to work on schedule and get their drafts to be corrected before submitting them to the instructor.

The ability to improve the drafts was attributed to the use of the CAWAR. Students PE-21, PE-5 and PE-13 claimed:

*It's something to guide us. What aspects of the paper to evaluate are all listed (in CAWAR). My peers will explain what is wrong asking me to change this and that. (PE-21, Interview)*

*I think the form (CAWAR) has included everything. So, it helps all areas of that paper. (PE-5, Interview)*

*It helps a lot. When we receive our paper we can see whether we are at the developing stage or still mastering. (PE-13, Interview)*

The first two students highlighted that the listing of assessment criteria in the CAWAR helped them be informed of what aspects made up a quality academic writing. The students could therefore focus on these criteria when checking and discussing each other's work in order to produce a better draft. Meanwhile, the third student acknowledged the development scale as very helpful to indicate one's level of ability to fulfill the listed criteria. The three labels (i.e. *emerging*, *developing* and *mastering*) signaled where they were at in the learning continuum in an encouraging tone. When asked if the use of the CAWAR could help her remember the criteria for a quality academic writing, student PE-21 affirmed:

*God willing, I believe I will. (PE-21, Interview)*



This confirmed that exposing the criteria to the students and asking them to evaluate all the drafts using the rubric helped them to internalise the criteria.

In general, the students perceived that peer evaluation using the CAWAR as a guide was very supportive for their learning and thought it should be more widely practised. This is supported in the comments of students PE-13 and PE-12 below:

*I feel that peer evaluation needs to be continued after this because it is really helpful.*  
(PE-13, Interview)

*I think the university should implement this programme because it can benefit a lot of students doing EAW subject. They can improve their skills in terms of writing the essay as well as share the ideas so they can get a new injection.* (PE-12, Interview)

The instructors' views on the potential of the peer evaluation using the CAWAR were also sought. The instructor in charge of the peer evaluation group, Instructor-1, commented:

*It is a good activity. I could see that the students were able to comment on each other.*  
(Instructor-1, Interview)

This instructor could see how the students were made to be critical when assessing each other's drafts. This was especially enhanced when the students were asked to identify each other's strengths and weaknesses in the term paper drafts and then discuss them. The use of the CAWAR was also greatly favoured. Instructor-1 asserted:

*It would give some indications to them. Perhaps the first stage when they started using the rubric, we would see maybe in terms of assessment, maybe it's more on the left side (ratings on the developing scale), but over a period of time during the second draft and the third draft, then they would be moving more towards the right side of the way (on the*

*mastering scale) showing that there were some improvement, I think. The rubric tells them these are the things expected from them. These are the specific things that we are looking for. So it keeps them on the right track. (Instructor-1, Interview)*

With the CAWAR, the students knew if their level of performance in the first draft had improved in the subsequent drafts. It was also taken as a means for the teachers to inform the students what exactly were expected (i.e. the assessment criteria) of them.

The instructor teaching the peer review group, Instructor-2 made a similar comment:

*I think students need to know what the teachers' expectations are. They know that they have to fulfill something. I think, it is important to make students see the skill, the technique not simply the content. (Instructor-2, Interview)*

Like the first instructor, the second also emphasised on the need for students to be informed and exposed to the skills they needed to develop in academic writing. When asked about the possibility of having students peer evaluate, she urged:

*I think it won't be a problem if they grade and then tell their friends the quality of their work. Not just that, they're going to be more conscious about the quality of their own work. They are, in fact also going to understand more about their own work when they look at people's work. They will say, "Oh, ok. This person is like this. It's good. Maybe I will follow it. Do it. Ok this is not, so I'll not do it." So, in a way, they will learn. And, to give, for example an eight, I think it's a good thing because that shows that they already decide what the quality of the paper is. (Instructor-2, Interview)*

Instructor-2 could see the advantages of the peer evaluation activity to the students' learning. Not only the students were made aware of the quality of their work, they also learned by comparing their own work with others'.

The likelihood of instructors' using the peer evaluation activity and the CAWAR in their teaching of academic writing was also investigated. They stressed that:

*I think that it (the peer evaluation activity) is good. In fact, after you leave, next semester if I were to teach (the English for Academic Writing course) maybe I'm going to do the same thing. (Instructor-2, Interview)*

*I would really recommend the rubric that was designed because it is really helping me as the assessor as well, not to miss any part of the paper when I am marking. The rubric was really helpful tool, I think for both the students and the teachers. I would definitely recommend it to be used again. (Instructor-1, Interview)*

These instructors' could see the advantages that peer evaluation with the use of the CAWAR could offer to both teachers and students, expressed their willingness to adopt it in their classes, and even recommended its use.

#### **6.4 Key Findings**

This study examined the potential of the CAWAR to promote critical thinking skills development via peer evaluation activity in comparison with self-evaluation and peer review. The research questions investigated are:

1. is there evidence of greater critical thinking skills when students use the CAWAR in peer evaluation exercises to assess academic writing than in self-evaluation activities and compared to using a checklist in peer review activities?
2. do students develop their critical thinking skills better when they use the rubric to assess their peers, or when being assessed by their peers?

3. to what extent do the teachers and students perceive the peer evaluation activity using the *CAWAR* as capable of fostering students' critical thinking skills in academic writing?

To answer these questions, data from interviews, questionnaire survey, classroom observations and students' written comments on the *CAWAR* and the checklist were analysed. Compared to the responses gathered from the survey, the validity of the responses gathered from the interviews was especially examined due to the greater tendency of respondents to please the interviewer (Marx, 2008; Houtkoop-Steenstra, 2000). This was done by checking the consistencies of what was said from the beginning of the interview until the end and with what was written in their written answers to the survey. At times when their expressions did not seem to agree with what was said, deeper questions were asked and the responses given indicated they answered the questions truthfully and honestly.

The answers to each question from the qualitative materials suggested that:

Question 1: There was evidence of greater critical thinking skills when students use the *CAWAR* in peer evaluation activity to assess academic writing provided that

- a. the students take it seriously;
- b. sufficient oral discussion follows the rating task and
- c. sufficient time is allowed for the activity.

Question 2: Despite the greater number of students admitting they benefitted more when being assessed, there was no conclusive evidence that there was greater evidence of critical thinking skills indicated by these students compared to those who believed benefitting more from assessing. This was supported by the contradicting findings from the other group who also

played similar roles (i.e. one who was doing the checking and the one who was being checked) during the interventions that was the peer review group who perceived benefitting equally by being both the reviewer and reviewee.

Question 3: Both the students and the instructors of the *English for Academic Writing* course perceived the peer evaluation activity using the *CAWAR* as having a great potential to help promote critical thinking in academic writing that they recommended both the activity and the rubric to be used in academic writing classes.

The qualitative analyses generated nine claims for a more effective promotion of critical thinking skills. They are:

1. Working collaboratively to evaluate or review the work of peers enhances critical thinking;
2. The use of a rubric or a checklist enhances students' awareness of the critical thinking skills listed;
3. A rubric has the potential to involve more detailed critical thinking than the checklist because it requires students to make judgments about achievement;
4. Both a rubric and a checklist are more valuable when followed by discussion;
5. A teacher's belief in the importance of a learning activity and the learning instrument and the teacher being able to adopt the activity and the instrument are important;
6. Sufficient knowledge of a particular topic allows more contribution from a peer doing an assessment. Peers who understand the writing they are critiquing are able to comment in greater depth than peers from a different kuliyyah/faculty;

7. Sufficient time for students to discuss ratings made of their peers' work must be provided;
8. Sufficient discussion time is needed to promote the development of critical thinking; and
9. Students who understand the value of a learning activity are better engaged in the activity.

Claim 5 ("A teacher's belief in the importance of a learning activity and the learning instrument and the teacher being able to adopt the activity and the instrument are important") and Claim 7 ("Sufficient time for students to discuss ratings made of their peers' work must be provided") are both concerned with teachers' role and attitude towards the importance of the learning activities. Therefore they are merged together as one claim that is "A teacher's belief in the importance of a learning activity and the learning instrument and the teacher being able to adopt the activity and the instrument are important. This includes allowing sufficient discussion opportunities for a more effective learning".

Claim 4 ("Both a rubric and a checklist are more valuable when followed by discussion") and Claim 8 ("Sufficient discussion time is needed to promote the development of critical thinking") relate to the value of discussion which should be conducted sufficiently to enhance thinking. The two claims are combined to be "Both a rubric and a checklist are more valuable when followed by sufficient discussion". This then leaves seven claims to discuss in Chapter Seven which are:

1. Working collaboratively to evaluate or review the work of peers enhances critical thinking;
2. Both a rubric and a checklist enhance students' awareness of the critical thinking skills listed;

3. A rubric has the potential to involve more detailed critical thinking than the checklist because it requires students to make judgments about achievement;
4. Both a rubric and a checklist are more valuable when followed by sufficient discussion;
5. A teacher's belief in the importance of a learning activity and the learning instrument and the teacher being able to adopt the activity and the instrument are important. This includes allowing sufficient discussion opportunities for a more effective learning;
6. Sufficient knowledge of a particular topic allows more contribution from a peer doing an assessment. Peers who understand the writing they are critiquing are able to comment in greater depth than peers from a different kuliyyah/faculty;
7. Students who understand the value of a learning activity are better engaged in the activity.

## **CHAPTER SEVEN: DISCUSSION AND CONCLUSIONS**

### **7.0 Introduction**

The purpose of this chapter is to summarise the research findings on the potential of the Critical Thinking for Academic Writing Analytical Rubric (*CAWAR*) in peer evaluation activities to foster critical thinking skills in academic writing. The three research questions are addressed and pedagogical considerations are presented to guide greater promotion of critical thinking skills through peer evaluation.

### **7.1 Summary of the Research Findings**

The study investigated the potential of peer evaluation activities using the *CAWAR* to promote critical thinking in academic writing by comparing the level of critical thinking of the peer evaluation group with that of the control group (using the traditional teacher-only assessment method), the peer review group and the self-evaluation group before and after the-three learning interventions. Below is the summary of the findings.

#### **7.1.1 Students' Learning Performance: Findings from Quantitative Analyses**

The findings about the relative benefits of peer evaluation over peer review and self-evaluation, and in comparison with the control group that received no intervention were not conclusive. It had been hypothesised that the peer evaluation group would perform better than the peer review group which in turn would perform better than the self-evaluation group. In addition, all groups would perform better than the control group. Using the two measures of critical thinking which were the Cornell Critical Thinking Test Level X (CCTT-X) and the term paper, what was found



was that the mean gain score of the peer review group on the CCTT-X was significantly higher than the peer evaluation group and the self-evaluation group for the *credibility of assertions* skill and *inductive reasoning* skill respectively, but not on the total mean gain score of the CCTT-X. No significant differences between the peer review group and the control group, and between the peer evaluation and the control and self-evaluation groups were shown. Comparing the students' performances on the term paper, there were no significant differences found between the performances of the groups.

A further check on whether the students' learning gains benefitted their academic writing performance by examining the correlation between the four groups' post-test scores on the CCTT-X and the term paper revealed a significant relationship only for the peer review and peer evaluation groups.

Based on the above comparisons of the four groups' performances on the CCTT-X and the term paper, generally the peer review group was more effective than the other two experimental groups in the gains made on the CCTT-X.

### **7.1.2 Possible Explanations for Students' Performance: Findings from the Qualitative**

#### **Analyses**

Possible explanations for the shortcomings of the self-evaluation and peer evaluation activities were explored through the analyses of the qualitative data especially by examining the collaborative nature of the learning activities, the use of a rubric as opposed to a checklist, the implementation of the learning activities, and students' previous experience with the learning

activities and motivation to learn. The analyses generated nine claims for promoting critical thinking skills but later reduced to seven claims after merging two pairs of related claims together. The seven claims are:

1. Working collaboratively to evaluate or review the work of peers enhances critical thinking;
2. The use of a rubric or a checklist enhances students' awareness of the critical thinking skills listed;
3. A rubric has the potential to involve more detailed critical thinking than the checklist because it requires students to make judgments about achievement;
4. Both a rubric and a checklist are more valuable when followed by sufficient discussion;
5. A teacher's belief in the importance of a learning activity and the learning instrument and the teacher being able to adopt the activity and the instrument are important. This includes allowing sufficient discussion opportunities for a more effective learning;
6. Sufficient knowledge of a particular topic allows more contribution from a peer doing an assessment. Peers who understand the writing they are critiquing are able to comment in greater depth than peers from a different kuliyyah/faculty;
7. Students who understand the value of a learning activity engage better with the activity.

These are further discussed below in section 7.3.

### **7.1.3 Students' Perceptions of the Learning Activities: Findings from Quantitative and Qualitative Analyses**

Questionnaires were used to investigate students' perceptions of the influence of the learning activities on (1) the development of critical thinking skills for academic writing on the basis of the 12 assessment criteria of the *CAWAR* and (2) their learning of academic writing.

Responses to questions about the development of critical learning skills indicated that the self-evaluation group had the greatest perception of improvement followed by the peer review group and then the peer evaluation group. This finding was counter to the hypothesis that the peer evaluation group would perceive the more improvement than other groups.

Investigation on the influence of the learning activities on the students' learning of academic writing revealed that prior to the interventions, the peer evaluation group had higher regard for the peer evaluation activity than the other groups' perceptions of their respective learning activities. After the interventions, students' perceptions of the extent to which their respective learning activity actually assisted them in learning academic writing skills and improving their critical thinking skills indicated that the peer review group had the greatest belief in the learning activity they had experienced to develop their critical thinking in academic writing, followed by the peer evaluation group and then the self-evaluation group. One reason for the lower perception of the peer evaluation group of the potential of peer evaluation activity to enhance learning was due in part to one student's low confidence in assessing other students' writing.

An in depth investigation of students' and instructors' perceptions of the peer evaluation activity through interviews nonetheless revealed that they all accepted that peer evaluation using the CAWAR had a great potential to promote critical thinking in academic writing. When the limitations had been discussed, all the four groups of students believed that the peer evaluation activity was the best approach to help develop their critical thinking in academic writing. All groups recommended the future use of both the activity and the rubric in academic writing classes.

## **7.2 Current and Previous Findings Compared and Contrasted**

An important finding of the study is that the evaluation experience benefits students' learning. This finding corroborates claims by advocates of students' involvement in assessment promoting their critical faculties (e.g. Brown et al., 1994; Black et al., 2003; Kay et al., 2007; and Evans, 2008). The finding also supports the studies by Todd and Hudson (2007) and Odom et al. (2009) on the potential of the peer evaluation activity for the development of critical thinking. Similar reasons were given by students in the present study and those of Odom et al.'s (2009) study on the strengths of the activity being to allow students (a) to compare viewpoints (b) to observe different styles of writing, and (c) to clarify ideas and acknowledge what should be included in the assignment (i.e. the term paper for the current study and a research critique in Odom et al.'s study).

The study supports Johnson and Johnson's (1986) claim that collaborative learning can enhance critical thinking better than individual learning which was also found true in studies by Skon et al. (1981), Gokhale (1995), Garside (1996) and Quitadamo et al. (2009). The present study

corroborates this through the significantly better performance of the peer review group than the self-evaluation group. No significant difference was revealed however, between peer evaluation and self-evaluation groups' learning performances. The study also supports Kagan and Kagan's (2010) assertion that students with different learning abilities could benefit from working together in the collaborative learning environment. Through interviews with the peer review and peer evaluation groups, the students claimed to benefit from each other. For example, students who were good in English helped improve their peers' language use. On the other hand, students who were weaker in English helped suggest ways to clarify ideas.

Furthermore, the study also supports Nicol and MacFarlane-Dick's (2006) statement that working in groups allows discussion and feedback by peers that are important to help produce self-regulated learners. This was especially observed in the students' responses to the interviews and questionnaires. They indicated that as they were assessing others' work, they were made to reflect on their own work. They tended to compare what was written by their peers with what they themselves had produced.

In the present study, the collaborative learning activities which allowed student to share their learning was absent in the self-evaluation group and therefore did not allow an opportunity for discussion between students. These students did not have peers to comment on their work or suggest improvements. They also lacked the opportunity to practise their existing critical thinking when assessing the feedback they gathered from peers, and missed the role as feedback provider or assessor which required critical analyses of their peers' work (Falchikov, 2001). These missing features of their activity hindered the development of their critical and judgmental

skills. In this study, the *credibility of assertion* skill among the self-evaluation group in particular was less developed than the peer review group. The absence of peers to share and argue viewpoints was a possible reason. Hence, this study supports the importance of peer feedback as revealed in other related studies including Reese-Durham (2005) and Li et al. (2008). Generally, students in these studies and the present study reported that their performance could be improved through interactions with peers, as they understood better what was to be done. They also affirmed that their critical assessment skills were better exercised.

For the peer evaluation group, on the other hand, although they were expected to acknowledge their peers' academic writing performance through written comments and discussion, the students got noticeably brief written comments from their peers and therefore lacked the opportunity for discussion. This may be a possible explanation for the better performance particularly for the *inductive reasoning* skill by the peer review group who provided more elaborative written feedback on their peers' work and spent longer time for discussion. This in a way corroborates the study by Praver et al. (2011) which highlighted the importance of rich feedback to foster learning. In their study, the students who received written comments as well as a grade assessment from peers appreciated the guidance they received to improve their work and confirmed for them what they did well. The present study was in line with this as it also indicated that detailed comments rather than just numerical grades were important for promoting greater learning.

The problem of lack of discussion opportunity arose from the limited time allocated by the instructor for the activity. The peer evaluation group complained about not having sufficient time

to discuss comments and suggestions they had for each other. This confirms Romney's (1996) statement that working collaboratively is slower than traditional methods of learning and thus, agrees with earlier studies on peer evaluation where students complained that the activity is time-consuming (e.g. Topping et al. (2000), Falchikov (2001) and Odom et al. (2009). This indicates that, for a peer evaluation activity to be effective in promoting learning, including the development of critical thinking skills, it requires more time to enable students to read and understand a peer's piece of writing, rate the writing using the rubric, and provide written and oral comments.

Another important finding is that grouping of students did not satisfy the students' need for audience who could understand the arguments discussed in their term paper. Working with a wider range of people allows students to be exposed to "more elaborative thinking, more frequent giving and receiving of explanations and greater perspective taking in discussion material, all of which increase the depth of understanding, the quality of reasoning and the accuracy of long-term retention" (Johnson et al., 1994, p. 26). In this study the peer evaluation group was not comfortable having students from different faculties as peers to check their work. Since the students tended to write on a subject-related topic for their term paper, the different subject knowledge was not particularly helpful for enabling them to understand the content. This may have, in turn hindered students from being able to provide constructive comments. This finding however, is not counter to the strengths of heterogeneous groupings as advocated by Johnson et al. (1994), Murdoch and Wilson (2004), Arends and Kilcher (2010) and many others. The present study only points out that when it comes to examining academic papers discussing

subject-related issues, the peer assessors who can provide more meaningful feedback are those who share some knowledge on the subject.

No conclusive evidence was found in the study to indicate whether one benefits more as an assessor or as an assessee in terms of developing critical thinking skills, because all the students involved in the peer assessment activities experienced both roles. Li et al.'s (2010) claim that despite the literature reporting learning gains via peer assessment, there is no clear indication of how the role as assessor or assessee affects learning.

Finally, a possible explanation for the weaker performance of both the peer and self-evaluation groups when compared with the peer review group was the lack of evaluation experience among the students. Students had little experience assessing their peers' or even their own performance prior to the interventions of this study. It was not common for the students to assess much less grade their peers' and/or their own work. This finding also corroborates previous studies where students who were used to teacher evaluation found it hard to tolerate peer assessment (e.g. Cheng & Warren (2003) and Sengupta (1998)). The present study also suggests that lack of self- and peer evaluation experience might have hindered a more effective use of the *CAWAR*. Unfamiliarity with the evaluation activity could have affected the students' perceived ability to evaluate which in turn might have also affected the students' motivation or confidence to engage in the evaluation activity. This in turn would affect their use of the *CAWAR*.



### **7.3 Implications for Practice**

Critical thinking has gained broad recognition as invaluable to students not only in school but also for their future. Stone (2001) advocates that “[w]ith essential learning pointing education toward such broader goals as communication, citizenship, and thinking skills it is our obligation as teachers to make thinking something that students do every day in our classrooms and not just on assessment day” (p. 527). Due to this, various pedagogical approaches to critical thinking development have been explored to understand and find effective ways to help promote critical thinking skills among students of different backgrounds and learning contexts. Comparing and contrasting the three learning activities led to the generation of seven claims. The pedagogical implications of these claims are discussed in section 7.3.1.

#### **7.3.1 Pedagogical Considerations**

This section explores the implications of the seven pedagogical claims and discusses teaching and learning strategies that are likely to enhance students’ critical thinking skills.

##### **Claim 1: Working collaboratively to evaluate or review the work of peers enhances critical thinking**

For the development of critical thinking, the classroom activities should allow team work to solve problems and should trigger metacognition (Ryder, 1994). Peer review and peer evaluation activities cater for this through the collaborative learning environment and the formative assessment task which requires active participation from students. This is consistent with Topping’s (1998) and Ammer’s (1998) assertion that collaborative assessment by peers benefits the development of the critical thinking of both the assessor and the assessee. Activities that

allow interactions with peers rather than having students passively trying to absorb lessons by themselves are possible with the peer review and peer evaluation activities. Discussion opportunities that are initiated through the interactions need to be used to reach the most benefit it can offer. One such benefit is to help students be critical thinkers. The teacher, however, needs to support students in learning by ensuring that the students participate as effectively as possible in the activities. This is especially important as students might go off task when given the chance to interact with each other. Thus, the teacher's presence to monitor the activity is vital.

**Claim 2: The use of a rubric or a checklist enhances students' awareness of the critical thinking skills listed**

Either a rubric or a checklist can guide assessment by providing a list of all the assessment criteria. Without the list, students do not have as much direction as to what to assess. Using the assessment instrument repeatedly develops students' awareness of the skills they need to develop. This corroborates Johnson's (2001) claim that an assessment guide helps students internalise the qualities of good writing. In this study, as students have to judge the quality of a learning performance according to the criteria, they become aware of the critical thinking skills listed in the *CAWAR* and the checklist. Hence, it will be beneficial for the students if they are provided with a rubric or checklist when assessing their own or their peer's work.

**Claim 3: A rubric has the potential to involve more detailed critical thinking than the checklist because it requires students to make judgments about achievement**

The quantitative data analyses showed that using either a grading instrument or a checklist helped promote critical thinking. A rubric, however, could have triggered greater thinking than a

checklist. Depka (2007) states that, unlike a rubric, a checklist does not provide a measure of quality of performance. It is providing the measure of the quality of performance that encourages more thinking that is through the double thinking tasks students are required to do which are (1) assessing the quality of a task according to each assessment criterion listed in the rubric and (2) assigning a grade for each criterion. Compared to the single task of commenting on the quality of a task without having to grade it potentially makes it a better learning instrument. Teachers are therefore, encouraged to provide sufficient opportunities for students to use rubrics in assessing either their own or their peer's work.

**Claim 4: Both a rubric and a checklist are more valuable when followed by sufficient discussion**

Assessment opportunities by students using a rubric or checklist are beneficial for fostering the students' development of critical thinking skills. The use of the assessment guide followed by discussions will be of more advantage to students. Although holding discussions means longer time is required as stated by Romney (1996), the greater benefits that discussions are capable of offering to students made them a necessity to retain as part of the learning activities. Through discussions, the students' critical judgment skills are sharpened when they voiced out their points of view which might be questioned, argued and needed further clarifications by peers. Sufficient time for effective engagement should be allowed. The quantitative analyses comparing the learning performance of the three experimental groups indicated that the peer review group showed the greatest development of critical thinking. This was surprising given the assumption that collaborative learning together with the grading task would promote greater learning gains. In Chapter Six, the qualitative analyses revealed that the peer evaluation group did not have

sufficient time to make the expected gains. This explained the quantitative finding. We cannot know what would happen if the peer evaluation group had had more time but evidence from the qualitative analyses indicated that the learning gains would have been as great if not greater. Sufficient time would enable the careful judgment of the drafts and enough discussion opportunities for a better promotion of critical thinking. This corroborates Stone's (2001) reminder that sufficient time should be provided before students can be expected to show their best thinking. Holding an activity in a rush will only distract students' concentration and hamper opportunities for the students to expand their cognitive abilities accordingly. Due to this, it is proposed that course design should also cater for the additional time required for such learning enhancement activities like peer evaluation.

**Claim 5: A teacher's belief in the importance of a learning activity and the learning instrument and the teacher being able to adopt the activity and the instrument are important. This includes allowing sufficient discussion opportunities for a more effective learning.**

Successful teaching requires some flexibility from the teacher. Instead of being the central figure that controls what, how and how much students learn, a teacher needs to allow some space for students to decide what is best for themselves. This however, does not mean that the teacher has less work to do. Black et al. (2003) state that assessment for learning does involve extra work but the benefits are worth making the commitment. Only teachers who value the importance of a learning activity and the learning instrument used for the activity will consider adopting them and work on the strategies to use to enable students to also see the value of the activity and the learning instrument which needs further investigation.

In the case of academic writing, peer evaluation demands effort from the teacher to: identify or design a suitable rubric for students' use, guide the students in the appropriate way to use the rubric, monitor the pair or group work, adjust the balance of time in a lesson for peer evaluation activities and lecturing, and address students' uncertainties. These, however, are just a little load if the teacher's teaching goal is the learning gains for the students. Another challenge imposed on teachers if they are to adopt peer assessment activities is allowing sufficient time for discussion activities which have been identified as capable of provoking critical thinking. As effective discussions depend on the amount of time allotted, a proper division of time for learning activities needs to be planned. Teachers who appreciate the importance of discussions will devote sufficient time for the activity in order for it to be effective. Students need to be given enough time to share their ideas with their peers rather than just listening to teachers' talking.

**Claim 6: Sufficient knowledge of a particular topic allows more contribution from a peer doing an assessment. Peers who understand the writing they are critiquing are able to comment in greater depth than peers from a different kuliyyah/faculty**

The grouping of students in the peer review and the peer evaluation groups in the study to consist of three or four students per group was made on the basis of the literature. Although working in pairs could also be useful, it was not adopted because multiple gradings were sought from each student. Not only the students get more experience assessing their peers' work despite the limited opportunities that they had to assess each other's work, comments from different perspectives accelerate learning (Johnson et al., 1994). The study suggested that when peer assessment is to be adopted for use with students with different subject matter knowledge, consideration to group those with the same background might be helpful for students' learning. Liu and Carless (2006)

explain in the literature that an assessment is unreliable when the assessor does not know what he or she is trying to assess. Topping (1998) also states that an assessor utilises his or her content knowledge when assessing. Sharing similar content knowledge, students will be better able to contribute to each other's learning as they have a better understanding of what is done by the peers.

**Claim 7: Students who understand the value of a learning activity engage better with the activity**

The implementation of a particular learning activity involving an active role to be played by students might not be easily accepted in certain learning settings. Eastern world education specifically, as stated by Hassan et al. (2010) has long-practised teacher-oriented approaches to teaching which limit the contributions students can make to accelerate learning. This needs to change if the development of students' fullest potential is sought. As this study shows, students' active engagement enhances effective learning of critical thinking skills. Therefore, students can benefit from activities that can stimulate their involvement.

In order to appreciate the value of a learning activity, students need to be aware of the value of the knowledge and skills to be learned and acquired. Elder and Cosgrove (n.d.) commented that in order for a society to keep improving they need to be critical. They stated that "A critical society is a community of people who value critical thinking and value those who practice it" (para 2). Thus, a prerequisite to fostering critical thinking widely in a community would be an understanding of the value of critical thinking. Realisation of the value of critical thinking encourages its practice. The practice in turn will lead to the formation of a critical society. A

clear understanding of why critical thinking can be of value helps ease acceptance of the means to develop it. For students, the association of critical thinking and other learning is obvious. It accelerates learning. Therefore, it is a clear indication of how important it is for students to develop their critical mind in order to be successful in their learning of various subjects. As mentioned in Chapter One, critical thinking contributes to self-actualisation. For school leavers or college graduates this is important in securing a job. At present, employers are seeking people who have the potential to do more than what is shown in their academic transcripts. This means that given the choice of two candidates with similar academic performances, one who can deal with and solve problems more efficiently is preferred over the other. Such a person will help promote the development of the nation. This supports Gabennesch's (2006) assertion that one of the most important assets to a society is critical thinking. In fact, critical thinking is fundamental for liberating oneself which is essential for developing a democratic society. Facione (2011) claims that "given a society that does not value and cultivate critical thinking, we might reasonably expect that in time the judicial system and the economic system would collapse. And, in such a society, one that does not liberate its citizens by teaching them to think critically for themselves, it would be madness to advocate democratic forms of government" (p. 23).

Having understood the potential of critical thinking, it is timely that students' critical thinking skills be heightened. One way is through instructional change. Peer evaluation is one such mechanism indicated by this and earlier research as potentially effective in promoting critical thinking particularly due to the provision of collaborative and autonomous learning environment which allows active learning through feedback and discussion among students. However,

introducing an approach alien to the existing learning condition includes some challenges. Again, if the value of the learning activity is understood, these challenges could be dealt with. There is an interplay between understanding the value of a peer evaluation activity and dealing with the issues that might hinder effective promotion of critical thinking via peer evaluation. When the value is understood, the challenges can be more readily dealt with. On the other hand, with the challenges addressed, the importance is better accepted. To illustrate, students' motivation may be affected by the time allowed for an activity but realising the value of an activity can help motivate students to use the available time to attain their learning objectives. Similarly, if students see that they are given enough time to complete a task, there will be fewer challenges that will obstruct their acceptance of the benefits they can gain from engaging themselves in the learning activity.

The study was designed with the expectation that all the three learning activities would have been implemented in the way that would allow the potential of each intervention to be fully realised. It was hypothesised that the peer evaluation activity would result in the greatest learning gain because it presented the assessment criteria and the rating scale, enabled the students to apply the scale and make comments. The peer review activity also had the potential but not as much as the peer evaluation activity because it also presented the assessment criteria and students could make comment on the strengths and weaknesses of a peer's work and discuss them but the students did not have the opportunity to grade while self-evaluation activity, despite the potential it could offer by familiarising students with the assessment criteria, was expected to bring about the least learning gain due to the missing collaborative learning experience and discussion opportunity. The reality was, the peer evaluation activity did not offer the greatest



learning gain for students due in large part to the fact that the peer evaluation group did not have enough time for discussion, was not familiar enough with the activity, and was not comfortable working in groups with members who came from different faculties. What happened in the classroom actually helped reveal the important elements of using peer evaluation and the *CAWAR* in this kind of context before the expected learning gains can be obtained. In particular, allowing sufficient discussion time and establishing effective grouping of students who can work comfortably with each other especially from the same faculty are important. In addition, familiarising students with using not only the criteria but also applying the scale of the grading instrument like the *CAWAR* helps students to reflect on the kind of growing development expected of them. The *CAWAR*, when its criteria do not match closely to the learning objectives of a course might need to be replaced with some other criteria or have some other criteria added to it.

This study, therefore, argues that while all the activities were valuable, peer evaluation, using the *CAWAR* could be the most effective intervention provided that students appreciate the value of the activity and the learning instrument, and enough discussion time is built in the activity.

### **7.3.2 How Peer Evaluation Using the *CAWAR* Can Be Effectively Implemented**

The adjusted *CAWAR* could be used in different ways by teachers in different settings. The following points are guidelines for teachers of academic writing in promoting critical thinking using the *CAWAR* or a similar evaluation tool for peer evaluation.

1. Give proper and adequate training especially on using the rubric. This is to familiarise students with the evaluation task and the list of criteria of assessment against which they will assess each others' work.
2. Follow the evaluation task with discussions about areas of strengths and weaknesses. Through the discussions students can seek clarifications, defend their work and ask for opinions. Peers, on the other hand, can explain viewpoints and ratings made.
3. Allocate sufficient time for the evaluation and discussion activities. More time is especially required when students work in larger groups.
4. Allow for repeated use of the rubric. This can be done by using the same rubric at all drafting stages and/or asking students to assess a few peers' work at a time using the rubric. This is important to help students internalise the criteria that are required for high quality academic writing.
5. Introduce the assessment criteria incrementally. Shade off the criterion/criteria which might not be relevant for certain drafting stage. This is to avoid confusion for students when they are not assessing the particular section/s of the academic paper.
6. Retain the same group members for each drafting stage particularly when the drafts are done following sections of a term paper. Having different group members will cause difficulties in them being able to follow the arguments made in the earlier draft/s.
7. Get students with similar subject knowledge to assess each others' work especially when students from different departments learn together in the same class. These students will better understand what is written by their peers.
8. Instead of group work, pair work can be a better option in a learning setting where collaborative activities are new. Pair work is less threatening, less onerous, more focused,

and less time consuming. It allows learning activities to be more focused between two students which then lead to rich discussions about the strengths and weaknesses of the work produced. Once students are comfortable with this form of interaction, multiple peer evaluations which can be very threatening and intimidating even to students in the western world who may have experienced a wider range of learning activities, can be adopted.

#### **7.4 Contributions of the Study**

The results of this study add to the body of knowledge on using a rubric in peer evaluation activity to improve critical thinking skills in academic writing. This is essential since critical thinking is a required student learning outcome of academic writing courses.

In particular, through the study, an objective, valid and user-friendly rubric named the Critical Thinking for Academic Writing Analytical Rubric (*CAWAR*) was designed. The rubric was generated in a collaborative way, by including the views of subject matter experts and students. Empirical trials provided evidence that the final version was satisfactory in terms of content, construct, criterion-related and face validity. Despite the weaker performance of the peer evaluation and self-evaluation groups, the students in both groups reported that they had no difficulty understanding and using the *CAWAR*. They did not report any part of the *CAWAR* as difficult or irrelevant thus, helping to confirm its validity. A peer evaluation student's remark on the need for written qualitative comments which helped identify aspects of their academic writing needing improvement led to the refinement of the *CAWAR* (i.e. adding a column at the right end of the rubric to provide space for comments on each assessment criterion). The inter-

rater reliability of the *CAWAR* was also checked and found to be of moderate level. Despite the *CAWAR* serving well in this study, it could still be improved and further researched so that the *CAWAR* can not only be used as a means to trigger students' critical thinking but also as a scoring tool.

Second, this study has shown that, when used in a peer evaluation activity, the *CAWAR* was more than a scoring mechanism. The *CAWAR* triggers deeper thinking in students as they work through writing the essay. Therefore, the way it can be better used with peer evaluation to achieve this goal is also suggested in this study.

Third, the study not only offers comparisons between the potential of the traditional approach to assessment (i.e. teacher evaluation) and the alternative approaches (i.e. self-evaluation, peer review and peer evaluation) to foster critical thinking development, but also comparisons between the alternative approaches themselves (i.e. self-evaluation vs. peer review vs. peer evaluation). Comparing and contrasting the approaches generate invaluable information on what makes one approach better or worse than the others and what can be done to improve any particular approach. To be exact, peer evaluation was found to be able to foster critical thinking in academic writing and ideas on how to refine its use were made by understanding the strengths of the other approaches.

Fourth, the quasi-experimental study using the four approaches to assessment involving two instructors and four groups of students provides both quantitative and qualitative evidence of the potential of the alternative approaches to teacher assessment to promote the development of

critical thinking in academic writing. Although it does not address the popular belief of the unreliability of peer and self-evaluation, the study attends to the dearth of research comparing peer, self- and teacher rating particularly to develop critical thinking skills in academic writing.

### **7.5 Limitations of the Study**

The study is however, not without some limitations. First, the design, validity and reliability of the *CAWAR* were confirmed through affirmation by those involved in the study. Consultation with different groups of subject matter experts, teachers and students might have produced a different instrument.

Second, the results of the present study were affected by the real class setting in which the study was conducted. The researcher did not have control over the lesson plan, the frequency and the length of time permitted for the use of the *CAWAR* in the *English for Academic Writing* (LE 4000) classroom. The study did not explore how the peer evaluation activity might be built into regular teaching programmes that allow enough time for the effective use of the activity.

Third, measures of students' level of critical thinking was based on their performance on the Cornell Critical Thinking Test Level X (CCTT-X) which assesses four critical thinking skills: namely, inductive reasoning, deductive reasoning, credibility of assertions and identification of assumptions, and the LE 4000 term paper. Other test instruments which are more closely aligned to critical thinking might yield different results.

Fourth, for studies involving pre- and post-testing, a setting that allows a better focus on the test questions is required. Both the place and time of tests therefore, need to be conducive to testing. Despite this, in this study for the peer evaluation group, the post-test was conducted on the day students were to submit their term paper for evaluation by the instructor. Some students were seen to be still busy putting the term paper materials together, hence spending a shorter amount of time on the post-test. This might have had some effects on the findings of the study. The situation was however, unplanned. Due to the limited contact time with the students, the submission of the term papers and the post-test had to be done on the same day.

Fifth, the degree of students' readiness to participate in the different learning activities might have affected the results. This was especially of concern among the experimental groups since the students were not familiar with the assessment approaches. Although they were given some training on the way to use the *CAWAR*, it might not have been sufficient for the students to appreciate the value of the activities to their learning or learn how to engage in the activities effectively.

Finally, generalisability is another issue to consider. Although the groups were made of students of mixed nationalities, the majority of the students were Malay Malaysians who use English as their second language. Prior to the study, the students also had little or no experience with peer evaluation. Thus, different findings might be gathered when different students are taken as participants.

### 7.6 Recommendations for Future Research

This study supports the potential of the use of the *CAWAR* in peer evaluation activities to promote critical thinking in academic writing. However, considering the limitations of the current study, to confirm this potential to develop critical thinking requires further research. Therefore, some recommendations for future research are presented below.

First, in this study, the use of the rubric was limited to three drafting stages within half a semester. The data therefore, had to be based on students' use of the *CAWAR* in at least two of three peer evaluation sessions. This might have not been enough to trigger much critical thinking among the students. Students also might not be able to see clearly the effect of using the *CAWAR* in the peer evaluation activity on their learning performance. Due to this, it is recommended that future research allows use of the rubric over a longer period of time.

Second, similar procedures to examine the extent to which peer evaluation using the *CAWAR* can be adopted, but using different measures of critical thinking. Instead of the Cornell Critical Thinking Test, other critical thinking test instruments that can be used include Ennis-Weir Critical Thinking Essay Test, California Critical Thinking Skills Test and Watson-Glaser Critical Thinking Appraisal. This will help see if similar results are yielded.

Third, the study found one possible reason for the weaker learning performance among the peer evaluation group could be due to the limited time allotted for the post-test. It is, therefore, advised that future research involving testing should find a suitable time scheduled separately for the tests so the participants can be more attentive to answering the questions.

Fourth, having students being both the assessor and assessee at the same time did not lead to findings about which role benefitted students' critical thinking the more. Perhaps the benefit of having both roles is much greater than being just the assessor or assessee due to their complimentary nature. Thus, instead of focussing narrowly on finding out which role is better at promoting the development of critical thinking skills, future investigations might focus on finding ways to maximise learning when students undertake either role.

Fifth, to increase the generalisability of studies, future research should try to involve more students with balanced blend of nationalities, language backgrounds and those with peer evaluation experience. Studies can also be conducted comparing the eastern and western students' use of the *CAWAR* in a peer evaluation activity.

Finally, future research focusing on teachers' belief and implementation of peer evaluation can be initiated. In particular, studies on how teachers can be enabled to value collaborative work, what particular strategies that might work for them and how they can deal with the challenges of collaborative work are among the possible area to explore.

### **7.7 Concluding Remarks**

Peer evaluation matches well to the process approach to writing, formative assessment and autonomous learning which all promote the development of critical thinking skills. Thus, it fits into the outcome-based approach to education which calls for a shift in the teaching paradigm to encourage the development of specific skills including critical thinking skills. Through peer



evaluation students are provided with room to be assessors of their own learning. Costa and Kallick (1992) urge that

We must constantly remind ourselves that the ultimate purpose of evaluation is to have students become self-evaluating. If students graduate from our schools still dependent upon others to tell them when they are adequate, good, or excellent, then we've missed the whole point of what education is about. (p. 280)

A rubric can be used in a peer evaluation activity not just as a tool to promote standardisation but also to help trigger thinking during the peer evaluation activity. The *CAWAR* is a valuable tool for students to use as they develop their critical skills through academic writing.

Two substantial elements of peer evaluation are the provision of collaborative learning environment and the use of a rubric to guide assessment and thinking. It is important to note that comparisons of the three learning activities suggest that learning is enhanced if either of these elements is present. This means that they can be adopted separately according to what is best for both the teacher and students. To clarify, when a teacher chooses not to have students working collaboratively, a rubric for students to self-evaluate their own performance can be used. Likewise, students will develop critical thinking skills if they discuss their work but do not evaluate it. This study has, however, shown that students benefit most when both elements are adopted together as in peer evaluation activities.

In addition, attempts to develop complex skills such as critical thinking require more time before the expected learning outcomes can be gained. It is therefore also suggested that sufficient time is allowed to make the most of peer evaluation activities particularly when the activities are

relatively new in the curriculum. Adapting to new expectations requiring students to undertake evaluation roles and teachers to play facilitative roles takes long time to build into a culture.

The potential of peer evaluation to promote critical thinking outweighs the challenges to its use. Thus, if students' learning is the goal of instruction, overcoming the challenges is worth the effort.

This study has shown that the use of a rubric such as the *CAWAR* can foster critical thinking within an academic writing course. Further, it has provided evidence of the importance of collaborative activities within academic writing courses where students are engaged in thinking critically about the quality of their own work and that of their peers. This thesis suggests that in order to promote critical thinking teachers should consider: using the *CAWAR* or a similar rubric which will identify the skills being fostered and require the students to make judgments on the quality of the work being evaluated; giving adequate time for peer discussion of the rubric; ensuring that they and their students understand both the value of critical thinking and of the how the evaluation and discussion activities will foster the skills of critical thinking. This is particularly the case in eastern settings where students are less used to classroom discussion and evaluation activities than students in the West.

## REFERENCES

- Ab Kadir, M. A. (2007). Critical thinking: A family resemblance in conceptions. *Journal of Education and Human Development*, 1(2), 1-11.
- Acharya, C. (2003). *Outcome-based education (OBE): A new paradigm for learning*, CDTLink. 7(3), 1. Retrieved August 24, 2009, from <http://www.cdtl.nus.edu.sg/link/nov2003/ohe.htm>
- Acton, W. (1984). Some pragmatic dimensions of ESL writing tutorials. *TECFORS*, 7, 107.
- Adodo, S. O., & Agbayewa, J. O. (2011). Effect of homogenous and heterogeneous ability grouping class teaching on student's interest, attitude and achievement in integrated science. *International Journal of Psychology and Counselling*, 3(3), 48-54.
- Allen, S., & Knight, J. E. (2009). A methodology for collaboratively developing and validating an assessment rubric. *International Journal for the Scholarship of Teaching and Learning*, 3(2), 1-17.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Ammer, J. J. (1998). Peer evaluation model for enhancing writing performance of students with learning disabilities. *Reading & Writing Quarterly*, 14(3), 263-282.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing*. New Jersey: Prentice-Hall.
- Anderson, L. W., Krathwohl, D. R., & Airasian, P. W. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

- Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 13-18.
- Andrade, H. L., Wang, X., Du, Y., & Akawi, R. L. (2009). Rubric-referenced self-assessment and self-efficacy for writing. *Journal of Educational Research*, 102(4), 287-302.
- Arends, R., & Kilcher, A. (2010). *Teaching for student learning: Becoming an accomplished teacher*. New York: Routledge.
- Armstrong, S. L., & Paulson, E. J. (2008). Whither "peer review"? Terminology matters for the writing classroom. *Teaching English in the two year college*, 35(4), 398-407.
- Arthur, J., & Mc Tighe, J. (2000). *Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance*. Thousand Oaks, CA: Corwin Press.
- Asmus, P. E. (1999). *Creating rubrics: Scale construction*. Retrieved July 7, 2009, from <http://www.music.miami.edu/assessment/rubricsScale.html>
- Assessment Reform Group. (1999). *Assessment for learning*. Cambridge: University of Cambridge School of Education.
- Atherton, J. S. (2011). *Learning and teaching: Constructivism in learning*. Retrieved September 25, 2011, from <http://www.learningandteaching.info/learning/constructivism.htm>
- Bacon, D. R., Stewart, K. A., & Anderson, E. S. (2001). Methods of assigning players to teams: A review and novel approach. *Simulation & Gaming*, 32(1), 6-17.
- Badger, R., & White, G. (2000). A process genre approach to teaching writing. *ELT Journal*, 54(2), 153-160.
- Bailin, S., Case, R., Coombs, J. R., & Daniels, L. B. (1999). Conceptualising critical

- Thinking. *Journal of Curriculum Studies*, 31, 269-283.
- Barkley, E., Cross, P., & Major, C. (2005). *Collaborative learning techniques: A handbook for college faculty*. San Francisco: Jossey-Bass.
- Basheti, I. A., Ryan, G., Woulfe, J., & Bartimote-Aufflik, K. (2010). Anonymous peer assessment of medication management reviews. *American Journal of Pharmaceutical Education*, 74(5), 1-8.
- Bouzidi, L., & Jaillet, A. (2009). Can online peer assessment be trusted? *Educational Technology & Society*, 12(4), 257–268.
- Bean, J. (2001). *Engaging ideas: The professor's guide to integrating writing, critical thinking, and active learning in the classroom*. San Francisco: Jossey-Bass.
- Biggs, J. B. (1988). Approaches to learning and essay writing. In R. R. Schmeck (Ed.), *Learning strategies and learning styles*, (pp. 185-226). New York: Plenum.
- Black, B. (2008). *Critical thinking – a definition and taxonomy for Cambridge assessment: Supporting validity arguments about critical thinking assessments administered by Cambridge Assessment*. Retrieved February 23, 2009, from [www.iaea2008.cambridgeassessment.org.uk/ca/digitalAssets/164791\\_Black.pdf](http://www.iaea2008.cambridgeassessment.org.uk/ca/digitalAssets/164791_Black.pdf)
- Black, P., Harrison, C., Lee, C., Marshall, B., & William, D. (2003). *Assessment for learning: Putting it into practice*. Berkshire: Open University Press.
- Black, P., & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment*. Retrieved May 20, 2009, from <http://www.pdkintl.org/kappan/kbla9810.htm>

- Bloom, B. S. (1994). Reflections on the development and use of the taxonomy. In L. W. Anderson & L. A. Sosniak, (Eds.), *Bloom's taxonomy: A forty-year retrospective* (pp. 1-8). Chicago: University of Chicago Press.
- Bloom, B. S., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain*. New York, Longmans.
- Blumberg, P., & Weimer, M. (2009). *Developing learner-centered teaching: A practical guide for faculty*. San Francisco, CA: Jossey-Bass.
- Bostock, S. (2001). *Student peer assessment*. A commissioned article for the ILT now in the HE Academy resources database. Retrieved May 15, 2009, from [http://www.heacademy.ac.uk/assets/York/documents/resources/resourcedatabase/id422\\_student\\_peer\\_assessment.pdf](http://www.heacademy.ac.uk/assets/York/documents/resources/resourcedatabase/id422_student_peer_assessment.pdf)
- Boyer, E. L. (1983). *High school: A report on secondary education in America*. New York: Harper & Row.
- Brookfield, S. D. (1997). Assessing critical thinking. *New Directions for Adult and Continuing Education*, 75, 17-29.
- Brown, D., & Thomson, C. (2000). *Cooperative learning in New Zealand schools*. Palmerston North: Dunmore Press.
- Brown, G. (1997). *Assessing student learning in higher education*. London: Routledge.
- Brown, J. D. (2000). What is construct validity? *Shiken: JALT Testing & Evaluation SIG Newsletter*, 4 (2), 8-12.
- Brown, S., Rust, C., & Gibbs, G. (1994). *Strategies for diversifying assessment in higher education*. Oxford: Oxford Centre for Staff and Learning Development.

- Butler, D. L., & Winne, P. H. (1995). Feedback and self regulated learning: A theoretical synthesis. *Review of Educational Research*, 65(3), 245-281.
- Campbell, K. S., Mothersbaugh, D. L., Brammer, C., & Taylor, T. (2001). Peer versus self-assessment of oral business presentation performance. *Business Communication Quarterly*, 64(3), 23-42.
- Campbell, M. (2010). Learner-centered assessment: A new tool. *Journal of Research in Innovative Teaching*, 3(1), 169-179.
- Caroll, R. T. (2004). *Becoming a critical thinker: A guide for the new millennium*. Boston, MA: Pearson Custom.
- Cassidy, S. (2006). Learning style and student self-assessment skill. *Education + Training*, 48(2/3), 170-177.
- Center for Teaching, Learning & Technology at Washington State University. (n.d.). *Guide to rating critical and integrative thinking*. Retrieved June 18, 2009, from Washington State University at [https://my.wsu.edu/portal/page?\\_pageid=177,276578&\\_dad=portal&\\_schema=PORTAL](https://my.wsu.edu/portal/page?_pageid=177,276578&_dad=portal&_schema=PORTAL)
- Chang, C., Tseng, K., Chou, P., & Chen, Y. (2011). Reliability and validity of web-based portfolio peer assessment: A case study for a senior high school's students taking computer course. *Computers & Education*, 57, 1306-1316.
- Chapman, K. J., Meuter, M., Toy, D., & Wright, L. (2006). Can't we pick our own groups? The influence of group selection method on group dynamics and outcomes. *Journal of Management Education*, 30(4), 557-569.
- Chappuis, S., Stiggins, R. J., Arter, J., & Chappuis, J. (2003). *Assessment for learning: An action guide for school leaders*. Portland, Oregon: Assessment Training Institute.

- Chatterji, M. (2003). *Designing and using tool for educational assessment*. Boston: Allyn and Bacon.
- Chaves, J. F., Baker, C. M., Chaves, J. A., & Fisher, M. L. (2006). Self, peer and tutor assessments of MSN competencies using the PBL-Evaluator. *Journal of Nursing Education*, 45 (1), 25-31.
- Cheng, W., Lam, S., & Chan, J. C. (2008). When high achievers and low achievers work in the same group: The roles of group heterogeneity and processes in project-based learning. *British Journal of Educational Psychology*, 78, 205-221.
- Cheng, W., & Warren, M. (1999). Peer and teacher assessment of the oral and written tasks of a group project. *Assessment and Evaluation in Higher Education*, 24(3), 301-314.
- Cheng, W., & Warren, M. (2003). Having second thoughts: student perceptions before and after a peer assessment exercise. In P. Stimpson, P. Morris, Y. Fung & R. Carr (Eds.), *Curriculum, learning and assessment: The Hong Kong experience* (pp. 261-272). Hong Kong: Open University of Hong Kong Press.
- Cherry, K. (n.d.) *What is validity*. Retrieved 23 May, 2011, from <http://psychology.about.com/od/researchmethods/f/validity.htm>
- Chisholm, R. M. (1991). Introducing students to peer review of writing. *Writing Across the Curriculum*, 3(1), 4-19.
- Cho, K., Schunn, C. D., & Charney, D. (2006). Commenting on writing: Typology and perceived helpfulness of comments from novice peer reviews and subject matter experts. *Written Communication*, 23, 260-294.



- Cho, K., Schunn, C. D., & Wilson, R. W. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, 98(4), 891-901.
- Chu-you, C. (2008). *An investigation of peer evaluation in EFL college writing*. 25<sup>th</sup> International Conference of English Teaching and Learning. 2008 International Conference on English Instruction and Assessment. Retrieved March 17, 2009, from <http://www.ccu.edu.tw/fllcccu/2008EIA/English/C09.pdf>
- Code of Practice: Quality Assurance in Public Universities of Malaysia*. (2004). Quality Assurance Division, Higher Education Department, Ministry of Education.
- Condon, W., & Kelly-Riley, D. (2004). Assessing and teaching what we value: The relationship between college-level writing and critical thinking abilities. *Assessing Writing*, 9, 56-75.
- Constantine, M. G., & Ponterotto, J. G. (2006). Evaluating and selecting psychological measures for research purposes. In F. T. L. Leong & J. T. Austin (Eds.), *The psychology research handbook: A guide for graduate students and research assistants* (pp. 104-113). CA: Sage Publications, Inc.
- Costa, A. L., & Kallick, B. (1992). Reassessing assessment. In A. L. Costa, J. A. Bellanca, & R. Fogarty (Eds.), *If minds matter: A forward to the future*. Vol. 2 (pp. 275-280). Arlington Heights, IL: SkyLight Training & Publishing.
- Cousins, C. J., & Ross, J. A. (1993). Improving higher order thinking skill by teaching "with" the computer: A comparative study. *Journal of Research on Computing in Education*, 26(1), 94-116.
- Creswell, A. (2000). Self-monitoring in student writing: Developing learner responsibility. *ELT Journal*. 54 (3), 235-244.

- Delius, C., Gatzemeier, M., Sertcan, D., & Wünscher, K. (2000). *The story of philosophy: from antiquity to the present*. Cologne: Könnemann.
- Department of Education and Training. (2006). *Investigating contemporary pedagogical knowledge in thinking, ICT and Science in Victoria: Intertwining knowledge and practice*. Melbourne, Victoria: The Office of Learning and Teaching, Department of Education and Training.
- Depka, E. (2007). *Designing assessment for mathematics*. Thousand Oaks, California: Corwin Press.
- Dewey, J. (1938). *Experience and education*. NY: Collier.
- Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, 24, 331-350.
- Earl, L., Katz, S., & Western and Northern Canadian Protocol for Collaboration in Education (WNCPE) Assessment Team. (2006). *Rethinking classroom assessment with purpose in mind*. Retrieved May 22, 2009, from <http://www.wncpe.ca/media/40539/rethink.pdf>
- Elander, J., Katherine, H., Norton, L., Robinson, H., & Reddy, P. (2006). Complex skills and academic writing: a review of evidence about the types of learning required to meet core assessment criteria. *Assessment and Evaluation in Higher Education*, 31(1), 71-90.
- Elder, L., & Cosgrove, R. (n.d.). Critical thinking, the educated mind, and the creation of critical societies... Thoughts from the past. Retrieved October 31, 2011, from <http://www.criticalthinking.org/pages/critical-societies-thoughts-from-the-past/762>
- Emig, J. (1971). *The composing processes of twelfth graders*. Champaign, IL: National Council of Teachers of English.

- Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educational Leadership*, (43), 44-48.
- Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In Baron, J. B. & Sternberg, R. J. (Eds.), *Teaching thinking skills: Theory and practice* (pp. 9-26). New York: W. H. Freeman and Company.
- Ennis, R. H., Millman, J., & Tomko, T. N. (1985). *Cornell Critical Thinking Test Level X and Level Z manual* (3rd ed). CA: Midwest Publications, Pacific Grove.
- Ennis, R. (1991). Critical thinking: A streamlined conception. *Teaching Philosophy*, 14, 5-25.
- Ennis, R. H., Millman, J., & Tomko, T. N. (2005). *Cornell critical thinking tests: Administration manual*. Seaside, CA: The Critical Thinking Co.
- Ennis, R. H., & Weir, E. (1985). *The Ennis-Weir critical thinking essay test*. Pacific Grove, CA: Midwest Publications.
- Entwistle, N. (2001). Promoting deep learning through teaching and assessment. In L. Suskie (Ed.), *Assessment to promote deep learning* (pp. 9 – 20). Washington, D.C.: American Association for Higher Education.
- Ertmer, P. A, Richardson, J. C., Belland, B., Camin, D., Connolly, P., Coulthard, G., ...Mong, C. (2007). Using peer feedback to enhance the quality of student online postings: An exploratory study. *Journal of Computer-Mediated Communication*, 12(2), article 4. Retrieved September 14, 2011, from <http://jcmc.indiana.edu/vol12/issue2/ertmer.html>
- Esperet, E., & Piolat, A. (1990). Production: planning and control. In G. Denhiere & J. P. Rossi (Eds.), *Texts and Texts Processing* (pp. 317-333). North – Holland Amsterdam.

- Evans, D. (2008). Reflections on peer evaluation in an English language course. *J Nurs Studies N C N J*, 7(1), 41-49.
- Facione, P. A. (1990a). *Executive summary – critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. CA: The California Academic Press.
- Facione, P. A. (1990b). *The California Critical Thinking Skills Test: College level, technical report #1. Experimental validation and content validity*. Millbrae, CA: Insight Assessment.
- Facione, P. A. (1998). *Critical thinking: What it is and why it counts*. California: Academic Press.
- Falchikov, N. (1986). Product comparison and process benefits of collaborative peer group assessment. *Assessment and Evaluation in Higher Education*, 10(3), 146-166.
- Falchikov, N. (2001). *Learning together: Peer tutoring in higher education*. New York: Routledge.
- Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70(3), 287-322.
- Farrar, E., Evans, R., & Kirk, M. (2003). The effect of ability grouping on student attitudes and achievements in Science Laboratory work. In L. P. McCoy (Ed.), *Studies in Teaching 2003 Research Digest: Research Projects Presented at Annual Research Forum* (pp. 51-55). Winston-Salem, NC: Wake Forest University.
- Ferris, D. (2008). *Response to student writing: Implications for second-language students*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

- Fisher, A. (2001). *Critical thinking: An introduction*. Cambridge: Cambridge University Press.
- Flateby, T. L. (2011). *Improving writing and thinking through assessment*. Charlotte, N.C.: Information Age Publishing.
- Flower, L., & Hayes, J. (1981). A cognitive process theory of writing. *College Composition and Communication*, 32, 365-387.
- Foundation for Critical Thinking. (2009). *A brief history of the idea of critical thinking*. Retrieved September 14, 2011, from <http://www.criticalthinking.org/aboutCT/briefHistoryCT.cfm>
- Freeman, M. (1995). Peer assessment by groups of group work. *Assessment and Evaluation in Higher Education* 20, 289-99.
- Gabennesch, H. (2006). Critical thinking: What is it good for? (In fact what is it?). *Skeptical Inquirer*, 30 (2). Retrieved October 10, 2011, from [http://www.csicop.org/si/show/critical\\_thinking\\_what\\_is\\_it\\_good\\_for\\_in\\_fact\\_what\\_is\\_it/](http://www.csicop.org/si/show/critical_thinking_what_is_it_good_for_in_fact_what_is_it/)
- Gadzella, B. M., & Masten, W. G. (1998). Relation between measures of critical thinking and learning styles. *Psychological Reports*, 83, 1248-1250.
- Galbraith, D., & Rijlaarsdam, G. (1995). Effective strategies for the teaching and learning of writing. *Learning and Instruction*, 9(2), 93-108.
- Gammil, D. M. (2006). Learning the write way. *The Reading Teacher*, 59 (8), 754-762.
- Garside, C. (1996). Look who's talking: A comparison of lecturing and group discussion teaching strategies in developing critical thinking skills. *Communication Education*, 45(3), 212-227.

- Gibson, W. & Brown, A. (2009). *Working with qualitative data*. London: Sage Publications Ltd.
- Glaser, E. (1941). *An experiment in the development of critical thinking*. New York: J. J. Little and Ives Company.
- Gliem, J. A., & Gliem, R. R. (2003). *Calculating, interpreting, and reporting Cronbach's Alpha reliability coefficient for likert-type scales*. 2003 Midwest Research to Practice Conference in Adult, Continuing, and Community Education. The Ohio State University, Columbus, OH. Retrieved August 17, 2009, from <https://scholarworks.iupui.edu/bitstream/handle/1805/344/Gliem%20&%20Gliem.pdf?sequence=1>
- Goh, C. T. (June, 1997). *Shaping our future: Thinking schools, learning nation*. Speech presented at The 7<sup>th</sup> International Conference on Thinking, The Suntec City Convention Centre Ballroom, Singapore. Retrieved February 15, 2009, from <http://www1.moe.edu.sg/speeches/1997/020697.htm>
- Gokhale, A. A. (1995). Collaborative learning enhances critical thinking. *Journal of Technology Education*, 7(1). Retrieved 29 April, 2009, from <http://scholar.lib.vt.edu/ejournals/JTE/jte-v7n1/gokhale.jte-v7n1.html>
- Goldson, O. B. (1990). *The relationship between cognitive style, critical thinking, and moral reasoning among adolescents* (Unpublished EdD dissertation). Loma Linda University, Riverside CA.
- Gopinath, C. (1999). Alternatives to instructor assessment of class participation. *Journal of Education for Business*, 75(1), 10-14.

- Graham, C., & Misanchuk, M. (2004). Computer-mediated learning groups: benefits and challenges to using groupwork in online learning environments. In T. S. Roberts (Ed.), *Online collaborative learning: theory and practice* (pp. 181–96). Hershey, PA: Information Science Publishing.
- Grant, J., & Kinney, M. (1992). Using the Delphi technique to examine the content validity of nursing diagnoses. *Nursing Diagnosis*, 3, 12–22.
- Green, C., Christopher, E., & Lam, J. (1997). Towards learner autonomy in postgraduate language studies: Self and peer-evaluation in the development of academic discussion skills. In D. Kember, *et al.* (Eds.), *Case studies of improving teaching and learning from the action learning project* (pp. 263-273). Hong Kong: Hong Kong Polytechnic University.
- Groeber, J. F. (2007). *Designing and using rubrics for reading and language arts K-6*. Thousand Oaks, CA: Corwin Press.
- Groth-Marnat, G. (2009). *Handbook of psychological assessment*. New Jersey: John Wiley & Sons Inc.
- Haaga, D. A. F. (1993). Peer review of term papers in graduate psychology courses. *Teaching of Psychology*, 20(1), 28–32.
- Hafner, J. C., & Hafner, P. M. (2003). Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. *International Journal of Science Education*, 25(12), 1509-1528.
- Hager, P., Sleet, R., Logan, P., & Hooper, M. (2003). Teaching critical thinking in undergraduate science courses. *Science & Education*, 12(3), 303-313.
- Halx, M. D., & Reybold, L. E. (2006). A pedagogy of force: Faculty perspectives of critical thinking in undergraduate students. *The Journal of General Education*, 54(4), 293-315.

- Hashim, R., & Hussein, S. (2003). *The teaching of thinking in Malaysia*. Kuala Lumpur: International Islamic University Malaysia.
- Hassan, A., Jamaludin, N. S., Sulaiman, T., & Baki, R. (2010). *Western and eastern educational philosophies*. 2010 Philosophy of Education Society of Australasia Conference Presentation. Retrieved February 25, 2011, from <http://www.pesa.org.au/papers/2010-papers/PESA%202010%20Paper%2006.pdf>
- Haswell, R. H. (2005). NCTE/CCCC's recent war on scholarship. *Written Communication*, 22(2), 198-223.
- Hatcher, D. L. (2000). *Critical Thinking: A new definition and defense*. Retrieved February 26, 2009, from [http://www.bakeru.edu/crit/literature/dlh\\_ct\\_defense.htm](http://www.bakeru.edu/crit/literature/dlh_ct_defense.htm)
- Hiltz, S. R., & Benbunan-Fich, R. (1997). *Supporting collaborative learning in synchronous learning networks*. Invited Keynote Address for the UNESCO/ Open University Symposium on Virtual Learning Environments and the role of the Teacher. Milton Keynes, England. Retrieved November 12, 2011 from <http://web.njit.edu/~hiltz/CRProject/unesco.htm>
- Hochberg Y. (1988). A sharper Bonferonni procedure for multiple tests of significance. *Biometrika*, 75 (4), 800–802.
- Houtkoop-Steenstra, H. (2000). *Interaction and the standardized survey interview*. Cambridge: Cambridge University Press.
- Hughes, A. (2003). *Testing for language teachers*. Canbridge: Cambridge University Press.
- Huitt, W. (2004). *Bloom et al. 's taxonomy of the cognitive domain*. *Educational Psychology Interactive*. Valdosta, GA: Valdosta State University. Retrieved March 26, 2009, from <http://chiron.valdosta.edu/whuitt/col/cogsys/bloom.html>



- Hyland, K., & Hyland, F. (2006). Contexts and issues in feedback on L2 writing; An introduction. In K. Hyland and F. Hyland (Eds.), *Feedback in second language writing* (pp.1-20). Cambridge: Cambridge University Press.
- Ingleton, C., Doube, L., Rogers, T., & Noble, A. (2000) *Leap into ... collaborative learning*. (2005). Centre for Learning and Professional Development (CLPD).The University of Adelaide, Australia.
- Jacobs, G. M. (2006). Issues in implementing cooperative learning. In S. G. McCafferty, G. M. Jacobs & A. C. DaSilva Iddings (Eds.), *Cooperative learning and second language teaching* (pp. 20-46). Cambridge: Cambridge University Press.
- Jacobs, H.J., Zinkgraf, S. A., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). *Testing ESL composition: A practical approach*. Rowley, MA: Newbury House.
- Jarvis, P., & Gibson, S. (1997). *The teacher practitioner and mentor*. Cheltenham: Stanley Thornes Ltd.
- Jenicek, M., & Hitchcock, D. (2005). *Evidence-based practice: Logic and critical thinking in medicine*. Atlanta, GA: AMA Bookstore.
- Johnson, D. W., Johnson, R. T., & Holubec, E. J. (1994). *Cooperative learning in the classroom*. Alexandria, Virginia: Association for Supervision and Curriculum Development.
- Johnson, D. W., Johnson, R. T., & Holubec, E. J. (1998). *Cooperation in the classroom* (7th ed.). Edina, MN: Interaction Book.
- Johnson, D. W., Johnson, R. T., & Smith, K. A. (1991). *Cooperative learning: Increasing college faculty instructional productivity* (ASHE-ERIC Higher Education Report No. 4). Washington DC: The George Washington University, School of Education and Human Development.

- Johnson, N., & Winterbottom, M. (2011). Supporting girls' motivation in science: A study of peer- and self-assessment in a girls' only class. *Educational Studies*, 37(4), 391-403.
- Johnson, R. (2001). *The next frontier of the student-centered classroom: Teaching students to recognize quality writing through the use of peer evaluation*. ERIC Doc. No. ED463813. Retrieved May 14, 2009, from [http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content\\_storage\\_01/0000019b/80/19/f8/45.pdf](http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/19/f8/45.pdf)
- Johnson, R. T., & Johnson, D. W. (1986). Action research: Cooperative learning in the science classroom. *Science and Children*, 24, 31-32.
- Jonsson, A., & Svingy, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2 (2), 130-144.
- Jurado, J. A. (2011). Group projects in interior design studio classes: Peer feedback benefits. *Journal of Family and Consumer Sciences*, 103(1), 34-34-39.
- Kagan, S. (1989). The structural approach to cooperative learning. *Educational Leadership*, 47(4), 12-15.
- Kagan, S. (1994). *Cooperative learning*. San Clemente, CA: Kagan Cooperative Learning.
- Kagan, S., & Kagan, M. (2010). *Kagan cooperative learning*. San Clemente, CA: Kagan Publishing.
- Kay, J., Li, L., & Fekete, A. (2007). *Learner reflection in student self-assessment*. Proceeding from The Ninth Australasian Computing Education Conference (ACE 2007) (pp. 81-95). Ballarat, Australia: Australian Computer Society, Inc.

- Kim, Y., & Kim, J. (2005). Teaching Korean University Writing Class: Balancing the Process and the Genre Approach. *Asian EFL Journal*, 7(2). Retrieved May 14, 2009, from [http://www.asian-efl-journal.com/June\\_05\\_yk&jk.php](http://www.asian-efl-journal.com/June_05_yk&jk.php)
- Knight, L. V., & Steinbach, T. A. (2011). Adapting peer review to an online course: An exploratory case study. *Journal of Information Technology Education*, 10, 81-100.
- Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher*, 28(2), 16-46.
- Langan, A. M., Shuker, D. M., Cullen, W. R., Penney, D., Preziosi, R. F., & Wheeler, C. P. (2008). Relationships between student characteristics and self-, peer and tutor evaluation of oral presentations. *Assessment and Evaluation in Higher Education*, 33(2), 179-190.
- Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York: Routledge.
- Lattuca, L. R. (2005). Making learning visible: Student and peer evaluation. *Journalism & Mass Communication Educator*, 60 (3), 247-251.
- Laughlin, P. R., Hatch, E. C., Silver, J. S., & Boh, L. (2006). Groups perform better than the best individuals on letters-to-numbers problems: effects of group size. *Journal of Personality and Social Psychology*, 90(4), 644-651.
- Lavelle, E. Smith, J., & O’Ryan, L. (2002). The writing approaches of secondary Students. *British Journal of Educational Psychology*, 72, 399-418.
- Learning and Teaching Scotland. (2007). *Assessment of learning*. Retrieved May 25, 2009 from <http://www.ltscotland.org.uk/assess/of/keyfeatures/index.asp>

- Lee, C. (2006). *Language for learning mathematics: Assessment for learning in practice*. Berkshire: Open University Press.
- Leki, I. (1990). Potential problems with peer responding in ESL writing classes. *CATESGL Journal*, 3, 5-17.
- Li, L., Liu, X., & Steckelberg, A. L. (2010). Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology*, 41(3), 525-536.
- Li, L., Steckelberg, A. L., & Srinivasan, S. (2008). Utilizing peer interactions to promote learning through a web-based peer assessment system. *Canadian Journal of Learning and Technology*, 34(2), 8. Retrieved 16 September, 2011, from <http://www.cjlt.ca/index.php/cjlt/article/viewArticle/497/228>
- Liftig, R. A. (1990). Feeling good about student writing: Validation in peer evaluation. *English Journal*, 79(2), 62-65.
- Lindblom-Ylänne, S., Pihlajamäki, H., & Kotkas, T. (2006). Self-, peer- and teacher-assessment of student essays. *Active Learning in Higher Education*, 7(1), 51-62.
- Liu, N., & Carless, D. (2006). Peer feedback: the learning element of peer assessment. *Teaching in Higher Education*, 11(3), 279-290.
- Lu, R., & Bol, L. (2007). A comparison of anonymous versus identifiable e-peer review on college student writing performance and the extent of critical feedback. *Journal of Interactive Online Learning*, 6(2), 100-115.
- Marin-Garcia, J. A., Miralles, C., & Marín, M. (2008). Oral presentation and assessment skills in engineering education. *International Journal of Engineering Education*, 24(5), 926-935.

- Marx, G. T. (2008). Surveys and surveillance. In F. G. Conrad & M. F. Schober (Eds.), *Envisioning the survey interview of the future* (pp.254-266). New Jersey: John Willey & Sons, Inc..
- Marzano, R. J. (1988). *Dimensions of thinking: A framework for curriculum and instruction*. Alexandria, VA: Association for Supervisio and Curriculum Development.
- Maslow, A.H. (1943). A theory of human motivation. *Psychological Review*, 50(4), 370-96.
- Mat Daud, N. Abu Kassim, N. L., & Mat Daud, N. S. (2011). Reliability of peer and self-assessment as opposed to teacher assessment. In Nuraihan Mat Daud, Isarji Sarudin, Ainon Jariah Muhamad, Engku Haliza Engku Ibrahim & Khairiah Othman (Eds.), *Language Studies in the Muslim World* (pp. 7-24). Gombak: IIUM Press.
- Mat Daud, N., Mat Daud, N. S., & Md. Zamin, A. A. (2011). English teachers' language skills: Possible problems in the implementation of the Malaysian language policy. In Nuraihan Mat Daud, Isarji Sarudin, Ainon Jariah Muhamad, Engku Haliza Engku Ibrahim & Khairiah Othman (Eds.), *Language Studies in the Muslim World* (pp. 57-82). Gombak: IIUM Press.
- Mat Daud, N., & Hussin, Z. (2004). Developing critical thinking skills in computer-aided extended reading classes. *British Journal of Educational Technology*, 35(4), 477-487.
- Matsuda, P. K. (2003). Process and post-process: A discursive history. *Journal of Second Language Writing*, 12(1), 65-83.
- Matsuda, P. K. (2009). A Review of the book "Review of feedback in second language writing: Contexts and issues" by K. Hyland & F. Hyland (Eds.). Cambridge, UK: Cambridge University Press (2006), (pp. 261), *Journal of English for Academic Purposes*, 8 (1), 75-77.

- Matthews, R. S., Cooper, J. L., Davidson, N., & Hawkes, P. (1995). Building bridges between cooperative and collaborative learning. *Change: The Magazine of Higher Learning*, 27(4), 35-40.
- McCombs, B., & Miller, L. (2008). *The school leader's guide to learner-centered education*. Thousand Oaks, CA: Corwin Press.
- McKeachie, W. (1987). Can evaluating instruction improve teaching? *New Directions for Teaching and Learning*, 31, 3-7.
- McMillan, J. H., & Hearn, J. (2008). Student self-assessment: The key to stronger motivation and higher achievement. *Educational Horizons*, 87(1), 40-49.
- McNamara, M. J., & Deane, D. (1995). Self-assessment activities: Toward autonomy in language learning. *TESOL Journal*, 5(1), 17-21.
- Mertens, D. M. (1998). *Research methods in education and psychology: Integrating diversity with quantitative and qualitative approaches*. Thousand Oaks, CA: Sage Publications.
- Miller, L., & Ng, R. (1994). Peer assessment of oral language proficiency. *Perspectives (City University of Hong Kong)*, 6, 41-56. Retrieved January 8, 2011, from <http://sunzi1.lib.hku.hk/hkjo/article.jsp?book=10&issue=100010>
- Ministry of Education. (2005). *Key competencies in tertiary education: Developing a New Zealand framework*. Wellington: Ministry of Education.
- Ministry of Education. (2007). *The New Zealand curriculum for English-medium teaching and learning in years 1-13*. Wellington: Learning Media Limited.

- Moody, J. D., & Gifford, V.D. (1990). The effect of grouping by formal reasoning ability levels, group size, and gender on achievement in laboratory chemistry. Paper presented at the annual meeting of the Mid-South Educational Research Association, New Orleans.
- Moskal, B. M. (2000). Scoring rubrics: what, when and how? *Practical Assessment, Research & Evaluation*, 7(3). Retrieved April 21, 2009, from <http://PAREonline.net/getvn.asp?v=7&n=3>
- Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research & Evaluation*, 7(10). Retrieved October 31, 2011, from <http://PAREonline.net/getvn.asp?v=7&n=10>.
- Mowl, G., & Pain, R. (1995). Using self and peer assessment to improve students' essay writing- A case study from geography. *Innovations in Education and Training International*, 32, 324-335.
- Murdoch, K., & Wilson, J. (2004). *How to succeed with cooperative learning*. Carlton South, Victoria: Curriculum Cooperation.
- Murdoch, K., & Wilson, J. (2008). *Creating a learner-centred primary classroom*. Oxon: Routledge.
- Nanna, M. J., & Sawilowsky, S. S. (1998). Analysis of likert scale data in disability and medical rehabilitation research. *Psychological Methods*, 3, 55-67.
- Nelson, C. E. (1994). Critical thinking and collaborative learning. *New Directions for Teaching and Learning*, 59, 45-58.
- Newstead, S., & Dennie, I. (1990). Blind marking and sex bias in student assessment. *Assessment and Evaluation in Higher Education*, 15, 132-139.

- Nicol, D. J. & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218.
- North Carolina Curriculum Team. (1994). The Marzano model for critical thinking skills. Retrieved March 19, 2009 from <http://www.wcu.edu/ceap/houghton/learner/Think94/homeNCthink94.html>
- Odom, S., Glenn, B., Sanner, S., & Cannella, K. A. S. (2009). Group peer review as an active learning strategy in a research course. *International Journal of Teaching and Learning in Higher Education*, 21(1), 108-117.
- Olson, C. B. (1992). *Thinking writing: Fostering critical thinking through writing*. New York: Harper Collins Publishers.
- Omelycheva, M. (2004). *Self and peer evaluation in undergraduate education: Are promises worth risking the perils?* Paper presented at the annual meeting of the APSA Teaching and Learning Conference. NA, Wasington, DC.
- Onwuegbuzie, A. J., Jiao, Q. G., & Bostick, S. L. (2004). *Library anxiety: Theory, research and applications*. Maryland:Scarecrow Press inc.
- Oosterhof, A. (2009). *Developing and using classroom assessments*. Upper Saddle River, New Jersey: Pearson Education, Inc.
- Orsmond, P, Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education*, 25(1), 23-38.
- Oxford, R. L. (1997). Cooperative learning, collaborative learning, and interaction: Three Communicative strands in the language classroom. *The Modern Language Journal*, 81(4), 443-456.



- Paul, R., & Elder, L. (2005). Critical thinking... and the art of substantive writing, (Part 1). *Journal of Developmental Education*, 29(1), 40-41.
- Peirce, W. (2006). Designing rubrics for assessing higher order thinking. Association of Faculties for Advancement of Community College Teaching, Howard University. Retrieved June 9, 2009 from <http://academic.pg.cc.md.us/~wpeirce/MCCCTR/Designingrubricsassessingthinking.html>
- Pintrich, P. R., & Zusho, A. (2002). Student motivation and self-regulated learning in the college classroom. In J. C. Smart & W.G. Tierney (Eds.), *Higher Education: Handbook of theory and research* (pp. 731–810). New York: Agathon Press.
- Potosky, D., & Duck, J. (2007). Forming teams for classroom projects. *Developments in Business Simulation and Experiential Learning*, 34, 144-148.
- Praver, M., Rouault, G., & Eidswick, J. (2011). Attitudes and affect toward peer evaluation in EFL reading circles. *The Reading Matrix*, 11(2), 89-101.
- Presseisen, B. Z. (2001). Thinking skills: Meanings and models revisited. In Costa, A. L. (Ed.), *Developing Minds: A resource book for teaching thinking* (pp. 47-43). Alexandria, VA: Association for Supervision and Curriculum Development.
- Quitadamo, I.J., Brahler, C.J., & Grouch, G. (2009). Peer led team learning: A potential method for increasing critical thinking in advanced undergraduate science courses. *Science Educator*, 18(1), 29-39.
- Radin Umar R. S., Salleh, Wahid, Jamaluddin, Haslinda, & Nor Azizah. (2006). *Modul pembangunan kemahiran insaniah (soft skills) untuk institusi pengajian tinggi Malaysia*. Serdang: Penerbit Universiti Putra Malaysia.

- Rau, W., & Heyl, B. S. (1990). Humanizing the college classroom: Collaborative learning and social organization among students. *Teaching Sociology*, 18, 141-155.
- Rauch, M., & Fillenworth, C. (1993). *Learning through self-evaluation: Writing the essay answer*. *Journal of Reading*, 37(1), 54-55.
- Reese-Durham, N. (2005). Peer evaluation as an active learning technique. *Journal of Instructional Psychology*, 32, 338-346.
- Reynolds, J., & Moskovvitz, C. (2008). Calibrated peer review assignments in science courses: Are they designed to promote critical thinking? *Journal of College Science Teaching*, 38(2), 60-66.
- Rieber, L. J. (2006). Using peer review to improve student writing in business courses. *Journal of Education for Business*, 81(6), 322-26.
- Rollins, T. J. (1990). Levels of critical thinking of secondary agriculture students. *Journal of Agricultural Education*, Fall 1990, 47-53.
- Romney, C. (1996). The benefits of collaborative learning. Retrieved August 15, 2011, from <http://www.ucalgary.ca/pubs/Newsletters/Currents/Vol3.6/Benefits.html>
- Royalty, J. (1995). The generalisability of critical thinking: Paranormal beliefs versus statistical reasoning. *The Journal of Generic Psychology*, 156(4), 477-490.
- Russell, B. (2000). *Critical thinking exam at AS-level is launched*. Retrieved 9 March 2009 from <http://www.independent.co.uk/news/education/education-news/critical-thinking-exam-at-aslevel-is-launched-697761.html>

- Rust, C., Price, M., & O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria & processes. *Assessment and Evaluation in Higher Education*, 28, 147-164.
- Rutherford, A. (2001). *Introducing ANOVA and ANCOVA: A GLM Approach*. London: SAGE Publications.
- Ryan, G. J., Marshall, L. L., Porter, K., & Jia, H. (2007). Peer, professor and self-evaluation of class participation. *Active Learning in Higher Education*, 8(1), 49-61.
- Ryder, R. J. (1994). Using frames to promote critical writing. *Journal of Reading*, 38(3), 210-218.
- Saddler, B., & Adrade, H. (2004). The writing rubric. *Educational Leadership*, 62(2), 48-52.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144.
- Şahin, S. (2008). An application of peer assessment in higher education. *The Turkish Online Journal of Educational Technology – TOJET*, 7(2), 5-10.
- Saito, H., & Fujita, T. (2004). Characteristics and user acceptance of peer rating in EFL writing classrooms. *Language Teaching Research*, 8(1), 31-54.
- Saltourides, E. (2006). Discovering language strategies through self-assessment: Students becoming independent learners. Arizona working papers in SLAT, 13 55-76. Retrieved August 25, 2011, from <http://www.coh.arizona.edu/AWP/AWP13/AWP13%5BSaltourides%5D.pdf>

- Santos, J. R. A. (1999). Cronbach's alpha: A tool for assessing the reliability of scales. *Journal of Extension*, 37(2). Retrieved September 4, 2010, from <http://www.joe.org/joe/1999april/tt3.php>
- Schafersmen, S. D. (1991). An introduction to critical thinking. Retrieved March 31, 2009, from <http://smartcollegeplanning.org/wp-content/uploads/2010/03/Critical-Thinking.pdf>
- Scheele, D. S. (1975). Reality construction as a product of Delphi interaction. In H. A. Linstone & M. Turoff (Eds.), *The Delphi method: Techniques and applications* (pp. 37-71). Reading, MA: Addison-Wesley.
- Scriven, M., & Paul, R. (2004). *Defining critical thinking*. Retrieved November 28, 2005, from <http://www.criticalthinking.org/aboutCT/definingCT.shtml>
- Sengupta, S. (1998). Peer evaluation: 'I am not the teacher'. *ELT Journal*, 52 (1), 19-28.
- Shephard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.
- Shuttleworth, M. (2009). *Face validity*. Retrieved May 19, 2011, from Experiment Resources: <http://www.experiment-resources.com/face-validity.html>
- Skon, L., Johnson, D. W., & Johnson, R. T. (1981). Cooperative peer interaction versus individual competition and individualistic efforts: Effects on the acquisition of cognitive reasoning strategies. *Journal of Educational Psychology*, 73(1), 83-92.
- Slavin, R. E. (1995). *Cooperative learning: Theory, research and practice*. Boston: Allyn and Bacon
- Smith, B., & MacGregor, J. (1992). What is collaborative learning? In Goodsell, A., M. Mahler, V. Tinto, B.L.Smith, and J. MacGreger, (Eds), *Collaborative Learning: A Sourcebook for*

- Higher Education* (pp. 9–22). University Park, PA: National Center on Postsecondary Teaching, Learning and Assessment.
- Sorell, J. M., Brown, H. N., Silva, M. C., & Kohlenberg E. M. (1997). Use of writing portfolios for interdisciplinary assessment of critical thinking outcomes of nursing students. *Nurse Forum*, 32(4), 12–24.
- Spandel, V. (2001). *Creating writers through 6-trait writing assessment and instruction*. Longman: New York.
- Srimavin, W., & Darasawang, P. (2004). Developing self-assessment through journal writing. In Reinders, H., Anderson, H., Hobbs, M. & Jones-Parry, J. (Eds.), *Supporting independent learning in the 21st century* (pp. 134–140). Proceedings of the inaugural conference of the Independent Learning Association, Melbourne September. Auckland: Independent Learning Association Oceania.
- Stagliano, A. A. (2004). *Rath and Strong's six sigma advanced tools pocket guide: How to use design of experiments, analysis of variance, regression analysis and 25 other powerful tools*. New York: McGraw-Hills Companies, Inc.
- Stanton, K. (1999). Involving students in setting criteria for assessing written work. In Hinett, K., & Thomas, J. (Eds.), *Staff Guide to Self and Peer Assessment* (pp. 51–53). Oxford: Oxford Brookes University.
- Stefani, L. A. J. (1994). Peer, self and tutor assessment: Relative reliabilities. *Studies in Higher Education*, 19(1), 69–75.
- Stiggins, R. J., Arter, J. A., Chappuis J., & Chappuis, S. (2004). *Classroom assessment for student learning: Doing it right-using it well*. Portland, Oregon: Assessment Training Institute.

- Stone, R. (2001). How teachers can assess the teaching skills they are teaching. In A. L. Costa (Ed.), *Developing minds: A resource book for teaching thinking* (pp.525-527). Alexandria, VA: Association for Supervision and Curriculum Development.
- Strijbos, J., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction*, 20(4), 265-269.
- Sung, Y., Chang, K, Chang, T., & Yu, W. (2010). How many heads are better than one? The reliability and validity of teenagers' self- and peer assessments. *Journal of Adolescence* 33, 135–145.
- Susser, B. (1994). Process approaches in ESL/EFL writing instruction. *Journal of Second Language Writing*, 3, 31-47.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. Boston: Allyn and Bacon.
- Tarnopolsky, O., & Kozhushko, S. (2007). *Teaching academic writing in English to tertiary students in Ukraine*. Retrieved May 7, 2009 from [http://www.readingmatrix.com/conference/pp/proceedings2007/tarnopolsky\\_kozhushko.pdf](http://www.readingmatrix.com/conference/pp/proceedings2007/tarnopolsky_kozhushko.pdf)
- Temple, C. (2005). Critical thinking and critical literacy. *Thinking Classroom*, 6(2), 15-20.
- The Assessment Reform Group. (1999). *Assessment for learning. Beyond the black box*. Cambridge: University of Cambridge school of Education.
- The balanced View, Research-based Information on Timely Topics (RBITT). (2002). *Ability grouping*. Westchester Institute for Human Services Research.7-11 South Broadway, White Plain,NYW10601 Vol.6, No2. July 2002. Retrieved August 15, 2011 from <http://www.sharingsuccess.org/code/bv/abilitygrouping.pdf>

- Todd, V., & Hudson, J. C. (2007). Using graded peer evaluation to improve students' writing skills, critical thinking ability, and comprehension of material in a Principles of Public Relations course. *Journal of College Teaching and Learning*, 4(10), 39-46.
- Tomlinson, B. (2003). Introduction: Are materials developing? In B. Tomlinson (Ed.), *Developing materials for language teaching* (pp. 1-11). London: Continuum.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249-276.
- Topping, K. J., Smith, E. F., Swanson, I., & Elliot, A. (2000). Formative assessment of academic writing between postgraduate students. *Assessment & Evaluation in Higher Education*, 25(2), 149-169.
- Totten, S., Sills, T., Digby, A., & Russ, P. (1991). *Cooperative learning: A guide to research*. New York: Garland
- Tribble, C. (1996). *Writing*. Oxford: Oxford University Press.
- Trochim, W. M. K. (2006). *Research methods knowledge base*. Retrieved March 3, 2009, from <http://www.socialresearchmethods.net/kb/desexper.php>
- Tsui, C. (2003). Fostering critical thinking through effective pedagogy. *The Journal of Higher Education*, 73(6), 740-763.
- van Gelder, T. J. (2005). Teaching critical thinking: some lessons from cognitive science. *College Teaching*, 53, 41-6.
- Vardi, I. (1999). *Developing critical writers at the undergraduate level: Some insights from critical thinking pedagogy and linguistics*. HERDSA Annual International Conference. Melbourne, Australia.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.

Waern, Y. (1982). What makes a task difficult? *Psychological Research*, 44(4), 393-408.

Wagaman, J. (2008). *Student grouping in the classroom*. Retrieved October 20, 2011, from <http://jenniferwagaman.suite101.com/student-grouping-in-the-classroom-a68563>

Washington State University. (n.d.). *WSU's critical and integrative thinking rubric*. Retrieved June 15, 2009, from [https://my.wsu.edu/portal/page?\\_pageid=177,276578&\\_dad=portal&\\_schema=PORTAL](https://my.wsu.edu/portal/page?_pageid=177,276578&_dad=portal&_schema=PORTAL)

Webb, N. M. (2009). The teacher's role in promoting collaborative dialogue in the classroom. *British Journal of Educational Psychology*, 79(1), 1-28.

Weimer, M. (2002). *Learner-centered teaching: Five key changes to practice*. San Francisco: Jossey-Bass.

Wenji Qu, & Shuyi Yang. (2010). A peer- and self-assessment project implemented in practical group work. *Journal of Language Teaching and Research*, 1(6), 776-781.

Westervelt, L. (1998). Teaching writing using the process-oriented approach. Eric Doc no. ED420864. Retrieved May 13, 2009, from [http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content\\_storage\\_01/0000019b/80/15/96/a8.pdf](http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/15/96/a8.pdf)



- Whatley, J., Bell, F., Shaylor, J., Zaitseva, E., & Zakrzewska, D. (2005). CAB – Collaboration across borders: Peer evaluation for collaborative learning. *Proceedings of the 2005 Informing Science and Information Technology Education Joint Conference, Flagstaff, Arizona, USA*, 33-48. Retrieved September 25, 2011, from <http://proceedings.informingscience.org/InSITE2005/I04f44What.pdf>
- Wiggins, G. P. (1993). *Assessing student performance: Exploring the purpose and limits of testing*. San Francisco: Jossey-Bass.
- Willey, K., & Gardner, A. (2010). Investigating the capacity of self and peer assessment activities to engage students and promote learning. *European Journal of Engineering Education*, 35 (4), p429-443
- Wilson, L. O. (2006). *Leslie Owen Wilson's curriculum pages – Beyond Bloom – A new version of the cognitive taxonomy*. Retrieved February 23, 2009 from <http://www.uwsp.edu/education/lwilson/curric/newtaxonomy.htm>
- Wilson, M. (2006). *Rethinking rubrics in writing assessment*. Portsmouth, NH: Heinemann.
- Wragg, E. C. (1999). *An introduction to classroom observation*. London: Routledge.
- Wright, S. L., Burt, C. D. B., & Strongman, K. T. (2006). Loneliness in the workplace: Construct definition and scale development. *New Zealand Journal of Psychology*, 35(2), 59-68.
- Xiao, Y., & Lucking, R. (2008). The impact of two types of peer assessment on students' performance and satisfaction within a wiki environment. *Internet and Higher Education*, 11(3-4), 186-186-193.
- Xu, X. (2005). An integrated approach to the teaching of English writing. *Sino-US English Teaching*, 2(12), 37-41.

- Yancher, S. C., & Slife, B. D. (2003). *The contribution of theoretical psychology to critical thinking*. Paper presented at the meeting of The American Psychological Association. Toronto, Canada.
- Zaitseva, E., Bell, F., Whatley, J., & Shaylor, J. (2004). *Peer-evaluation in multi-cultural context: Language and culture issues in international collaborative project*. IADIS International Conference Cognition and Exploratory Learning in Digital Age (CELDA 2004). Lisbon, Portugal. Retrieved December 23, 2009, from [www.iadis.net/dl/final\\_uploads/200407L016.pdf](http://www.iadis.net/dl/final_uploads/200407L016.pdf).
- Zamel, V. (1983). The composing processes of advanced ESL students: six case studies. *TESOL Quarterly*, 17, 165-187.

## APPENDICES

### APPENDIX 1. CHECKLIST FOR PEER REVIEW

Writer :

Reviewer:

Read your friend's draft and provide your feedback by writing your comments on the space provided. Use the criteria listed below as a guide.

1. Clearly states the thesis
2. Explains difficult terms, concepts, facts or/and ideas clearly
3. Properly breaks down the issue into parts for detailed analysis
4. Supports arguments well
5. Uses only reliable literature
6. Has organised ideas and/or information coherently
7. Integrates other people's ideas accurately
8. Demonstrates a clear stand on the issue
9. Concludes the essay strongly
10. Has used the appropriate academic writing style
11. Cites the literature accurately
12. Writing is free from grammatical, spelling and punctuation errors

Comments:

---

---

---

---

---

---

---

---

**APPENDIX 2. QUESTIONNAIRE**

Name:

Instructor:

Student ID no. :

Section:

SPM English Grade (Malaysians):

1. For each of the academic writing skills listed below, please tick the box that best describes how good you think your skills are.

	<b>Excellent</b>	<b>Good</b>	<b>Average</b>	<b>Poor</b>
Stating the thesis/argument	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Explaining difficult terms, concepts, facts or/and ideas	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Breaking down issues for analysis	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Supporting arguments	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Using reliable literature	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Organising ideas and/ or information	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Integrating others' ideas	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Demonstrating a stand on the issue	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Concluding an essay	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Using academic writing  
style

☐☐☐☐

Citing the literature

☐☐☐☐

Using correct grammar,  
spelling and punctuation

☐☐☐☐

2. Please indicate how well you think the learning activity you have experienced could assist you in learning academic writing skills and improving your critical thinking skills.

**Not at all**

**Small extent**

**Moderate extent**

**Large extent**

☐☐☐☐

3. When do you think you benefit more from, when assessing or when being assessed?

when assessing

☐

when being assessed

☐

Why?

---

---

---

---

---

---

---

---

---

---

### APPENDIX 3. INTERVIEW QUESTIONS

#### Questions for Teachers:

##### A. Teacher A

1. What have you 'taken' from the experience?
  - a. generally about writing
  - b. about teaching writing
  - c. about assessing writing
2. What do you think about teacher evaluation vs. peer review activities?
3. In what ways does the peer review support students' critical thinking in writing?
4. What is/are the difficult thing/s about peer reviewing?
5. What do you suggest the better way for diagnosing where a writer is at in his or her writing skill?
6. Any comment on the tasks as they are currently framed?

##### B. Teacher B

1. What have you 'taken' from the experience?
  - a. generally about writing
  - b. about teaching writing
  - c. about assessing writing
2. How likely would you be to use the rubric in teaching academic writing in the future? Why?
3. Which do you think the rubric is better used for, self or peer evaluation? Why?
4. How helpful is the rubric for diagnosing where a writer is at in each of the different skills in writing?
5. What is/are the difficult thing/s about using the rubric?
6. How could the rubric be improved?
7. Any comment on the peer review activity as they are currently framed?
8. Other comments on the rubric?

**Interview Questions for Students:****A. Control Group**

1. How well-supported were you when you were working on your drafts?
2. How sufficient is your teacher's feedback to help you improve your work?
3. What is /are the difficult thing/s about relying only on your teacher's comments to improve your work- if any?
4. Do you see other ways of getting feedback to help you refine your work? What are they?

**B. Peer Review Group**

1. What did you learn when reviewing your peer's work?
2. How helpful is it for you when your peers review your work and give feedback?
3. How helpful are the peer reviewing activities for diagnosing what your strengths and weaknesses are in writing?
4. When do you think you benefit more from, when reviewing or when being reviewed? Why?
5. What is/are the difficult thing/s about peer reviewing?

**C. Self Evaluation Group**

1. What did you learn when assessing your own work?
2. How helpful do you think the rubric is in the assessment task?
3. How helpful is the rubric for diagnosing where you are at in each of the different skill in writing?
4. What is/are the difficult thing/s about self evaluating?
5. What is/are the difficult thing/s about using the rubric?

**D. Peer Evaluation Group**

1. What did you learn when assessing your peer's work?
2. How helpful do you think the rubric is in the assessment task?
3. How helpful is it for you when your peers evaluate your work and give feedback?
4. How helpful is the rubric for diagnosing where you are at in each of the different skill in

writing?

5. When do you think you benefit more from, when assessing or when being assessed? Why?
6. What is/are the difficult thing/s about peer evaluating?
7. What is/are the difficult thing/s about using the rubric?



## APPENDIX 4. SAMPLES OF TRANSFORMATIONS OF INTERVIEW TRANSCRIPTIONS

### 1. English to English

ORIGINAL	REFINED
That is very helpful. And also when if we... if there are other people evaluate ourselves, it will helpful because we can... we can not only just err... look at about ourselves, but we can compare about others' work...so we can know which one is better, and then we can improve it.	That (peer checking) is very helpful. And also if there are other people evaluate our [work], it will [be] helpful because we can not only just - look at our [work], but we can compare [with] others' work, so we can know which one is better, and then we can improve it.
I've learnt that we're not only learn more from other people's opinions. You know, not just see our own opinion and the view of other people, u know, is so quite important as well. It's not just for our lecturer things, we should ask other people for their views and opinions.	I've learnt that we do not only learn more from other people's opinions. You know, [we do] not just see our own opinion [but also] the view[s] of other people [which are] quite important as well. It's not just for our lecturer [to comment], we should ask other people for their views and opinions.
I can experience in... many level of writing. If let's say I need to assess three papers, so from that... that assessment, that... that writing, that... that... the marking that I've given, I can know, which type is good, which type is not good... which type is medium because... because different people will...will write, will come out with different ideas. So, from... from that assessment I can... can identify which one... which one is the best and which one is not the best. So	I can experience many levels of writing. If let's say I need to assess 3 papers, from the marking that I've given, I can know which is good, which is not good [and] which is medium because different people will write [differently and] come out with different ideas. So, from that assessment I can identify which one is the best and which one is not the best. So that later in my writing I will implement that. I will implement the best, not to implement the worse thing.

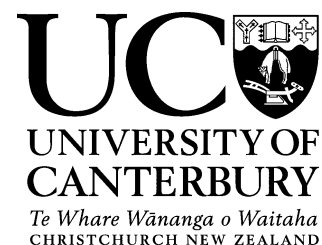
that in, later in the... in my writing I will implement that. I will implement the best, not to implement the... the worse thing.	
-----------------------------------------------------------------------------------------------------------------------------------	--

## 2. Malay to English

ORIGINAL	TRANSLATED	REFINED
Ok, saya rasa baguslah jugak sebab biasanya bila semua kita buat, kita tunggu lecturer comment. Macam tu je lah. Bila kita buat evalution group, so student kena ....err apa?.....evaluate kita punye essay. Jadi dekat situ bila kawan kita evaluate kita punye essay, kita dapat tahulah kelemahan kita... kita lemah kat mana so, dari situ kita boleh improve...err...apa yang kita lemah untuk kita cuba elakkanlah bila buat essay...err...untuk buat benda yang sama kan. So, dari situ dapat improvementlah. Dapat help saya untuk improve writing skill saya.	Ok, I think it is good because normally when we do everything, we will only then wait for the lecturer to comment. Just that. If we do group evaluation, so the students have to... err, what?... evaluate our essay, so when peers evaluate our essay, we can know our weaknesses... where our weaknesses are... so, from there we can improve... err... try to avoid doing what we are weak in when doing the essay... err... to do the same thing, so from there I can improve further... can help me improve my writing skills.	Ok, I think it is good because normally when we do everything, we will only then wait for the lecturer to comment. Just that. If we do group evaluation, other students have to- evaluate our essay, so when peers evaluate our essay, we can know our weaknesses, where our weaknesses are. From there we can improve - try to avoid doing what we are weak in when doing the essay, so from there I can improve further- can help me improve my writing skills.
<i>Dalam tu ada banyak items err, yang perlu ada kat dalam writing kan? So, kat situ kita tengok and then bila kita dapat tahu yang kita, “Oh...saya kurang</i>	<i>Listed were a lot of items err, which we needed to have in the writing, didn't we? So, we studied them and when we knew that we “Oh, I don't do</i>	<i>Listed (in the rubric) were a lot of items - which we needed to have in the writing, didn't we? So, we studied them and we knew that we, “Oh, I don't do</i>

<i>benda ni,” macam citation ke apa tak betul kat situ kan, so lepas tu kita improvelah. Kita buat try dapatkan yang excellent punya ranking dalam list ni.</i>	<i>well in this,”- like the citation. Something was wrong there, so later we corrected it. We tried to get the excellent ranking in the list.</i>	<i>well in this,” ... like the citation - something was wrong there, so later we corrected it. We tried to get the excellent ranking in the list.</i>
<i>When evaluating, saya tak tahu nak letak 4, 5. Saya baca balik. Dia ni elok ke tak elok... saya macam berfikir.</i>		<i>When evaluating, I didn't know whether to put 4, 5. I reread it, “Is she good or not?” I seemed to think.</i>

## APPENDIX 5. INFORMATION LETTER FOR THE DEAN



Prof. Dr. Nuraihan Mat Daud  
Centre for Languages and Pre-University Academy Development (CELPAD)  
International Islamic University Malaysia,  
Jalan Gombak 53100,  
Kuala Lumpur,  
Malaysia.

Date :

Dear Prof.

Your department is invited to participate in a research project entitled **DEVELOPING CRITICAL THINKING SKILLS IN ACADEMIC WRITING THROUGH THE USE OF AN INSTRUCTIONAL RUBRIC FOR PEER EVALUATION** which is a requirement for my PhD thesis under the supervision of Assoc. Prof. Alison Gilmore and Dr. Elaine Mayo from the School of Educational Studies and Human Development, College of Education, University of Canterbury, New Zealand. The aim of the project is to investigate the potential of an analytical rubric to develop students' critical thinking. Due to this, I hope to get two instructors and four groups of students taking the English for Academic Purposes (LE 4000) course to be my participants. Prior to the selection of the participants, I would like to, with your consent, gain access to information about them from the course coordinator.

As participants in this study, the instructors and students will be involved in a quasi-experimental study. Each teacher will teach two classes. One will handle the control group (no intervention) and peer review group (without rubric) and the other will teach the self evaluation group and peer evaluation group both of which will adopt the critical thinking rubric in their classes (3 times/3 periods at least). These interventions will be at the point where the students have developed certain aspects of the project paper drafts. To allow deeper insights into the activities, your permission for me to observe the classes as they work on their drafts is highly appreciated. A multiple choice test (50 minutes each – 71 questions) and a short questionnaire (10 minutes) will also be administered to all the students at the beginning and end of the semester. Later, all the instructors and some selected students from all groups will be interviewed (audio-taped) for about 15 minutes each to elicit their perception of the extent to which the different activities help

to develop students' critical thinking skills in the writing classes. The invitation for the interview among the students will be made randomly based on the group registration for the course list (e.g. every alternate fifth person in the list). Access to students' project papers marks scored by both teachers is also sought to allow comparisons.

Your department's participation in this research is completely voluntary. Therefore, you have the right to withdraw from the project at any time. If you choose to withdraw, I will use my best endeavours to remove any of the information relating to your department from the project, including any final publication, provided that this remains practically achievable.

All participants will be anonymous as names will not be used in the thesis, nor published at any time. Instead a numbering system will be used to identify one participant from another. All data will be securely stored in password protected facilities and/or locked storage at the University of Canterbury for a minimum period of 5 years following completion of the project and then destroyed.

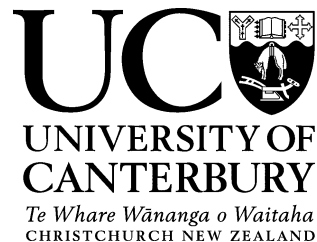
I would be happy to clarify any queries you may have in relation to this research. If you are able to participate, please complete the consent form attached and return to me in the envelope provided (or alternatively fax me at +6433437790 or e-mail me at [nsm44@student.canterbury.ac.nz](mailto:nsm44@student.canterbury.ac.nz)) by (date). Contact phone number should you wish to ring is +6433642987 Ext. 44525. Do let me know if you wish to receive a copy of the results from this study.

Yours sincerely,

Nor Shidrah Binti Mat Daud  
School of Educational Studies and Human Development  
College of Education  
University of Canterbury

This project has been received and approved by the  
University of Canterbury Educational Research Human Ethics Committee.  
Complaints may be addressed to:  
Dr Missy Morton, Chair, Educational Research Human Ethics Committee  
University of Canterbury, Private Bag 4800, CHRISTCHURCH  
Telephone: +64 345 8312  
[missy.morton@canterbury.ac.nz](mailto:missy.morton@canterbury.ac.nz)

## APPENDIX 6. CONSENT LETTER FOR THE DEAN



Tel: +64 3642987 ext. 44525, Fax: +64 3437790  
Email: [nsm44@student.canterbury.ac.nz](mailto:nsm44@student.canterbury.ac.nz)

**DEVELOPING CRITICAL THINKING SKILLS IN ACADEMIC WRITING THROUGH  
THE USE OF AN INSTRUCTIONAL RUBRIC FOR PEER EVALUATION****Declaration of Consent to Participate**

I have read and understood the description of the above named project. On this basis I agree to let the department, particularly two language instructors and four groups of students taking the English for Academic Purposes (LE 4000) course, to participate in the project. I also consent to publication of the results of the project to national or international journals or presented at educational conferences with the understanding that anonymity will be preserved.

I understand that I may at any time ask for additional information or results from the study. I also understand that I may at any time withdraw the instructors and students from the project, including withdrawal of any information they have provided relating to the department from the project, including any final publication, provided that this remains practically achievable to you.

I understand too that all data from this research will be stored securely at the University of Canterbury for five years following the study and then destroyed.

By signing below, I agree to participate in this research project.

\_\_\_\_\_ (Name)

\_\_\_\_\_ (Signature)

\_\_\_\_\_ (Date)

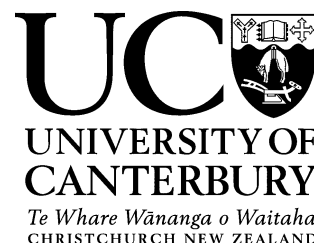
☐ By ticking this box, I will receive a report on the findings of this study and have provided my email details below for this purpose.

Email address for report on study:

*Please **return this completed consent form** in the envelope provided by (date)*

*Thank you for your department's contribution to this study.*

## APPENDIX 7. INFORMATION LETTER FOR INSTRUCTOR 1



Date :

Dear Sir/Madam

You are invited to participate in a research project entitled **DEVELOPING CRITICAL THINKING SKILLS IN ACADEMIC WRITING THROUGH THE USE OF AN INSTRUCTIONAL RUBRIC FOR PEER EVALUATION** which is a requirement for my PhD thesis under the supervision of Assoc. Prof. Alison Gilmore and Dr. Elaine Mayo from the School of Educational Studies and Human Development, College of Education, University of Canterbury, New Zealand. The aim of the project is to investigate the potential of an analytical rubric to develop students' critical thinking. Due to this, I would appreciate it very much if you are willing to be the instructor for two classes of students taking the English for Academic Purposes (LE 4000) course who will be the participants of this study.

As the instructor, your two classes will be involved in a quasi-experimental study. Both of your classes will form the self-evaluation and peer evaluation experimental groups. Your involvement in the study is expected throughout the drafting stage (three class meetings) of the course project paper. During each meeting, for the self-evaluation group, all the drafts that students prepare for the project paper you assign them to do will be self-assessed by the students using an analytical rubric that I have developed before their consultation with you. In contrast, the peer evaluation group will be asked to work in groups of three to peer evaluate each others' work using the same rubric that I have developed. Then, they will be engaged in a group discussion to give and listen to each other's feedback. This guides them in improving the work before submitting it to you.

To collect evidence on the impact of the activities, I would like to observe the classes as they work on their drafts. Students will be asked to do a multiple choice test (50 minutes -71 questions) and answer a short questionnaire (10 minutes) first at the beginning and later at the end of the semester. Later, you and some selected students will be interviewed (audio-taped) for about 15 minutes each to elicit your and their perception of the extent to which the activities help develop students' critical thinking skills in the writing classes. The invitation for the interview among the students will be made randomly based on the group registration for the course list (e.g. every alternate fifth person in the list). To see the effect of the activities on students' writing, with your permission, I would also like access to the students' project paper marks. I would also like to ask you to mark 10 extra papers from each of the other two classes involved in this study for which you will be paid accordingly.



Your participation in this research is completely voluntary. Therefore, you have the right to withdraw from the project at any time. If you choose to withdraw, I will use my best endeavours to remove any of the information relating to you from the project, including any final publication, provided that this remains practically achievable.

Your identity will be kept anonymous as names will not be used in the thesis, nor published at any time. Instead, a numbering system will be used to identify one participant from another. All data will be securely stored in password protected facilities and/or locked storage at the University of Canterbury for a minimum period of five years following completion of the project and then destroyed.

I would be happy to clarify any queries you may have in relation to this research. If you wish to participate, please complete the consent form attached and return to me in the envelope provided (or alternatively fax me at +6433437790 or e-mail me at [nsm44@student.canterbury.ac.nz](mailto:nsm44@student.canterbury.ac.nz)) by (date). My phone number should you wish to ring me is +6433642987 Ext. 44525. Do let me know if you wish to receive a copy of the results from this study.

Yours sincerely,

Nor Shidrah Binti Mat Daud  
PhD Candidate  
School of Educational Studies and Human Development  
College of Education  
University of Canterbury

This project has been received and approved by the  
University of Canterbury Educational Research Human Ethics Committee.

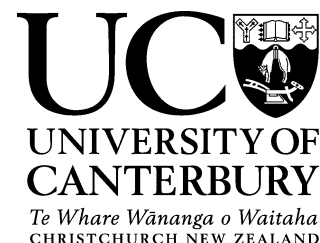
Complaints may be addressed to:

Dr Missy Morton, Chair, Educational Research Human Ethics Committee  
University of Canterbury, Private Bag 4800, CHRISTCHURCH  
Telephone: +64 345 8312  
[missy.morton@canterbury.ac.nz](mailto:missy.morton@canterbury.ac.nz)

or

Noor Amili Abdul Ghani, English for Specific Purposes (LE 4000) Course Coordinator  
Centre for Languages and Pre University Academic Development,  
International Islamic University Malaysia, Jalan Gombak,  
53100 Kuala Lumpur, Malaysia.  
Telephone: +603-61964901  
[nuraihan@iiu.edu.my](mailto:nuraihan@iiu.edu.my)

## APPENDIX 8. INFORMATION LETTER FOR INSTRUCTOR 2



Date :

Dear Sir/Madam,

You are invited to participate in a research project entitled **DEVELOPING CRITICAL THINKING SKILLS IN ACADEMIC WRITING THROUGH THE USE OF AN INSTRUCTIONAL RUBRIC FOR PEER EVALUATION** which is a requirement for my PhD thesis under the supervision of Assoc. Prof. Alison Gilmore and Dr. Elaine Mayo from the School of Educational Studies and Human Development, College of Education, University of Canterbury, New Zealand. The aim of the project is to investigate the potential of an analytical rubric to develop students' critical thinking. Due to this, I would appreciate it very much if you are willing to be the instructor for two classes of students taking the English for Academic Purposes (LE 4000) course who will be the participants of this study.

As the instructor, your two classes will be involved in a quasi-experimental study. One of the classes will be the control group (no intervention). However, the other class will form the peer review experimental group. Your involvement in the study is expected throughout the drafting stage (three class meetings) of the course project paper. During each meeting, your cooperation is required to instruct students to work in groups of three to peer review each others' drafts. Then, they will be engaged in a group discussion to give and listen to each other's feedback. This guides them in improving the work before submitting it to you.

To collect evidence on the impact of the activities, I would like to observe the classes as they work on their drafts (during the three lesson periods). Students will be asked to complete a multiple choice test (50 minutes- 71 questions) and a questionnaire (10 minutes) at the beginning and end of the semester. Later, you and some selected students will be interviewed (audio-taped) for about 15 minutes each to elicit your and their perception of the extent to which the activities help develop students' critical thinking skills in the writing classes. The invitation for the interview among the students will be made randomly based on the group registration for the course list (e.g. every alternate fifth person in the list). To see the effect of the activities on students' writing, with your permission, I would also like access to the students' project paper marks. I would also like to ask you to mark 10 extra papers from each of the other two classes involved in this study for which you will be paid accordingly.

Your participation in this research is completely voluntary. Therefore, you have the right to withdraw from the project at any time. If you choose to withdraw, I will use my best endeavours to remove any of the information relating to you from the project, including any final publication, provided that this remains practically achievable.

Your identity will be kept anonymous as names will not be used in the thesis, nor published at any time. Instead, a numbering system will be used to identify one participant from another. All data will be securely stored in password protected facilities and/or locked storage at the University of Canterbury for a minimum period of five years following completion of the project and then destroyed.

I would be happy to clarify any queries you may have in relation to this research. If you wish to participate, please complete the consent form attached and return to me in the envelope provided (or alternatively fax me at +6433437790 or e-mail me at [nsm44@student.canterbury.ac.nz](mailto:nsm44@student.canterbury.ac.nz)) by (date). My phone number should you wish to ring me is +6433642987 Ext. 44525. Do let me know if you wish to receive a copy of the results from this study.

Yours sincerely,

Nor Shidrah Binti Mat Daud  
PhD Candidate  
School of Educational Studies and Human Development  
College of Education  
University of Canterbury

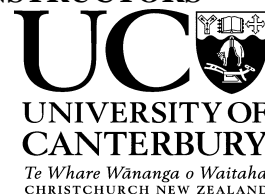
This project has been received and approved by the  
University of Canterbury Educational Research Human Ethics Committee.

Complaints may be addressed to:

Dr Missy Morton, Chair, Educational Research Human Ethics Committee  
University of Canterbury, Private Bag 4800, CHRISTCHURCH  
Telephone: +64 345 8312  
[missy.morton@canterbury.ac.nz](mailto:missy.morton@canterbury.ac.nz)

or

Noor Amili Abdul Ghani, English for Specific Purposes (LE 4000) Course Coordinator  
Centre for Languages and Pre University Academic Development,  
International Islamic University Malaysia, Jalan Gombak,  
53100 Kuala Lumpur, Malaysia.  
Telephone: +603-61964901  
[nurairhan@iiu.edu.my](mailto:nurairhan@iiu.edu.my)

**APPENDIX 9. CONSENT LETTER FOR THE INSTRUCTORS**

Tel: +64 3642987 ext. 44525, Fax: +64 3437790

Email: [nsm44@student.canterbury.ac.nz](mailto:nsm44@student.canterbury.ac.nz)

**DEVELOPING CRITICAL THINKING SKILLS IN ACADEMIC WRITING THROUGH  
THE USE OF AN INSTRUCTIONAL RUBRIC FOR PEER EVALUATION****Declaration of Consent to Participate**

I have read and understood the description of the above named project. On this basis I agree to participate as the instructor for the two groups of students taking the English for Academic Purposes (LE 4000) course who will be the participants of this study and I agree to mark the 20 other two classes project papers (10 per class) with payments and to reveal the students' marks with the students' consent. I also consent to you observing the activities and later to the publication of the results of the project to national or international journals or presented at educational conferences with the understanding that anonymity will be preserved.

I understand that I may at any time ask for additional information or results from the study. I also understand that I may at any time withdraw from the project, including withdrawal of any information I have provided and any final publication, provided that this remains practically achievable to you. I understand too that all data from this research will be stored securely at the University of Canterbury for five years following the study and then destroyed.

By signing below, I agree to participate in this research project.

\_\_\_\_\_ (Name)

\_\_\_\_\_ (Signature)

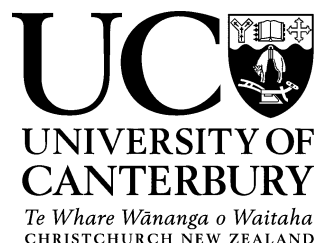
\_\_\_\_\_ (Date)

☐ By ticking this box, I will receive a report on the findings of this study and have provided my email details below for this purpose.

Email address for report on study:

*Please return this completed consent form in the envelope provided by (date)*

*Thank you for your contribution to this study.*

**APPENDIX 10. INFORMATION LETTER FOR CONTROL GROUP**

Date :

Dear Student,

You are invited to participate in a research project entitled **DEVELOPING CRITICAL THINKING SKILLS IN ACADEMIC WRITING THROUGH THE USE OF AN INSTRUCTIONAL RUBRIC FOR PEER EVALUATION** which is a requirement for my PhD thesis under the supervision of Assoc. Prof. Alison Gilmore and Dr. Elaine Mayo from the School of Educational Studies and Human Development, College of Education, University of Canterbury, New Zealand. The aim of the project is to investigate the potential of an analytical rubric to develop students' critical thinking. Due to this, I would appreciate it very much if you are willing to be one of the participants of the study.

As a participant, you will be asked to do a multiple choice test (50 minutes -71 questions) and answer a short questionnaire (10 minutes) first at the beginning and later at the end of the semester. Your learning activities in the classroom will not be intervened by this study. At the end of the semester you might be invited for an interview (audio-taped) for about 15 minutes to talk about your views on how useful the class activities were in developing your critical thinking skills in the course. The invitation for the interview will be made randomly based on the group registration for the course list (e.g. every alternate fifth person in the list). With your permission, I will observe the class as it works on the project paper drafts. I would also like to collect your project paper marks.

For being the participant, you will be given a copy of the rubric used in the study which is a scoring guide listing the main criteria of a good academic paper for your reference with a brief description of how it can be used. Your participation in this research is however, completely voluntary. Therefore, you have the right to withdraw from the project at any time. If you choose to withdraw, I will use my best endeavours to remove any of the information relating to you from the project, including any final publication, provided that this remains practically achievable.

Your identity will be kept anonymous as your name will not be used in the thesis, nor published at any time. Instead, a numbering system will be used to identify one participant from another. All data will be securely stored in password protected facilities and/or locked storage at the University of Canterbury for a minimum period of five years following completion of the project and then destroyed.

I would be happy to clarify any queries you may have in relation to this research. If you wish to participate, please complete the consent form attached and return to me in the envelope provided. Do let me know if you wish to receive a copy of the results from this study.

Yours sincerely,

Nor Shidrah Binti Mat Daud  
School of Educational Studies and Human Development  
College of Education  
University of Canterbury

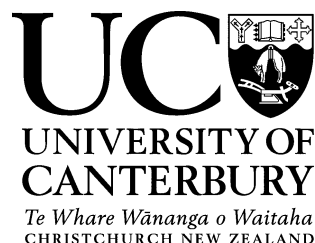
This project has been received and approved by the  
University of Canterbury Educational Research Human Ethics Committee.

Complaints may be addressed to:

Dr Missy Morton, Chair, Educational Research Human Ethics Committee  
University of Canterbury, Private Bag 4800, CHRISTCHURCH  
Telephone: +64 345 8312  
[missy.morton@canterbury.ac.nz](mailto:missy.morton@canterbury.ac.nz)

or

Noor Amili Abdul Ghani, English for Specific Purposes (LE 4000) Course Coordinator  
Centre for Languages and Pre University Academic Development,  
International Islamic University Malaysia, Jalan Gombak,  
53100 Kuala Lumpur, Malaysia.  
Telephone: +603-61964901  
[marsya@iiu.edu.my](mailto:marsya@iiu.edu.my)

**APPENDIX 11. INFORMATION LETTER FOR PEER REVIEW GROUP**

Date :

Dear Student,

You are invited to participate in a research project entitled **DEVELOPING CRITICAL THINKING SKILLS IN ACADEMIC WRITING THROUGH THE USE OF AN INSTRUCTIONAL RUBRIC FOR PEER EVALUATION** which is a requirement for my PhD thesis under the supervision of Assoc. Prof. Alison Gilmore and Dr. Elaine Mayo from the School of Educational Studies and Human Development, College of Education, University of Canterbury, New Zealand. The aim of the project is to investigate the potential of an analytical rubric to develop students' critical thinking. Due to this, I would appreciate it very much if you are willing to be one of the participants of the study.

As a participant, you will be involved in a quasi-experimental study. Together with your classmates taking the English for Academic (LE 4000) course, you will form the peer review experimental group. In this study, for the project paper that you will be assigned with, your class will be asked to work in groups of three to peer review each others' work based on a given checklist. Then, you will be engaged in a group discussion to give feedback to each others' essay and also to listen to their feedback on your essay. This study will benefit you as the activities are aimed at guiding you in improving the project paper drafts before submitting the assignment to your instructor.

With your permission, I will observe the class as it works on the drafts. You will be asked to do a multiple choice test (50 minutes -71 questions) and answer a short questionnaire (10 minutes) first at the beginning and later at the end of the semester. You might also be selected for a brief 15-minute interview (audio-taped) to talk about your views on how useful the peer review activities were in developing your critical thinking skills in the course. The invitation for the interview will be made randomly based on the group registration for the course list (e.g. every alternate fifth person in the list). I would also like to collect your project paper marks.

Your participation in this research is completely voluntary. Therefore, you have the right to withdraw from the project at any time. If you choose to withdraw, I will use my best endeavours to remove any of the information relating to you from the project, including any final publication, provided that this remains practically achievable.

Your identity will be kept anonymous as your name will not be used in the thesis, nor published at any time. Instead, a numbering system will be used to identify one participant from another. All data will be securely stored in password protected facilities and/or locked storage at the University of Canterbury for a minimum period of five years following completion of the project and then destroyed.

I would be happy to clarify any queries you may have in relation to this research. If you wish to participate, please complete the consent form attached and return to me in the envelope provided. Do let me know if you wish to receive a copy of the results from this study.

Yours sincerely,

Nor Shidrah Binti Mat Daud  
School of Educational Studies and Human Development  
College of Education  
University of Canterbury

This project has been received and approved by the  
University of Canterbury Educational Research Human Ethics Committee.

Complaints may be addressed to:

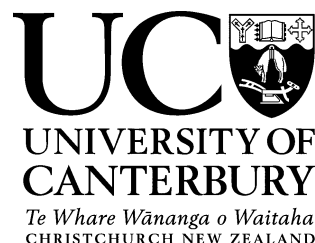
Dr Missy Morton, Chair, Educational Research Human Ethics Committee  
University of Canterbury, Private Bag 4800, CHRISTCHURCH  
Telephone: +64 345 8312  
[missy.morton@canterbury.ac.nz](mailto:missy.morton@canterbury.ac.nz)

or

Noor Amili Abdul Ghani, English for Specific Purposes (LE 4000) Course Coordinator  
Centre for Languages and Pre University Academic Development,  
International Islamic University Malaysia, Jalan Gombak,  
53100 Kuala Lumpur, Malaysia.  
Telephone: +603-61964901  
[nurairhan@iiu.edu.my](mailto:nurairhan@iiu.edu.my)



## APPENDIX 12. INFORMATION LETTER FOR SELF-EVALUATION GROUP



Date :

Dear Student,

You are invited to participate in a research project entitled **DEVELOPING CRITICAL THINKING SKILLS IN WRITING THROUGH THE USE OF AN INSTRUCTIONAL RUBRIC FOR PEER EVALUATION** which is a requirement for my PhD thesis under the supervision of Assoc. Prof. Alison Gilmore and Dr. Elaine Mayo from the School of Educational Studies and Human Development, College of Education, University of Canterbury, New Zealand. The aim of the project is to investigate the potential of an analytical rubric to develop students' critical thinking. Due to this, I would appreciate it very much if you are willing to be one of the participants of the study.

This study will benefit you as the activities are aimed at guiding you in improving the project paper drafts before submitting the assignment to your instructor. As a participant, you will be involved in a quasi-experimental study. Together with your classmates taking the English for Academic (LE 4000) course, you will form the self-evaluation experimental group. In this study, for the project paper that you will be assigned with, you will be asked to self-evaluate your work using a rubric that I have developed, before consulting your instructor.

With your permission, I will observe the class as it works on the drafts. You will be asked to do a multiple choice test (50 minutes -71 questions) and answer a short questionnaire (10 minutes) first at the beginning and later at the end of the semester. You might also be selected for a brief 15-minute interview (audio-taped) to talk about your views on how useful the self evaluation activities were in developing your critical thinking skills in the course. The invitation for the interview will be made randomly based on the group registration for the course list (e.g. every alternate fifth person in the list). I would also like to collect your project paper marks.

Your participation in this research is completely voluntary. Therefore, you have the right to withdraw from the project at any time. If you choose to withdraw, I will use my best endeavours to remove any of the information relating to you from the project, including any final publication, provided that this remains practically achievable.

Your identity will be kept anonymous as your name will not be used in the thesis, nor published at any time. Instead, a numbering system will be used to identify one participant from another. All data will be securely stored in password protected facilities and/or locked storage at the University of Canterbury for a minimum period of five years following completion of the project and then destroyed.

I would be happy to clarify any queries you may have in relation to this research. If you wish to participate, please complete the consent form attached and return to me in the envelope provided. Do let me know if you wish to receive a copy of the results from this study.

Yours sincerely,

Nor Shidrah Binti Mat Daud  
School of Educational Studies and Human Development  
College of Education  
University of Canterbury

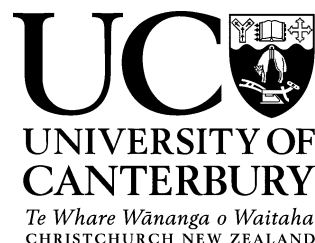
This project has been received and approved by the  
University of Canterbury Educational Research Human Ethics Committee.

Complaints may be addressed to:

Dr Missy Morton, Chair, Educational Research Human Ethics Committee  
University of Canterbury, Private Bag 4800, CHRISTCHURCH  
Telephone: +64 345 8312  
[missy.morton@canterbury.ac.nz](mailto:missy.morton@canterbury.ac.nz)

or

Noor Amili Abdul Ghani, English for Specific Purposes (LE 4000) Course Coordinator  
Centre for Languages and Pre University Academic Development,  
International Islamic University Malaysia, Jalan Gombak,  
53100 Kuala Lumpur, Malaysia.  
Telephone: +603-61964901  
[marsya@iiu.edu.my](mailto:marsya@iiu.edu.my)

**APPENDIX 13. INFORMATION LETTER FOR PEER EVALUATION GROUP**

Date:

Dear Student,

You are invited to participate in a research project entitled **DEVELOPING CRITICAL THINKING SKILLS IN WRITING THROUGH THE USE OF AN INSTRUCTIONAL RUBRIC FOR PEER EVALUATION** which is a requirement for my PhD thesis under the supervision of Assoc. Prof. Alison Gilmore and Dr. Elaine Mayo from the School of Educational Studies and Human Development, College of Education, University of Canterbury, New Zealand. The aim of the project is to investigate the potential of an analytical rubric to develop students' critical thinking. Due to this, I would appreciate it very much if you are willing to be one of the participants of the study.

As a participant, you will be involved in a quasi-experimental study. Together with your classmates taking the English for Academic (LE 4000) course, you will form the peer evaluation experimental group. In this study, for the project paper that you will be assigned with, your class will be asked to work in groups of three to peer evaluate each others' work using a rubric that I have developed. Then, you will be engaged in a group discussion to give feedback on each others' essay and also to listen to their feedback on your essay. This study will benefit you as the activities are aimed at guiding you in improving the project paper drafts before submitting the assignment to your instructor.

With your permission, I will observe the class as it works on the drafts. You will be asked to do a multiple choice test (50 minutes -71 questions) and answer a short questionnaire (10 minutes) first at the beginning and later at the end of the semester. You might also be selected for a brief 15-minute interview (audio-taped) to talk about your views on how useful the peer evaluation activities were in developing your critical thinking skills in the course. The invitation for the interview will be made randomly based on the group registration for the course list (e.g. every alternate fifth person in the list). I would also like to collect your project paper marks.

Your participation in this research is completely voluntary. Therefore, you have the right to withdraw from the project at any time. If you choose to withdraw, I will use my best endeavours to remove any of the information relating to you from the project, including any final publication, provided that this remains practically achievable.

Your identity will be kept anonymous as your name will not be used in the thesis, nor published at any time. Instead, a numbering system will be used to identify one participant from another. All data will be securely stored in password protected facilities and/or locked storage at the University of Canterbury for a minimum period of five years following completion of the project and then destroyed.

I would be happy to clarify any queries you may have in relation to this research. If you wish to participate, please complete the consent form attached and return to me in the envelope provided. Do let me know if you wish to receive a copy of the results from this study.

Yours sincerely,

Nor Shidrah Binti Mat Daud  
School of Educational Studies and Human Development  
College of Education  
University of Canterbury

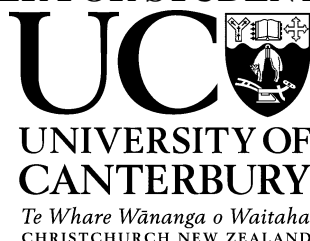
This project has been received and approved by the  
University of Canterbury Educational Research Human Ethics Committee.

Complaints may be addressed to:

Dr Missy Morton, Chair, Educational Research Human Ethics Committee  
University of Canterbury, Private Bag 4800, CHRISTCHURCH  
Telephone: +64 345 8312  
[missy.morton@canterbury.ac.nz](mailto:missy.morton@canterbury.ac.nz)

or

Noor Amili Abdul Ghani, English for Specific Purposes (LE 4000) Course Coordinator  
Centre for Languages and Pre University Academic Development,  
International Islamic University Malaysia, Jalan Gombak,  
53100 Kuala Lumpur, Malaysia.  
Telephone: +603-61964901  
[marsya@iiu.edu.my](mailto:marsya@iiu.edu.my)

**APPENDIX 14. CONSENT LETTER FOR STUDENTS**

Tel: +64 3642987 ext. 44525, Fax: +64 3437790  
Email: [nsm44@student.canterbury.ac.nz](mailto:nsm44@student.canterbury.ac.nz)

**DEVELOPING CRITICAL THINKING SKILLS IN WRITING THROUGH THE USE  
OF AN INSTRUCTIONAL RUBRIC FOR PEER EVALUATION****Declaration of Consent to Participate**

I have read and understood the description of the above named project. On this basis I agree to participate in the project, and I give permission to my LE 4000 instructor to reveal my final project score. I also consent to you observing my learning and later the publication of the results of the project to national or international journals or presented at educational conferences with the understanding that anonymity will be preserved.

I understand that I may at any time ask for additional information or results from the study. I also understand that I may at any time withdraw from the project, including withdrawal of any information I have provided and any final publication, provided that this remains practically achievable to you. I understand too that all data from this research will be stored securely at the University of Canterbury for five years following the study and then destroyed.

By signing below, I agree to participate in this research project.

\_\_\_\_\_ (Name)

\_\_\_\_\_ (Signature)

\_\_\_\_\_ (Date)

☐ By ticking this box, I will receive a report on the findings of this study and have provided my email details below for this purpose.

Email address for report on study:

*Please return this completed consent form in the envelope provided by (date)*

*Thank you for your contribution to this study.*

## APPENDIX 15. SAMPLES OF STUDENTS' WRITTEN COMMENTS ON PEERS' WORK

## 1. Peer Evaluation Group

Comments:

- Good
- improve!

with first, really to see how it flows, a, and really, and so  
then more at depth and find out more interesting

## Comments:

- Look back at the conclusion.
- Where is your stand? Are there any arguments included?
- Recommendation.
- The abstract maybe can talk more about the paper.

## 2. Peer Review Group

Comments:

\* Argument 2 & counter argument was not separated.

- Argument 2: confusing - as regards to the stages of prohibition.

- may be simplified & "stages" can be part of & rational of prohibition.

- Authorities are firm.

- Counter-argument: the <sup>arguments are</sup> relation is more focusing on Hadd punishment as a whole → not a punishment for alcoholic drinkers.

suggestions: Scope of punishment: for muslims - emphasized it more.

1. compare with punishment under civil law → difference of effect < punishment preventive measure



## Comments:

1. Thesis statement is not clearly stated, as it does not reflect the stand of the writer. It only tells us that consumption of alcohol ~~is~~ gives negative effects - which is a common knowledge.
2. Difficult terms are explained but no authority cited for that definition.
3. Issues are broken into parts, however the flow lacks. there is no flow of ideas.
4. Arguments are <sup>not</sup> supported <sup>well. Bring more logical supporting details.</sup> by ~~only one authority~~ (World Health organization).
5. <sup>and</sup> WHO is reliable, however more authorities needed.
6. Ideas & information are ~~not~~ organised coherently.
7. No comment, as no idea of other person is ~~cited~~. cited.
8. Does not demonstrate a clear stand on the issue.
10. Yes, APA style used.
11. Yes.
12. Grammatical, spelling and punctuation errors are available.  
I would suggest to the writer to be ~~are~~ really careful in this aspect.

## APPENDIX 16. THE CRITICAL THINKING IN ACADEMIC WRITING ANALYTICAL RUBRIC (CAWAR)

**Writer:****Assessor:**

Please read the essay and then rate how well each skill is demonstrated in the essay. Use the 1-6 scale from the skill 'emerging' through 'developing' to the skill being 'mastered'. Circle one number on the scale. You are welcome to use the available space to write any comments.

<div>Emerging    Developing    Mastering</div> <div><div></div></div>										Details (e.g. page reference, explanations and suggestions)
1.	Does not clearly state the thesis	1	2	3	4	5	6	Clearly states the thesis		
2.	Explains difficult terms, concepts, facts and/or ideas poorly	1	2	3	4	5	6	Explains difficult terms, concepts, facts or/and ideas clearly		
3.	Doesn't break down the issue into parts for detailed analysis	1	2	3	4	5	6	Properly breaks down the issue into parts for detailed analysis		
4.	Does not support arguments	1	2	3	4	5	6	Supports arguments well		
5.	Does not use reliable literature	1	2	3	4	5	6	Uses only reliable literature		
6.	Has not organised ideas and/or information coherently	1	2	3	4	5	6	Has organised ideas and/or information coherently		
7.	Integrates other people's ideas poorly	1	2	3	4	5	6	Integrates other people's ideas accurately		

