



**DEVELOPING A VALID AND RELIABLE ANALYTIC  
RUBRIC TO ASSESS CRITICAL THINKING IN  
ENGLISH ARGUMENTATIVE ESSAYS OF EFL  
UNIVERSITY STUDENTS**

**BY**

**MR. NATTAWUT NAKKAEW**

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE  
REQUIREMENTS FOR THE DEGREE OF MASTER OF  
ARTS IN ENGLISH LANGUAGE STUDIES  
DEPARTMENT OF ENGLISH AND LINGUISTICS  
FACULTY OF LIBERAL ARTS  
THAMMASAT UNIVERSITY  
ACADEMIC YEAR 2018  
COPYRIGHT OF THAMMASAT UNIVERSITY**

**DEVELOPING A VALID AND RELIABLE ANALYTIC  
RUBRIC TO ASSESS CRITICAL THINKING IN  
ENGLISH ARGUMENTATIVE ESSAYS OF EFL  
UNIVERSITY STUDENTS**

**BY**

**MR. NATTAWUT NAKKAEW**

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE  
REQUIREMENTS FOR THE DEGREE OF MASTER OF  
ARTS IN ENGLISH LANGUAGE STUDIES  
DEPARTMENT OF ENGLISH AND LINGUISTICS  
FACULTY OF LIBERAL ARTS  
THAMMASAT UNIVERSITY  
ACADEMIC YEAR 2018  
COPYRIGHT OF THAMMASAT UNIVERS**

THAMMASAT UNIVERSITY  
FACULTY OF LIBERAL ARTS

THESIS

BY

MR. NATTAWUT NAKKAEW

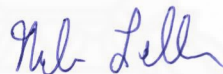
ENTITLED

DEVELOPING A VALID AND RELIABLE ANALYTIC RUBRIC TO ASSESS  
CRITICAL THINKING IN ENGLISH ARGUMENTATIVE ESSAYS OF EFL  
UNIVERSITY STUDENTS

was approved as partial fulfillment of the requirements for  
the degree of Master of Arts

on July 31, 2019

Chairman



---

(Nunthika Puthikanon, Ph.D.)


Member and Advisor



---

(Associate Professor Dumrong Adunyarittigun, Ph.D.)

Member



---

(Jiraporn Dhanarattigannon, Ph.D.)

Dean



---

(Assistant Professor Passapong Sripicharn, Ph.D.)

Thesis Title	DEVELOPING A VALID AND RELIABLE ANALYTIC RUBRIC TO ASSESS CRITICAL THINKING IN ENGLISH ARGUMENTATIVE ESSAYS OF EFL UNIVERSITY STUDENTS
Author	MR. NATTAWUT NAKKAEW
Degree	Master of Arts
Major Field/Faculty/University	English Language Studies (ELS) Faculty of Liberal Arts Thammasat University
Thesis Advisor	Associate Professor Dumrong Adunyarittigun, Ph.D.
Academic Years	2018

## **ABSTRACT**

This study aimed to develop and validate Critical-Thinking-in-Argumentative-Essay Rubric (CTER) for EFL college students. Two research questions were: 1) Is the CTER a valid and reliable assessment instrument? and 2) What are the attitudes of the raters and the writing teachers after they use the CTER? This study used a mixed methods research design and had three groups of participants: five experts, three raters and three writing teachers. Data sources included the experts' validation survey for the CTER, interview, field notes and questionnaire. Two levels of validation of the rubric were conducted, and the evaluative descriptors were revised based on the experts' comments. The writing teachers and the researcher used the CTER to evaluate the writing samples. Then, the scores obtained from the evaluation were analyzed to examine the inter-rater reliability of the rubric. The analysis of the quantitative and qualitative data showed that the CTER is a valid and reliable assessment. That is, it contained six valid and clear indicators of critical thinking in

argumentative writing of EFL students and had the moderate inter-rater reliability (Kendall's  $W = 0.5$ , ICC = 0.70) that was statistically significant at the 0.05 level. The raters and the writing teachers also perceived that the rubric can be used to promote teaching, learning and critical thinking of EFL college students. Pedagogical implications and recommendations for future research were presented based on the findings of this study.

**Keywords:** the CTER, analytic rubric, critical thinking, argumentative essay, validation, validity, inter-reliability



## ACKNOWLEDGEMENTS

“[E]ven though we do not see the shores, we still have to persist in our swimming in the wide ocean. Any individual who practices perseverance, even in the face of death, will not be in any debt to relatives or gods or father or mother. Furthermore, any individual who does his duty like a man will enjoy Ultimate Peace in the future.”

from the book *The story of Mahajanaka* (1997, pp.21-23)

Earning a master degree in ELS Program at Thammasat University is a long and challenging process for me, but at the same time, also a rewarding and invaluable experience. Through the process of creating this thesis, I have been grateful for help and support from many people.

First, I wish to express my sincere gratitude and appreciation to my advisor, Associate Professor Dr. Dumrong Adunyarittigun, for his kindness, guidance, patience, support, inspiration and wisdom. During the process of supervising this thesis, he does not give me fish but teaches me to fish by myself. He encourages me to pursue my own ideas and provides valuable guidance for me to grow professionally like a researcher. He challenges me to think critically, provides me with key ideas and insightful comments and is always available for me when I have problems. He introduces me to the conferences and the distinguished scholars in the field of teaching and language assessment. He also provides me with the opportunity to gain working experience and knowledge of the world. He is my role model of working, teaching and conducting research professionally. I am very lucky to have him as my advisor.

My sincere gratitude and appreciation also goes to Dr. Nunthika Puthikanon, the chair of the committee, and Dr. Jiraporn Dhanarattigannon, a thesis defense committee member, for their assistance and constructive suggestions on my thesis. Both of them encourage me to think critically about my research and conduct and write research efficiently. Their kindness,

profound knowledge and academic excellence inspire me to keep developing myself into a better writer, teacher, researcher and life-long learner.

I also owe a huge debt to many people whose warm support, cheerful encouragement, genuine guidance and kind assistance have been incalculable: Associate Professor Dr. Rachel Grant, Associate Professor Dr. Nitaya Suksaeresup, Assistant Professor Dr. Saneh Thongrin, Assistant Professor Nattha Kaewcha, Dr. Walaiporn Chaya, Dr. Sirinan Srinaowaratt, Dr. Siri-on Lada, Dr. Korawan Deekawong, Dr. Khwanchanok Seubsuk, Dr. Pranee Srinark, Dr. Manachai Inkaew, Asst. Prof. Thanis Bunsom, Ajarn Garry Mcfeeter, Ajarn Nikom Tolongthong and Ajarn Hasana Binma-oung.

I am also indebted to Associate Professor Dr. Viphawee Vongpumivitch, Associate Professor Dr. Wararit Panichkitkosolkul, Dr. Pornpimol Sukavatee, Assistant Professor Dr. Suriyan Panlay, Dr. Chalida Janenoppakarn, Ajarn Marisa Phongsirikul, Ajarn Taweewat Inree, Ajarn Yossiri Yossathorn, Nelobol Nipakornkitti (Pee Bow), Ajarn Chatchawan Chaiyasat (Pee Wood), Trirat Phokanoey (Pee Go), Apisit Klaopitak (Pee Bie), Chanapha Muaeng-ngamsomboon (Pha), Thanit Neelayotin (Stang) and Nareupun Sartjeanphong (Go) for their kind assistance and constructive comments on my thesis. Their help in various forms also meant a lot to me.

I am also grateful to Luangpor P. A. Payutto for cultivating a thirst for knowledge, persistence for education and a life-long learning into me. I have benefitted so much from his teachings concerning learning, working, conducting research and solving problems. From-the-bottom-of-my-heart thanks go to Ajarn Orachun Saw-seang-mongkon, my late English teacher, who inspired me to love studying and teaching English. I also extend my sincere gratitude to Ajarn Amara Phathanaphonkornsahun, a yoga teacher at Saranrom garden, and other members in a yoga club for teaching me to do yoga. Yoga helps me release stress from studying and working and gives me strength to complete my studies. I also would like to take

this opportunity to thank all professors in the ELS program for providing me knowledge, encouragement and inspiration in learning, teaching and creating academic works. My thesis would not have been possible without these people.

Besides, my warmest thanks which cannot put into words go to ELS classmates, Valin Lay (Pee Free), Thanisorn Reungsri (L), Abhisit Teeworawat (Pee Sit), Jirash Piasiri (Pee Pleang), Jingjo Kongmun (Pee Joe), Chartbun Saipradit (Pong), Thanyathorn Parnroch (Aern), Natthanon Phakawinyou (Tom), Phatcharakit Panyakaew (Lek), Thanyaporn Deelai (Pee Nud), Puwit Watecho (Bung), Nattakrit Intaraprasert (Kit), Apiwat (Nase), Amornchai Rungsuwanrat (Phong), Baramee Suthampaisarn (Ome), Apiboon Buranavej (Mon), Chatnarong Sawangjang (Pee Dum), Supak Toochemroen (Cassi) and many other friends. Their friendship and encouragement have made this long process a pleasant one for me.

Heartfelt thanks also go to Khun Waraporn Prommanee (Pee Ice) and Khun Saowalak Suwonnabut (Pee Jum) for their kind help and excellent educational, administrative service. I also thank Khun Chinnakrit Kongyoo for his useful guidance about a statistical analysis of data. Special thanks also go to TU librarians Khun Chaiyasit Angkapunyadej and Khun Kittanun Dussadeemethanont (Class) for helping me download theses and research articles.

Last but certainly not the least, I would like to express my deepest gratitude and appreciation to my parents and my brother for their unconditioned love, care and support. My parents have instilled endurance and perseverance into me. My brother always helps me when I have problems and always supports me to do what I love. My family has provided me with the best education they could get and nurtured my passion of doing research and my love of knowledge. Their understanding and confidence in me have carried me through every hardship along the way while I have completed the accomplishment of the thesis. Without them, my dream could not have come true. Thank you!

Nattawut Nakkaew

## TABLE OF CONTENTS

ABSTRACT	(1)
ACKNOWLEDGEMENTS	(3)
LIST OF TABLES	(12)
LIST OF FIGURES	(13)
LIST OF ABBREVIATIONS	(15)
CHAPTER 1 INTRODUCTION	1
1.1 Background of the study	1
1.2 Statement of the problem	2
1.3 Significance of the study	5
1.4 Definitions of terms	6
1.5 Conclusion	7
1.6 Chapter summary	7
CHAPTER 2 REVIEW OF RELATED LITERATURTE	8
2.1 Theories and definitions of critical thinking	8
2.1.1 Critical thinking theories in the field of philosophy.	8
2.1.2 Critical thinking theories in the field of psychology.	9
2.1.3 Critical thinking theories in the field of education.	10
2.2 Critical thinking and English argumentative writing	11
2.2.1 The relationship between critical thinking and argumentative writing	11

	(7)
2.2.2 The development of an argument	12
2.2.3 Development from Toulmin’s model to other models of arguments	14
2.3 Critical thinking in EFL/ESL context	16
2.3.1 The differences of culture of learning and teaching between Asians and westerners	17
2.4 Contrastive rhetoric and EFL/ESL students’ critical thinking abilities	18
2.4.1 The relationship between CR and critical thinking of ESL/EFL students	19
2.4.2 Many factors affecting L2 writing	19
2.4.2.1 The positive transfer	20
2.4.2.2 The developmental and educational factor	20
2.4.2.3 The cross-cultural values	21
2.5 Existing critical thinking tests	22
2.5.1 Three types of the existing critical thinking tests	22
2.5.1.1 The multiple-choice critical thinking tests	22
2.5.1.2 The open-ended critical thinking tests	23
2.5.1.3 The combination between multiple-choice and open-ended formats	23
2.5.2 Limitations of the existing critical thinking tests: turning crisis into opportunities	24
2.6 Principle of assessment and a rubric	25



	(9)
2.7.3.1 The nature of the analytic rubrics	41
2.7.3.2 Problems of the previous critical thinking analytic rubrics	41
2.7.3.2.1 The previous rubrics lack clear rating scales and/or lack clear evaluative descriptors in each scale level	42
2.7.3.2.2 The previous rubrics overlook some crucial, relevant aspects of critical thinking in English argumentative essays of EFL students	43
2.8 Chapter summary	43
<b>CHAPTER 3 RESEARCH METHODOLOGY</b>	<b>45</b>
3.1. Research design	45
3.2 Research questions	45
3.3 Participants and context	46
3.3.1 Three experts taking part in the first level of validation of the CTER	45
3.3.2 Five experts taking part in the second level of validation of the CTER	45
3.3.3 Three raters who implement the CTER	47
3.4 Ethic consideration	50
3.5 Research instruments	51
3.6 Research procedures	55
3.7 Data collection and data analysis for each research question	61
3.8 Chapter summary	67

CHAPTER 4 RESULTS AND DISCUSSION	68
4.1 Research Question 1: Is the CTER a valid and reliable assessment instrument?	69
4.1.1 Findings from first level of validation of the CTER	70
4.1.2 Findings from the second level of validation of the CTER	70
4.1.3 The inter-rater reliability of the CTER	90
4.1.4 Summary of the results about the validity and the inter-rater reliability of the CTER	93
4.2 Research Question 2: What are the attitudes of the raters and the writing teachers after they use the CTER?	94
4.2.1 Summary of the raters' and the writing teachers' attitudes towards the CTER	102
4.3 Implications of the study	103
4.4 Limitations of the study	105
4.5 Recommendations for future research	106
4.6 Conclusions of the study	108
4.7 Chapter Summary	108
REFERENCES	109
APPENDICES	145
APPENDIX 1 A table summarizing the previous multiple choice critical thinking tests	146
APPENDIX 2 A table summarizing the domains of assessment of the previous	148

holistic critical thinking rubric	
APPENDIX 3 A table summarizing the previous analytic critical thinking rubrics which contain unclear evaluative descriptors	149
APPENDIX 4 A table summarizing the scoring criteria of the previous analytic critical thinking rubrics	149
APPENDIX 5 The detailed descriptions of the six domains of the CTER	150
APPENDIX 6 The first version of the CTER	156
APPENDIX 7 The second version of the CTER	157
APPENDIX 8 The final version of the CTER	159
APPENDIX 9 Rating manual for the CTER	161
APPENDIX 10 The checklist	163
APPENDIX 11 Anchor papers 1 – 8 for the final version of the CTER	167
APPENDIX 12 Experts' validation survey for the CTER	179
APPENDIX 13 Interview questions for three raters and three writing teachers	190
APPENDIX 14 Open-ended questionnaire for the writing teachers	191
BIOGRAPHY	193

## LIST OF TABLES

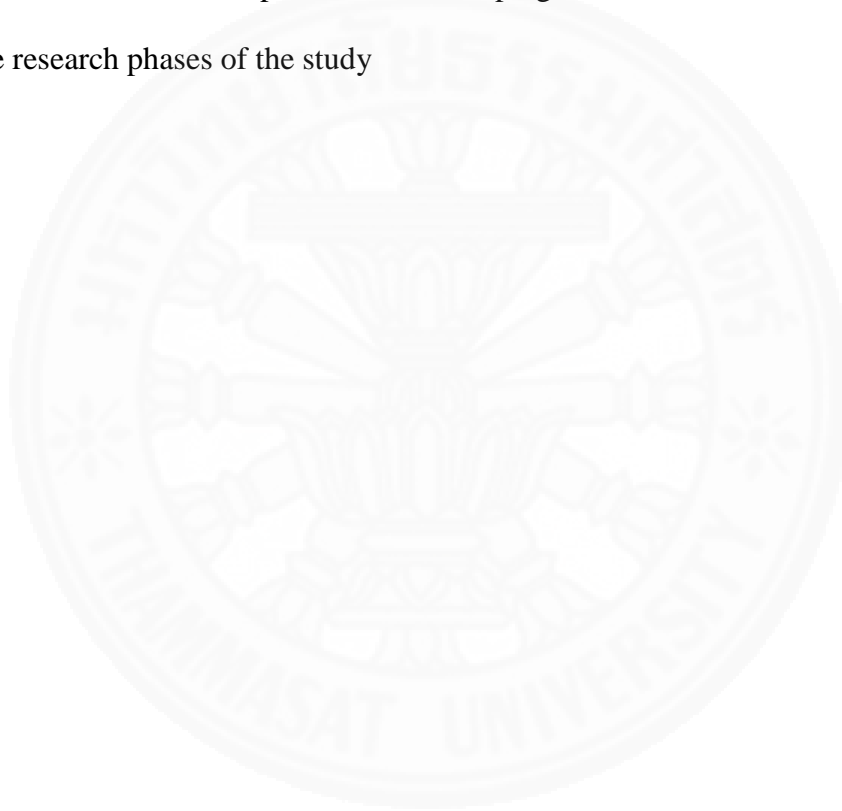
Tables	Page
2.1 The example of a formal logic	13
2.2 The example of an analytic rubric	33
2.3 The critical thinking checklist rubric of Liu, Wu and Shieh (2015)	37
2.4 The example of a holistic rubric	39
3.1 Criteria for interpreting the mean score from the experts' validation survey	63
3.2 A summary of the data sources and data analysis for each research question	67
4.1 The validation results of the six domains of the critical thinking assessment	71
4.2 The validation results of the descriptions of the six domains of the assessment	72
4.3 The validation results of the evaluative descriptors in each score level	72
4.4 The validation of the descriptors' ability to make a clear distinction among the four score levels	74
4.5 The second and the final versions of the descriptors of domain 1	77
4.6 The second and the final versions of the descriptors of domain 2	79
4.7 The second and the final versions of the descriptors of domain 3	81
4.8 The second and the final versions of the descriptors of domain 4	82
4.9 The second and the final versions of the descriptors of domain 5	84
4.10 The second and the final versions of the descriptors of domain 6	86

4.11 The results of the inter-rater reliability of the CTER via Kendall's W	90
4.12 The results of the inter-rater reliability of the CTER via ICC	91



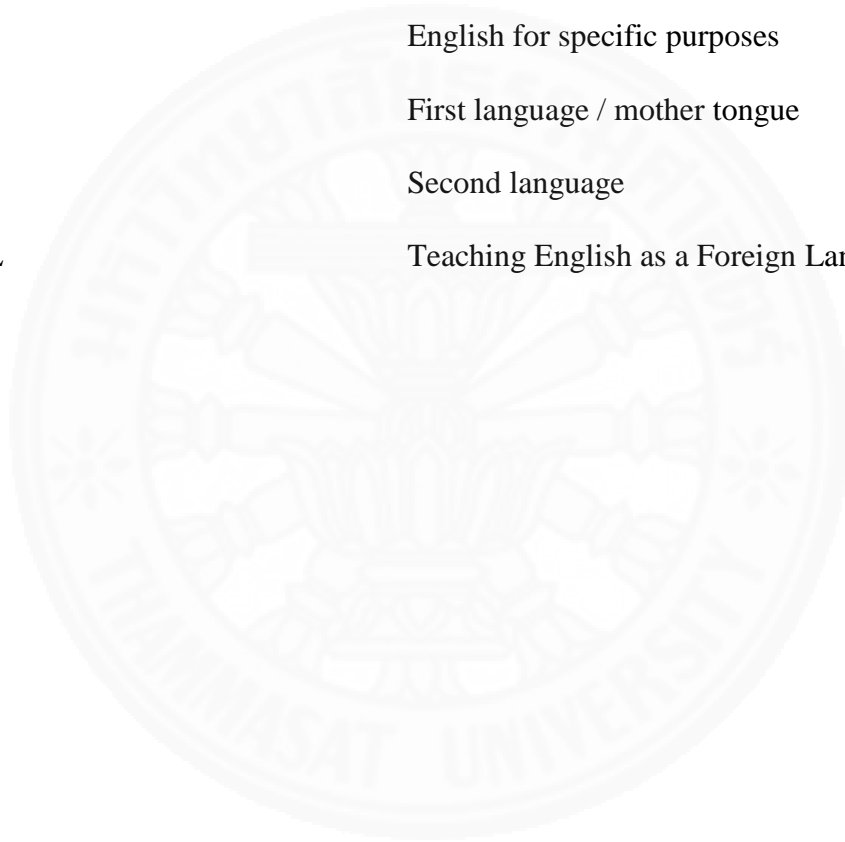
**LIST OF FIGURE**

Figures	Page
2.1 Toulmin's model	14
2.2 Three models of English argumentative writing	15
2.3 Stapleton's critical thinking elements in argumentative writing	16
2.4 Reliability and validity	29
2.5 Daud's (2012) model for developing a critical thinking rubric	34
3.1 Overview of the research phrases for developing the CTER	55
4.1 Three research phases of the study	69



## LIST OF ABBREVIATIONS

<b>Symbols/Abbreviations</b>	<b>Terms</b>
CR	Contrastive rhetoric
CT	Critical thinking
EFL	English as a foreign language
ESL	English as a second language
ESP	English for specific purposes
L1	First language / mother tongue
L2	Second language
TEFL	Teaching English as a Foreign Language



# CHAPTER 1

## INTRODUCTION

Education is the most powerful weapon which you can use to change the world.

Nelson Mandela

### 1.1 Background of the study

Critical thinking includes many crucial subskills, such as applying, analyzing, synthesizing and evaluating information (Chatfield, 2018; Fisher, 2011; Scriven & Paul, 1987), thinking clearly, carefully and precisely (Paul & Elder, 2006; Tittle, 2011), making a logical argument (Hillocks, 2010; Stapleton, 2001a, 2001b; Foresman et al., 2017), providing a constructive comment (Lau, 2011), creating new knowledge (Silva, 2008), solving problems (Stein & Haynes, 2011) and cultivating active curiosity and inquiry (Thomas, 2009; Vallis, 2010). Critical thinking is a higher-order thinking (Angelo, 1995; Liu & Stapleton, 2018), a life-long learning skill (Celuch & Slama, 2010; Wade, 1995) and a foundation of science (Lau, 2011). Thus, the cultivation of critical thinking should be the central aim of education (Lau, 2011; Thomas, 2009; Wilson, 2016), and assessing students' critical thinking abilities should be a main responsibility of teachers (Nair, 2011).

In Thailand, the English curriculum in all educational levels is in line with an educational reform measure; namely, the 1999 National Education Act (NEA) and the National Scheme on Education published by the Office of the National Education Commission (ONEC). One of the objectives of the NEA and the National Scheme on Education is that Thai students will be proficient at critical thinking (ONEC, 2002). As a result, it is important for Thai teachers to promote critical thinking of the students.

Although critical thinking is a desirable outcome of Thai education, many Thai students still lack critical thinking skills (Prommak, 2016; Puthikanon, 2009). That is firstly because many Thai teachers use traditional teaching methods which emphasize rote learning, while critical thinking is minimally treated (Puthikanon, 2009). Secondly, most exams are in the form of multiple choice tests rather than performance-based assessments (Jantrasakul, 2004). This does not promote critical thinking of the students (Blattner & Frazier, 2002). Thirdly, Thai culture of learning does not stimulate critical thinking of the students. Thai culture is hierarchical; that is, to be decent Thais, students should respect parents, teachers and elders (Thongrin, 2002). Thus, the students are hardly encouraged to question or argue against older people because in Thai culture, such behavior is regarded as a challenging behavior on the part of the students (Thongrin, 2002). As a result, many students scarcely dare to express their ideas critically or think differently from the elderly (Jantrasakul, 2004). For these reasons, teachers should find ways to help Thai students develop critical thinking.

Writing an argument for or against a particular position is an activity that can foster critical thinking of students because the students need to analyze, synthesize and evaluate information from various perspectives (Puthtikanon, 2009). They also need to indicate a claim and support it with reasons and evidence, state and refute an opposing view, write a conclusion and organize ideas in a logical way. These skills require the use of critical thinking (Burkhalter, 1993; Puthtikanon, 2009). Thus, it can be said that argumentative writing is an analytical and critical writing (Barnet, & Bedau, 1993; Cheng, 2003; Malcolm, 2011; Pawlowski, 1997), is a task-based language teaching (Yang, Lu, & Weigle, 2015) and is a “valid method” for evaluating critical thinking of the students (Hughes, 2000, p.44).

## **1.2 Statement of the problem**

A rubric can help teachers assess critical thinking in students' argumentative essays. A good critical thinking rubric should have the following qualities: 1) have clear domains of assessment, clear rating scales, and clear evaluative descriptors in each score level, 2) have inter-rater or intra-rater reliability and 3) promote teaching and learning by serving as an instructional tool, providing students with useful, diagnostic feedback, helping them monitor their own learning and helping teachers plan and improve their own instruction (Bauer, 2016; Dickinson, & Adams, 2017; Moskal & Leydens, 2000; Stevens & Levi, 2005; Williams, 1999). As Penny, Johnson and Gordon (2000) stated, a critical thinking rubric that is valid, reliable, accessible and can promote students' learning is the heart of educational assessment.

However, the previous rubrics suffer from many important problems. To explain, there are three types of critical thinking rubrics: checklist, holistic and analytic. A checklist rubric of Stapleton (2001a; 2001b) and that of Liu and Stapleton (2014) count the frequency of each critical thinking element (e.g., claim, reasons, evidence, and conclusion) in English argumentative writing of EFL students. However, these rubrics do not take the quality of the critical thinking elements into consideration. Also, a rubric of Liu, Wu and Shieh (2015) just checks the presence of each critical thinking element in argumentative writing of EFL students (e.g., stating a position and giving reasons and evidence to support a claim). However, this rubric does not evaluate the students' levels of performance in writing each element of critical thinking. In other words, the previous checklist rubrics have domains of critical thinking assessment; however, they lack levels of performance and evaluative criteria in each performance level. Therefore, they cannot offer detailed, diagnostic information about each critical thinking element and thus should not be used with EFL students.

The holistic rubrics of the previous educators (e.g. Blattner & Frazier, 2002; Facione, & Facione, 1996; Angeli & Valanidies, 2009) cannot provide an accurate picture of a student's critical thinking ability in argumentative essays. To explain, the existing holistic

rubrics evaluate the critical thinking abilities; namely, 1) writing a claim, 2) giving reasons and evidence to support the claim, 3) stating a counter-argument and a refutation, 4) organizing ideas logically, 5) using effective language and 6) providing a suggestion, a recommendation or a solution. The previous rubrics assume that these aspects of critical thinking develop at the same rate and can be graded in a single score (Saxton, Belanger & Backer, 2012). However, this assumption is faulty because a student who is proficient in one critical thinking skill may not be always proficient in the other one (Saxton et al., 2012). Generalizing a student's critical thinking ability levels to one holistic score, therefore, fails to provide the students with valid, diagnostic feedback (Saxton et al., 2012).

An analytic rubric is the most appropriate tool for use with EFL students because it specifically evaluates each domain of critical thinking assessment and offers diagnostic feedbacks for each domain. Unfortunately, the previous critical thinking analytic rubrics suffer from two important problems. First, descriptors of the existing rubrics (e.g., Dong, 2015; Saxton et al., 2012) are subjective and cannot make a clear distinction between score levels. This problem threatens the validity of these rubrics. For example, in a rubric of Saxton, Belanger and Becker (2012), domain of the assessment, namely Inference, contain unclear evaluative descriptors as follows: (Score Level 6) “Demonstrates **confident** ability to apply or extend key concepts”... and “demonstrates **surprising/insightful** ability to take concepts further” ... vs. (Score Level 5) “Demonstrates a **clear** ability to apply or extend key concepts” ... and “demonstrates **strong** ability to take concepts further” ... (p.268, emphasis added). Obviously, the use of subjective adjectives (i.e., confident, surprising and insightful in Score Level 6 and clear and strong in Score Level 5) cannot clearly differentiate the descriptors among the score levels. Also, rubrics of Billups (2006), Carbone (2009), Daud (2012) and McLaughlin and Moore (2012) do not have evaluative descriptors in each level of performance. As Weigle (2002) emphasized, “the development of a scale and descriptors of

each scale level are of critical importance for the validity of the assessment” (p.109). Lacking these crucial components in a rubric, therefore, calls the validity of the previous rubrics into question.

Second, the rubrics of the previous researchers (e.g., Hughes, 2000; Brookhart, 2010; Metcalfe, 2007) do not assess some crucial, relevant aspects of critical thinking construct in argumentative writing of EFL students. To explain, these rubrics include domains of critical thinking assessment; namely, claims, reasons, evidence, counter-arguments, refutations, organization and conclusions. However, they overlook other important dimensions of critical thinking in argumentative writing: 1) an ability to write an introduction to gain readers’ attention, 2) an ability to provide a final, constructive comment in a conclusion and 3) an ability to use word choices and sentence structures effectively. According to many researchers (e.g., Moskal & Leydens, 2000; Nair, 2011), if a rubric does not take all of the relevant, crucial aspects of critical thinking into consideration, that rubric will be invalid. In short, the existing rubrics do not embrace some relevant, crucial aspects of critical thinking in argumentative writing, so they lack a construct representation; that is, lack validity.

To date, there is no suitable critical thinking rubric for use with EFL university students. This study will develop a valid and reliable rubric for assessing critical thinking in English argumentative writing. The proposed rubric is called Critical-Thinking-in-Argumentative-Essay Rubric (CTER). Two research questions of this study are as follows:

- (1) Is the CTER a valid and reliable assessment instrument?
- (2) What are the attitudes of the raters and the writing teachers after they use CTER?

### **1.3 Significance of the study**

This study proposes a valid and reliable analytic rubric for assessing critical thinking in English argumentative essays of EFL university students. Since there has been a dearth of

research about assessing critical thinking in argumentative writing of EFL students (Preiss et al., 2013), especially in a Thai context (Bunsom, Vungthong, & Singhasri, 2002), this study will benefit writing teachers and EFL students in particular.

#### **1.4 Definitions of terms**

Argument refers to giving reasons and evidence to convince an audience to accept a claim or a position of a speaker or a writer (Hirschberg, 1996; Winkler & McCuen, 1997).

Argumentative writing refers to a type of writing which a writer takes a stance on a controversial issue and supports it with reasons and evidence in order to convince readers (Qin, 2009).

Critical Thinking refers to applying, analyzing, synthesizing and evaluating information and making a logical argument (e.g. Scriven & Paul, 1987; Rainbolt & Dwyer, 2012; MamourChoul, 2010; Puthikanon, 2009).

Rubric refers to a scoring tool which contains the domains of assessment, the rating scales and the evaluative descriptors in each rating scale (Weigle, 2002; Stevens & Levi, 2005).

Validity refers to the extent to which inferences and decisions made from assessment results are accurate, appropriate and meaningful in terms of the purpose of the assessment (Gronlund, 1998; Verlinden, 2005).

Reliability refers to the extent to which a rubric yields consistent scoring results (Brown & Abeywickama, 2010; Verlinden, 2005).

Inter-rater reliability refers to the measurement of the consistency among raters (Moss & Koziol, 1991; Simpkins, 2016)

Assessment refers to using various methods or approaches to collect information in order to make judgments or decisions about students' knowledge, ability or learning (Chapelle & Geoff, 2010; Katz, 2012).

Assessment for learning refers to a formative assessment—i.e., an assessment that provides useful, diagnostic feedback to students, can be used as an instructional tool and promotes critical thinking and learning of students (e.g., Shepard, 2000; Stiggins et al., 2004).

### **1.5 Conclusion**

Critical thinking is crucial for writing an argumentative essay, yet it remains unclear and challenging for language teachers to assess critical thinking. To date, there is no appropriate critical thinking rubric for use with EFL students. This study will develop such a valid and reliable rubric, namely the CTER, to assess critical thinking in argumentative essays of EFL students.

### **1.6 Chapter summary**

Chapter I presents an introduction to this study. The problem statement and the research questions are addressed. The significance of the study and the operational definitions of terms are provided.

## CHAPTER 2

### REVIEW OF RELATED LITERATURE

Thinking is at the heart of what it means to be human, so to fail to develop your [critical] thinking potential is to preclude the full expression of your humanity

(Hager & Kaye, 1992, p.27)

The chapter reviews key concepts related to assessing critical thinking in English argumentative writings. It discusses 1) theories and definitions of critical thinking, 2) critical thinking and argumentative writing, 3) critical thinking in EFL/ESL contexts, 4) contrastive rhetoric and EFL students' critical thinking, 5) existing critical thinking tests, 6) principle of assessment and a scoring rubric and 7) problems of the existing critical thinking rubrics.

#### **2.1 Theories and definitions of critical thinking**

The root of critical thinking was traced back about 2,400 years ago to the teaching of Socrates, a Greek philosopher, who used questioning technique to teach how to think critically to his students (Paul, 1985). His questioning technique is to question common beliefs and distinguish beliefs that are reasonable from those that lack adequate reasons and evidence (Paul & Elder, 2001; Dong, 2015). This technique is followed by many subsequent philosophers, such as Plato and Aristotle (Paul & Elder, 2001; Dong, 2015). Subsequently, a movement of critical thinking extended across centuries until in the early 1980s, the studies on critical thinking expanded into three fields, i.e., philosophy, psychology and education (Fasko, 2003; Sternberg, 1985).

##### **2.1.1 Critical thinking theories in the field of philosophy.**

Philosophy teaches people to think critically and try to arrive at the truth of something by using reasons or argumentation (Payutto, 2015). Thus, critical thinking theory in the philosophical field focuses on using logic, reasons and effective thinking (Dong, 2015).

Scriven and Paul (1987) defined critical thinking as applying, analyzing, synthesizing and evaluating information. Paul and Elder (2008) specifically elaborated that critical thinking consists of three elements: a standard, elements of thoughts and intellectual traits of mind.

- The standards (or criteria of effective thinking) include clarity, accuracy, relevance, logicalness, precision, completeness, significance (stating central ideas), depth (discussing issues thoroughly), breadth (considering others' opinions) and fairness (considering contexts and others' perspectives) (Paul & Elder, 2008).
- Eight elements of thought include purposes, questions, points of view, information, inferences, concepts, implications and assumptions (Paul & Elder, 2008).

To develop critical thinking, one needs to apply the standards to the elements of thought (Paul & Elder, 2008). For example, in making an argument, one needs to make his/her point of view, information and implication accurate, clear, logical and so on (Paul & Elder, 2008). This will help one develop the following intellectual traits of mind—an intellectual humility, an intellectual autonomy, an intellectual integrity, an intellectual courage, an intellectual perseverance, an intellectual empathy, fair-mindedness and a confidence in reasoning (Paul & Elder, 2008).

### **2.1.2 Critical thinking theories in the field of psychology.**

Sternberg (1985) believed that critical thinking involves three components: a meta-component (solving problems), a performance component (reasoning) and a knowledge-acquisition component (organizing ideas logically). Ennis (1993) further defined critical thinking as a reasonable reflective thinking. According to Ennis (1993), critical thinking includes many crucial skills as follows: 1) judging the credibility of sources, 2) identifying conclusions, reasons and assumptions, 3) judging the quality of an argument, reasons and evidence, 4) developing and defending a position on an issue, 5) asking questions clearly and appropriately, 6) planning and judging experimental designs, 7) defining terms that is

appropriate for a context, 8) being open-minded, 9) trying to be well informed and 10) drawing a logical conclusion.

### **2.1.3 Critical thinking theories in the field of education.**

Bloom and his colleagues (1956) proposed a taxonomy which consists of three higher-order thinking skills; namely, analysis, synthesis and evaluation. In the year of 2001, Anderson and Krathwohl revised Bloom's taxonomy by using three higher-order thinking skills as follows: 1) analyzing, 2) evaluating and 3) creating—i.e., writing, planning, designing, producing and/or solving problems. The revised taxonomy contains two dimensions--knowledge dimension and cognitive dimension (Krathwohl, 2002). The knowledge dimension contains four kinds of knowledge that are arranged from the most concrete to the most abstract; that is, factual, conceptual, procedural and metacognitive, respectively, whereas the cognitive dimension contains three crucial aspects: analyzing, evaluating and creating (Krathwohl, 2002; Mayer, 2002).

Other researchers also tried to define the term critical thinking. For example, Facione (1990) used an expert consensus approach to identify a definition of critical thinking. He asked forty-six experts in the field of critical thinking to provide their own list of critical thinking skills. After a final round of discussion, the experts reached an agreement on the six dimensions of critical thinking, i.e., an interpretation, an analysis, an evaluation, an inference, an explanation and a self-regulation (Facione, 1990). Lerner (2000) stated that Facione's study provides one of the most comprehensive and agreed-upon definitions of critical thinking construct. Similarly, other researchers (e.g., Chatfield, 2018; Fisher, 2011; Liu et al., 2016; Liu et al., 2018; Mayfield, 1994) defined critical thinking as reasoning, interpreting, analyzing and evaluating ideas and arguments. Likewise, Jones and his colleagues (1997) defined critical thinking as an interpretation, an analysis, an evaluation, an inference, a presentation of argument, a reflection and a disposition—a tendency to use critical thinking when necessary.

From the previous critical thinking theories and research, it can be concluded that critical thinking includes crucial subskills, such as reasoning, applying, analyzing, synthesizing, evaluating and solving problems (Butterworth & Thwaites, 2013; Eva, 2018; Chatfield, 2018; Pu & Evans, 2019; Taube, 1997; Wilson, 2016). Three of the core definitions of critical thinking (i.e., analyzing, evaluating and creating) are the skills for the 21<sup>st</sup> century, and thus students should learn and develop these crucial skills (Silva, 2008).

## **2.2 Critical thinking and English argumentative writing**

This section will discuss how critical thinking and argumentative writing relate to each other. It points out a development of argument and discusses the application of models of argumentative writing in writing classes.

### **2.2.1 The relationship between critical thinking and argumentative writing**

Critical thinking and argumentative writing are reciprocal. That is, to write an argument effectively, one needs to use critical thinking; therefore, argumentative writing is both a process that fosters thinking and a product that reflects critical thinking of a writer (Bean, 2011; Chen, 2017; Dong, 2015).

The argument is giving reasons and evidence to convince audience to accept a claim of a writer or a speaker (Hirschberg, 1996; Winkler & McCuen, 1997). Two types of the argument include: critiquing an argument and writing an argument (Fleming, 2014; Qin, 2009). For the former, a student evaluates and responds to a position of others, whereas for the latter, a student takes a stance on a controversial issue and supports it with reasons and evidence (Qin, 2009).

Additionally, there are two main patterns of reasoning: deduction and induction (West, 1994). For the former, a student explicitly states a claim in an introduction and supports it with reasons and evidence (Inch & Warnick, 2002; Qin, 2009). For the latter, a

student does not need to state a claim in the introduction (Inch & Warnick, 2002; Qin, 2009). He/she can begin with providing reasons and evidence that lead to a position in a conclusion (Inch & Warnick, 2002; Qin, 2009). (See Nussbaum and Schraw, 2007, Qin, 2009, Rosenwasser and Stephen, 2012, VanderMey et al., 2012, and Yeh, 1998a, for models of inductive and deductive reasoning). Mulnix and Mulnix (2010) added that a teacher should teach the reasoning patterns, so students can think critically and write an argument effectively.

Writing an argument can promote critical thinking of EFL students. That is because they need to synthesize information from various perspectives, evaluate the credibility of the information and write a position on a controversial issue (Bauer, 2016; Hammer, 2017; Puthikanon, 2009). Also, the students need to analyze reasons, evidence and a structure of an argument and plan how to write a conclusion logically (Bauer, 2016; Brukhalter, 1993, 1995; Fisher, 2011; Puthikanon, 2009). Thus, argumentative writing is an analytical and critical writing (Cheng, 2003; Barnet, & Bedau, 1993) and is a “valid method” that can be used for assessing critical thinking of students (Hughes, 2000, p.44).

### **2.2.2 The development of an argument**

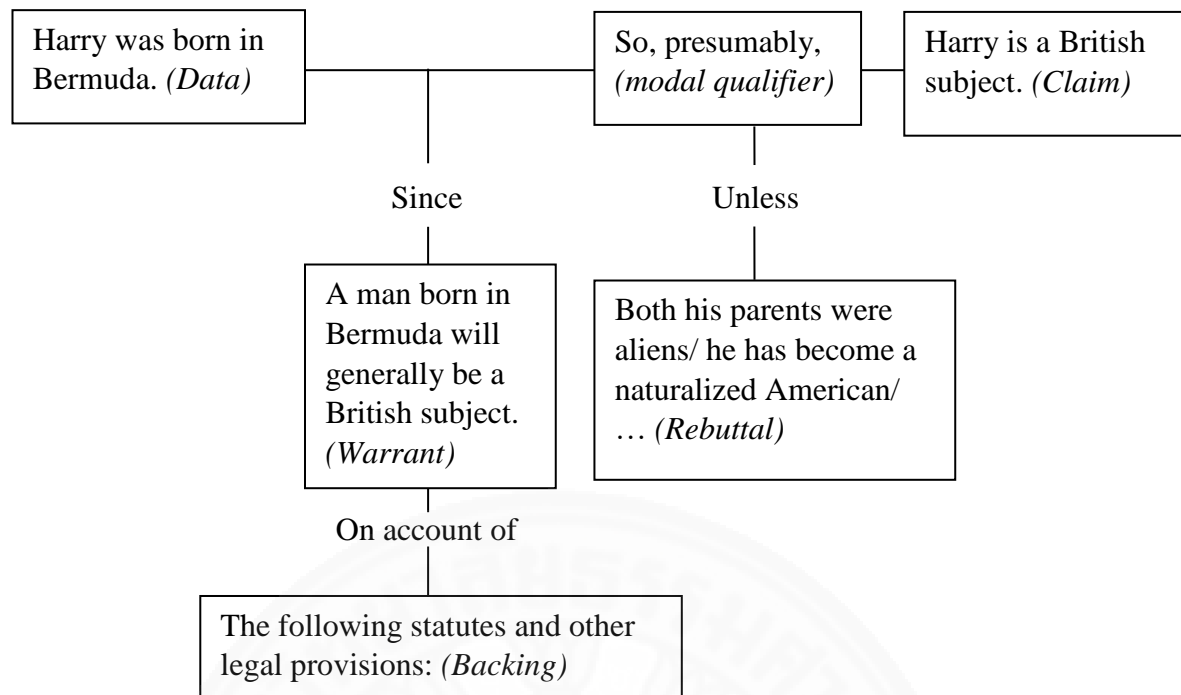
Invented by Aristotle, a formal logic or a syllogism contains three main parts: a major premise, a minor premise and a conclusion (Hillocks, 2010). The characteristic of the formal logic is that one or more statements (i.e., premises) support another statement (i.e., conclusion) (Annis, 1974). An example of a formal logic is presented in Table 2.1.

**Table 2.1:** The example of a formal logic

Major Premise:	All men are mortal.
	↓
Minor premise:	Socrates is a man.
	↓
Conclusion:	Therefore, Socrates is mortal (Hillocks, 2010)

The syllogism, however, has limitations. For example, the structure of the syllogism is too simplistic (Toulmin, 1958) and too narrow, so it is inappropriate for use as a strategy for argumentation (Vallis, 2010). Many writing teachers, therefore, used Toulmin's model of argument which focuses on a logic of possibility or a nature of real world argument (Fulkerson, 1996; Mateus, 2019). Elements in Toulmin's model of argument connect with each other (See Figure 2.1). The elements conclude claims, ground or data, warrants, backing, modal qualification and rebuttals (Toulmin, Rieke & Janik, 1979). Each element can be briefly explained as the followings.

- Claim is an opinion, a stance or a position.
- Ground is data or information that supports the claim. Ground can be observation, common knowledge, statistical data, personal testimony and so on.
- Warrant is a logical explanation of how ground supports the claim.
- Backing is a fact or additional information that back up the warrant.
- Modal qualifiers are adverbials, such as probably, possibly and presumably. Not all arguments support the claim with the same degree of certainty, so the modal qualifiers help modify the claim.
- Rebuttal tackles any possibilities that refute a writer's argument (Toulmin, et al. 1979).



**Figure 2.1** Toulmin's model (Toulmin, 1958)

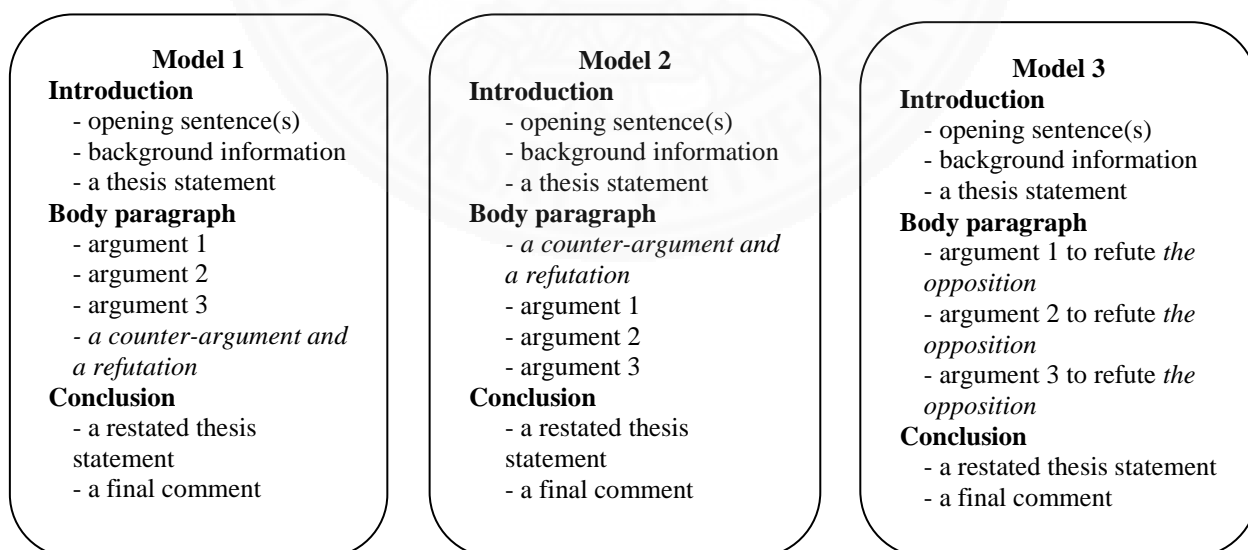
### 2.2.3 Development from Toulmin's model to other models of arguments

Although Toulmin's model was used as an instructional tool by many researchers (e.g., Hillocks, 2010; Zimmerbaum, 2014), the model itself still has limitations. First, it is difficult to clearly specify data, warrant and backing in argumentative essays. To explain, sometimes students write a claim and data without an explicit warrant; therefore, it is difficult for teachers to make a clear distinction between the data and the warrant or even between the data and the backing (Inch & Warnick, 2002). Second, Toulmin's model does not address a fallacy or an error in reasoning (Stapleton, 2001a). It is important for students to be aware of fallacies in argumentative writing, so they can think more critically and give reasons and evidence more effectively (Mulnix & Mulnix, 2010).

Due to the aforementioned limitations, many researchers adopted and adapted Toulmin's model. For example, Qin (2009) modified Toulmin's model and proposed six elements in argumentative writing of EFL students; that is, 1) a claim, 2) data (reasons and

evidence), 3) a counter-argument claim, 4) a counter-argument data, 5) a rebuttal claim (a statement that refutes the counter-argument claim) and 6) rebuttal data (reasons and evidence that support the rebuttal claim). However, although these elements were clear and easy for students to understand, they were developed with the purpose of evaluating students' general argumentative writing skills, not critical thinking. To assess critical thinking in argumentative writing, one needs to assess other crucial, relevant features of critical thinking, such as a conclusion and an organization. Also, although raters in Qin's research count the frequency of these specific elements, the raters do not evaluate students' level of performance in writing each element of argumentative writing.

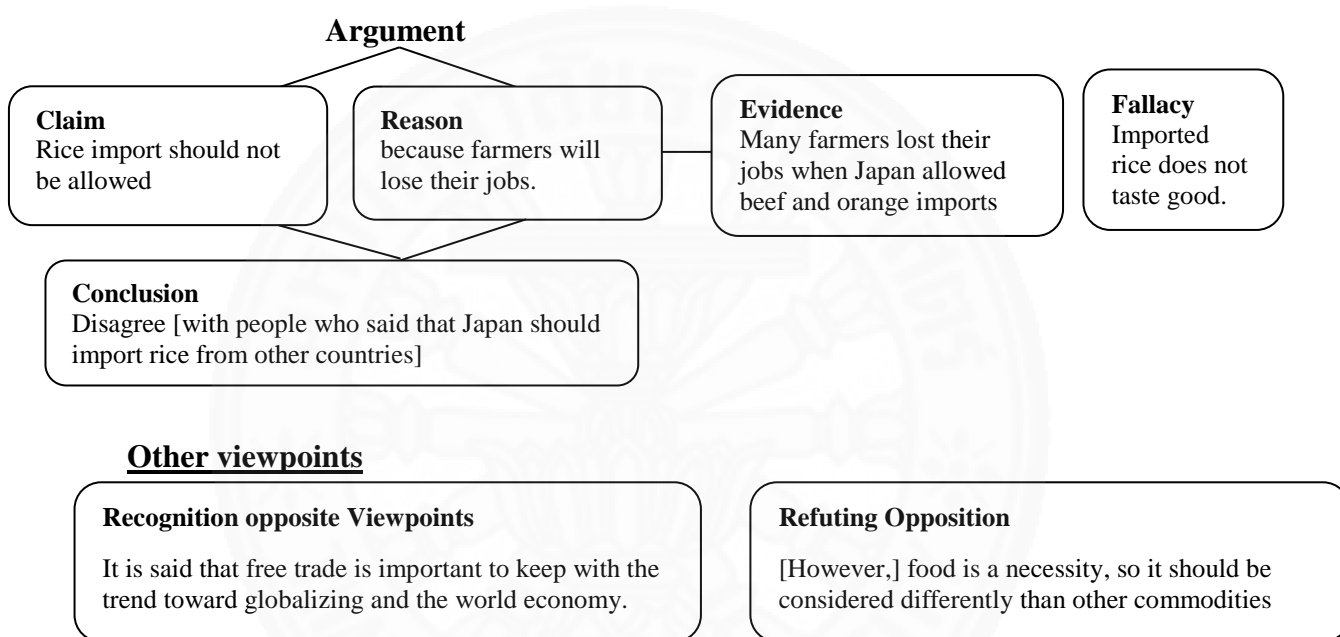
Many other researchers also modified Toulmin's model and proposed their own models of argument. For example, Liu and Stapleton (2014) proposed a diagram of argument which consists of a claim, data, a counter-argument, a rebuttal and a restatement of the claim in a conclusion, while Folse, Muchmore-Vokoun and Solomon (2002), Kalchayanant (2009) and Widodo (2012) proposed argument models which contain the same features of argumentative writing. Their models are summarized and presented in Figure 2.2.



**Figure 2.2** Three models of English argumentative writing

Stapleton (2001a) also adapted Toulmin's model and proposed a model of critical thinking in argumentative writing which consists of an argument (a claim and reasons), evidence, a counter-argument, a refutation, a conclusion and a fallacy. This model (See Figure 2.3) was used as a rubric for assessing critical thinking in argumentative writing of EFL students; that is, raters counted the frequency of each element of critical thinking in argumentative essays of the students (Stapleton, 2001a).

### Topic: Japan's rice import policy



**Figure 2.3** Stapleton's (2001a) critical thinking elements in argumentative writing

In sum, argumentative writing can be used to promote critical thinking abilities of EFL students, and many researchers adopted and adapted Toulmin's model and used it as a tool for teaching argumentative writing to EFL students.

### 2.3 Critical thinking in EFL/ESL context

This section firstly pinpoints the differences of a culture of learning and teaching between Asians and westerners. Then, it addresses two groups of scholars who have different perceptions about critical thinking abilities of EFL/ESL students. Finally, it addresses research to support the claim that the students have a potential to develop critical thinking.

### **2.3.1 The differences of the culture of learning and teaching between Asians and westerners**

Asians and westerners have different culture of learning and teaching. In EFL contexts, especially in Thailand, the culture of teaching and learning is hierarchical (Thongrin, 2002). That is, Thai teacher is perceived as a person who is most knowledgeable, so his or her role is a giver of knowledge, whereas the students' role is passive receivers (Dhanarattigannon, 2008). Thus, Thai EFL students are expected to listen and wait for knowledge from the teacher rather than participate in class actively (Dhanarattigannon, 2008). In addition, asking questions critically may be regarded as a challenging behavior on the part of the students (Dhanarattigannon, 2008). Conversely, in the American culture of learning, students are expected to think themselves, express ideas and participate in class actively (Dhanarattigannon, 2008). This, therefore, causes many western teachers to perceive that EFL students lack critical thinking.

### **2.3.2 Empirical research on critical thinking of EFL/ESL students**

In the field of English as a second language (ESL), many western educators (Atkinson, 1997; Ramanathan & Kaplan, 1996a; 1996b) perceived that only people who grow up in Western can acquire critical thinking naturally. They believed that Asian students were raised in a hierarchical culture, so it is very difficult for the students to learn and develop critical thinking (Atkinson, 1997; Ramanathan & Kaplan, 1996a, 1996b). However, findings from many research studies contradict the claim of the western educators. For example, Davison (1998) believed that critical thinking is not culturally-specific but is something universal, so EFL/ESL students can learn and develop critical thinking abilities. To investigate this reality, a study of Davidson and Dunham (1996) examined whether EFL students can learn and develop critical thinking in a content-based class. The investigators provided an experimental group (i.e., thirty six Japanese students) with a critical thinking instruction for one year. These

students learn how to identify an assumption, give reasons inductively and deductively and evaluate a credibility of source. Statistical results from the Ennis-Weir Critical Thinking essay test showed that due to the effect of the intervention, the experimental group gained significantly higher scores than a control group. The findings in Davidson and Dunham's study (1996) suggested that if EFL students receive a critical thinking instruction, they can think critically.

Likewise, Stapleton (2001b) perceived critical thinking as an attribute that all humans are equipped with at birth, so EFL students have the potential to think critically. He (2001a) developed the model of critical thinking (See Figure 2.3) and used it to evaluate argumentative essays of 45 Japanese students. The results indicated that EFL students' background knowledge and familiarity with a written topic affected the use of critical thinking elements in argumentative writing. That is, the more the students were familiar with the topic, the more frequently they wrote critical thinking elements in argumentative essays. Stapleton's (2001a) research results indicated that EFL students could express critical thinking, like their Western counterparts.

To conclude, ESL/EFL students may find it difficult to express critical thinking because they lack the opportunity to use critical thinking or lack background knowledge on the issue (Puthikanon, 2009). Thus, if they receive a critical thinking instruction which emphasizes reasoning, analyzing, synthesizing and evaluating, they can think critically and effectively (Davidson & Dunham, 1997; Puthikanon, 2009; Stapleton, 2001a, 2001b).

#### **2.4 Contrastive rhetoric and EFL/ESL students' critical thinking abilities**

Contrastive rhetoric (CR) is one area of research in the field of second language acquisition (Connor, 1996). CR believes that writing problems of L2 students stem from a negative transfer from L1 writing conventions (Connor, 1996, cited in Qin, 2009). CR research is vital since many research studies (e.g., Eggington, 2015; Tapinta, 2006; Thongrin,

2002, 2009; Uysal, 2006) reveal that EFL students transferred L1 rhetorical conventions into L2 writing. The next section will 1) discuss how the contrastive rhetoric relates to ESL/EFL students' critical thinking and 2) points out factors affecting their L2 writing.

#### **2.4.1 The relationship between CR and critical thinking of ESL/EFL students**

Kaplan (1996) analyzed 600 English expository texts of non-native English writers from different cultures. He found that English writers used a deductive, linear reasoning pattern, whereas Orient or Asian writers (e.g. the Japanese, Korean, Chinese, and Thais) used a circular reasoning pattern. Kaplan (1996) explained that these L2 writers could not write English texts logically and straightforwardly like native English speakers did because such concepts did not exist in their native, cultural thought patterns. Thus, when writing in English, the L2 writers transferred their L1 rhetorical conventions into English texts and produced the texts that were indirect and circuitous (Kaplan, 1996, as cited in Ismail, 2010). Kaplan's CR research implies that a lack of the clarity in L2 writers' texts indicates that they cannot think clearly and critically (Oda, 2008; Peirce, 2006).

However, Kaplan's study was seriously criticized of 1) being overgeneralizing that all Asian writers transfer their circuitous, cultural thought pattern to L2 writing, 2) lacking an analysis of the participants' relevant L1 data and 3) using only a text-based approach without other research methods for triangulation purposes (Ho, 2011). Also, findings of many research studies contradict the claim of Kaplan. These studies indicated that regardless of the L1 cultural thought pattern, other crucial factors do affect L2 writings of Asian students. The next section will discuss such factors in detail.

#### **2.4.2 Many factors affecting L2 writing**

Three main factors affect English writing of EFL students: 1) L1 positive transfer, 2) a developmental and educational factor and 3) cross-cultural values.

#### **2.4.2.1 The positive transfer**

Kubota's (1998) study examined both Japanese and English essays of 46 Japanese college students. The findings showed that although Japanese ESL students tended to use an inductive reasoning pattern when writing in English, they could change this mode of thinking and write L2 texts through using deductive reasoning pattern (Kubota, 1998). Similarly, findings of Kobayashi's study (1984) showed that when writing in English, one group of Japanese students in Japanese universities used L1 writing style (an inductive pattern), whereas the other group of Japanese students in American universities could shift L1 writing style into a deductive one and wrote English texts effectively. The results of these studies suggest that regardless of the L1 cultural thought pattern, skilled L1 writers can write L2 texts effectively and a positive transfer from L1 affects L2 writing of the students.

#### **2.4.2.2 The developmental and educational factor**

This factor refers to EFL students' English language proficiency, writing experience and/or the way they are taught in L1 writing classes. This factor dramatically affects L2 writing of Thai EFL learners, especially those from elementary and secondary schools, because many Thai writing teachers emphasize grammatical accuracy rather than ideas, organization and writing process (Dhanarattigannon, 2008). Also, topics that most Thai EFL students are assigned to write are only about Father's Day or Mother's Day (Dhanarattigannon, 2008), so many Thai students lack experience of writing an argument (Ka-kan-dee & Kaur, 2015). Conversely, in American writing classes both in a secondary school and in a university, writing teachers encourage students to make an argument by taking one side on a controversial issue and arguing strongly for their position (Ramanathan & Atkinson, 1999). Thus, it can be said that the developmental and educational factor potentially affects L2 writings of EFL students (Hua, 2014).

### 2.4.2.3 The cross-cultural values

The cross-cultural values involve assertiveness, different writing conventions and different perceptions about creativity in argumentative writing. First, regarding the assertiveness in L2 writing, people from a collectivist culture tend to care about others' feelings and a relationship within a group, so they tend to use the term *we* in writing (Wu & Rubin, 2002). Also, they concerned a social hierarchy and a face saving, so they seemed to avoid criticizing or arguing against others assertively (Thongrin, 2002). However, in western culture, people value the expression of their individual voice. That is, each person is regarded as equally competent to analyze, criticize and construct new knowledge, so the people feel free to argue or criticize others assertively (Young, 1994).

Besides the assertiveness, L1 writing conventions also affect L2 writing. English writing convention emphasizes conciseness, so a majority of English texts have a clear topic sentence and a strict, coherent organization throughout the text with a concise introduction, a body paragraph and a conclusion (Grabe & Kaplan, 1996; Thongrin, 2009). However, in a Thai context, a rhetorical convention of Thai essays does not require a strict organization, a clear topic sentence, concise paragraphs and a clear conclusion (Sudajit-apa, 2008; Thongrin, 2009). A writing convention affects how students write a text (Hyland, 2016), so it can be said that Thai L1 writing convention affects the way Thai students think and write in English.

Finally, the different perceptions about expressing creativity in L1 writing influence English writing of EFL students. To explain, English writing classes in the U.S. encourage writers to use their own original ideas (Cho, 1999). Thus, these writers tend to think and write in "their own voice using their own words" and perceive that using a well-known phrase or a proverb makes their text unoriginal (Wu & Rubin, 2002, p.156). Besides, when these writers take other people' ideas, they make citations (Dhanarattigannon, 2008).

Interestingly, however, in the cultural context of Asia, many people perceive that imitating classic works from other writers makes their texts more valuable, while a copyright is not seriously taken into consideration (Dhanarattigannon, 2008). Obviously, what is seen as good in one context is not always alike in the others (Sudajit-apa, 2008). Thus, teachers should teach the cross-cultural values in writing so that students have cultural awareness and knowledge of writing in L2 appropriately and effectively.

To recap, many factors affect L2 writing, so it is faulty to overgeneralize that all Asian students lack critical thinking due to their L1 cultural thought pattern. To investigate the reality about critical thinking in L2 writing of EFL students, researchers should take all crucial factors into consideration, use multiple research approaches and examine not only a final written product but also a writing process and an on-going development of the students. Findings from empirical research (e.g., Thongrin, 2002, 2009) revealed that after EFL students receive L2 writing instruction, they are aware of cultural differences between L1 and L2 writings, can think more critically and write an argument clearly and logically. This affirmed that EFL/ESL students can express critical thinking in English argumentative writing effectively.

## **2.5 Existing critical thinking tests**

This section describes three types of the critical thinking tests; that is, 1) multiple choices tests, 2) open-ended tests and 3) the tests that combine both multiple choices and open-ended format. Limitations of these tests are also discussed.

### **2.5.1 Three types of the existing critical thinking tests**

#### **2.5.1.1 The multiple-choice critical thinking tests**

Traditionally, there are two forms of the critical thinking tests in higher education: the multiple-choice format and the open-ended response format (US. Department

of Education. National Center for Education Statistics, 2000). Examples of the multiple-choice critical thinking tests include the followings:

- California Critical Thinking Skills Test (CCTST),
- Collegiate Assessment of Academic Proficiency (CAAP) Critical Thinking,
- Cornell Critical Thinking Test (CCTT),
- ETS Proficiency Profile (EPP) Critical Thinking, and
- Watson-Glaser Critical Thinking Appraisal (WGCTA) (Liu, Frankel, & Roohr, 2014).

Although the multiple-choice critical thinking tests have a high reliability and an ease of scoring (Liu et al., 2014), only three tests are suitable for college students; that is, California Critical Thinking Skills Test, Cornell Critical Thinking Test and Watson-Glaser Critical Thinking Appraisal (Al-Musaad, 2001). Further descriptions of the multiple-choice critical thinking tests are provided in Appendix 1.

#### **2. 5.1.2 The open-ended critical thinking tests**

Most open-ended critical thinking tests are in the form of essay tests (U.S. Department of Education, National Center for Education Statistics, 2000). Rhetorical tasks for the open-ended tests include narration, description, exposition, argumentation and creative writing. Specifically, for argumentative writing task, many researchers (e.g., Dong, 2015; Saxton, et al., 2012) developed a prompt and a rubric for evaluating critical thinking in English argumentative writing of students. The present study aimed to develop and validate a rubric for assessing critical thinking in argumentative writing. Thus, the strengths and weaknesses of these rubrics are thoroughly discussed in Section 2.5.2.

#### **2.5.1.3 The combination between multiple-choice and open-ended formats**

The current trend of a critical thinking test is combining the multiple-choice format and the open-ended format into one test (Halpern, 2010; Ku, 2009). Examples of such tests include: The Halpern Critical Thinking Assessment Using Everyday Situations (HCTAES) (Halpern, 2007) and Collegiate Learning Assessment + (CLA+) (Council for Aid to Education, 2017). HCTAES evaluates a test taker's abilities to give reasons verbally, analyze an argument, test a hypothesis, use a likelihood and an uncertainty, make a decision and solving problems (Ku, 2009). There are 25 questions, and each question contains both open-ended format and a multiple-choice format (Ku, 2009). Meanwhile, CLA+ is an online test which is comprised of one open-ended response and twenty-five multiple choices. The open-ended section evaluates a test taker's abilities to critique an argument, while the multiple choices section evaluates his/her scientific reasoning and critical reading skills (Zahner, 2013, as cited in Liu et al., 2014).

### **2.5.2 Limitations of the existing critical thinking tests: turning crisis into opportunities**

First, most previous critical thinking tests are costly and not easily accessible. According to many researchers (e.g., Nair, 2011; Yeh, 2001), examples of such tests include: 1) Watson-Glaser Critical Thinking Appraisal (WGCTA) (Watson & Glaser, 1980), 2) Cornell Critical Thinking Test (CCTT) (Ennis, Millman & Tomko, 2005), 3) California Critical Thinking Skills Test (CCTST) (Facione, Facione, Blohm & Gittens, 2008), 4) Ennis-Weir Critical Thinking Essay Test (EWCTET) (Ennis & Weir, 1985), 5) ICAT Critical Thinking Essay Test (Sonoma State University, 1996) and 6) Critical Thinking Test (University of Cambridge, 1996). The educators (Nair, 2011; Yeh, 2001), therefore, suggested that teachers should locally develop and field-test their own rubrics in order to assess the students' critical thinking abilities practically.

Second, the existing standardized tests are summative or one-point-in-time tests, so they only test the students' final product but cannot assess their on-going process of development (Marzano et al., 1993; Bauer, 2016). To promote students' critical thinking abilities, teachers should use a formative assessment because such assessment can 1) diagnose students' levels of critical thinking abilities, 2) provide useful, detailed feedbacks that help students learn their strengths and weaknesses, 3) motivate students to learn and improve critical thinking from the assessment and 4) inform teachers about their success or failure in teaching critical thinking (Bauer, 2016; Ennis, 1993).

Third, using the multiple-choice tests does not promote critical thinking of students due to two reasons. First, the tests cannot analyze reasoning behind a test taker's answer-choice (Davidson & Dunham, 1996), so they cannot provide the students with diagnostic feedback about reasoning skills. Second, the multiple-choice tests are not authentic tasks since it does not reflect the ability to use critical thinking in a real life situation. In most real life situations, there are no choice provided; instead, students need to think and solve problems themselves (Liu et al., 2014). The tests do not provide the students with an opportunity to think critically through doing real world tasks (Alfrey & Cooney 2009; Marzano et al., 1993; Ennis, 1993; Murray, 2016; Bauer, 2016), so it is inappropriate to use such tests for assessing critical thinking of EFL students.

To sum up, the existing critical thinking tests have many limitations, so educators should turn this crisis into opportunity by developing a critical thinking assessment that is valid, reliable and practical and fully promotes student learning and critical thinking.

## **2.6 Principle of assessment and a rubric**

This section explains the principle of assessment which consists of validity, reliability and practicality, addresses a concept of assessment for learning and discusses the components of a scoring rubric.

### **2.6.1 Validity**

Validity refers to a rubric's ability to "actually measure what it is supposed to measure" (Verlinden, 2005, p.293) and provide scoring results that are accurate, appropriate and meaningful (Gronlund, 1998). In developing a critical thinking rubric, a test developer needs to ask experts in the field to validate both content validity and construct validity of a rubric. The next section will briefly discuss content validity and construct validity.

#### **2.6.1.1 Content validity**

In developing a critical thinking rubric, content validity of the rubric must be ensured (Hughes, 2000; Rowicki, 2001). Content validity examines relevance, representativeness and adequacy of the test content or the sample of behavior which the test aims to measure (Akbari, 2012; Messick, 1996a). To achieve the content validity, a test needs to 1) include what students have learned in previous lessons and 2) match with the objective of the course (Brown & Abeywickrama, 2010). For example, if a teacher teaches English reading strategies, and the assignment is based on such strategies, then to be content valid, the test needs to include such strategies (Brown & Abeywickrama, 2010). Also, if a teacher wants to assess students' ability to speak English in a controversial setting, he or she needs to use a speaking test in an authentic context (Brown & Abeywickrama, 2010). Using a paper-pencil makes their test invalid (Brown & Abeywickrama, 2010).

Many researchers in the field of critical thinking and language assessment (e.g., Moskal & Leydens, 2000; Castilleja, 2012; Kaven, 2010; Chatterji, 2003, Daud, 2012;

Lada, 2016; Weigle, 2002) provided the questions as guidance for validating the content validity of an assessment instrument:

- Do all domains of the assessment in a rubric embrace all crucial, relevant aspects of the intent content?
- Do all grading criteria address any extraneous or unrelated content?
- Do the evaluative descriptors clearly represent the domain of the assessment? and
- Are the scoring criteria, the rating scales and the descriptions of each rating scale clear, relevant and appropriate?

#### **2.6.1.2 Construct validity**

The construct is an attribute, a skill or an ability that is processed in humans' brain, such as intelligence, problem solving, creativity and critical thinking (Moskal & Leydens, 2000). The construct is abstract and cannot be observed directly (Moskal & Leydens, 2000), so a rubric for evaluating a construct needs to be developed based on a theory (Akbari, 2012). A test taker's answers will be used as data to make inferences about his or her construct, and such inferences will be used for making an instructional decision (Saxton et al., 2012; Weigle, 2002).

Construct validity refers to the degree to which a rubric "accurately measures the theoretical construct it is designed to measure" (Cozby, 2001, p.374). More specifically, the construct validity examines whether a rubric covers all crucial, relevant aspects of the construct of interest (Moskal & Leydens, 2000). A "scoring guide" or scoring rubric is crucial because it identifies "operational definition of the construct to be measured" (Weigle, 1994, p.205). For example, if a teacher wants to assess a student's oral fluency, then to achieve the construct validity, a scoring rubric needs to include all crucial, relevant aspects of the construct of oral fluency, such as speed, rhythm and (lack of) hesitations (Brown &

Abeywickrama, 2010). To validate the construct validity of a critical thinking rubric, a test developer needs to provide clear, operational definitions of the construct or clear domains of the assessment (Williams, 1999). Also, the experts in the field of critical thinking need to review the rubric carefully to make sure that the rating scales and the descriptors of each rating scale are clear, accurate, adequate and appropriate (Moskal & Leydens, 2000; Williams, 1999). Akbari (2012) stated that a test developer should validate the construct validity of a rubric by using the theory and the statistical techniques, namely factor analysis (Akbari, 2012).

A body of research studies validated both content validity and construct validity of an assessment instrument. For instance, Goodfellow (1999) developed and validated a tool for evaluating critical thinking skills of respiratory therapists in their “clinical respiratory care practice” (p.1). Critical thinking skills in such practice include reflecting on patients, and communicating in a particular situation appropriately, for example (Goodfellow, 1999). Experts were asked to validate the scale by judging the level of importance of each item and then judging whether the items capture the dimensions of the construct (Goodfellow, 1999). Polston (2014) also developed a rubric for assessing college students’ creativity in designing clothes. Experts were asked to validate the rubric by judging whether domains of the assessment, evaluative criteria and levels of performance are clear and appropriate (Polston, 2014). Then, the rubric was revised based on the experts’ comments (Polston, 2014). Most recently, Adunyarittigun (2015) developed a self-perception reading scale based on Bandura’s self-efficacy theory and then asked experts to validate the scale by judging whether each item in the scale captures the dimensions of Bandura’s theory. The scale was subsequently revised based on the experts’ comments (Adunyarittigun, 2015). In light of these studies, it is important to validate both content validity and construct validity of a rubric.

### 2.6.2 Reliability

Reliability refers to a rubric's ability to "consistently give the same results when measuring the same phenomenon" (Verlinden, 2005, p.293). The reliability includes two definitions relevant to critical thinking assessment: reliability across raters (inter-rater reliability) and reliability across time (intra-rater reliability) (Williams, 1999). To achieve the reliability, a test should be conducted under controlled conditions, have items or tasks that are clear to test takers and have a clear scoring rubric (Brown & Abeywickrama, 2010). Also, having a high degree of the reliability will increase confidence in interpreting the scoring results (Stobart, 2012).

A good rubric should have both reliability and validity (Bachman & Palmer, 1996). Ongiem and Vichitvejpaisal (2018) stated that evaluating one's ability reliably and validly by using a scoring rubric is like shooting an arrow to a target accurately. Figure 2.4 illustrates the ability to measure one's ability reliably and validly.

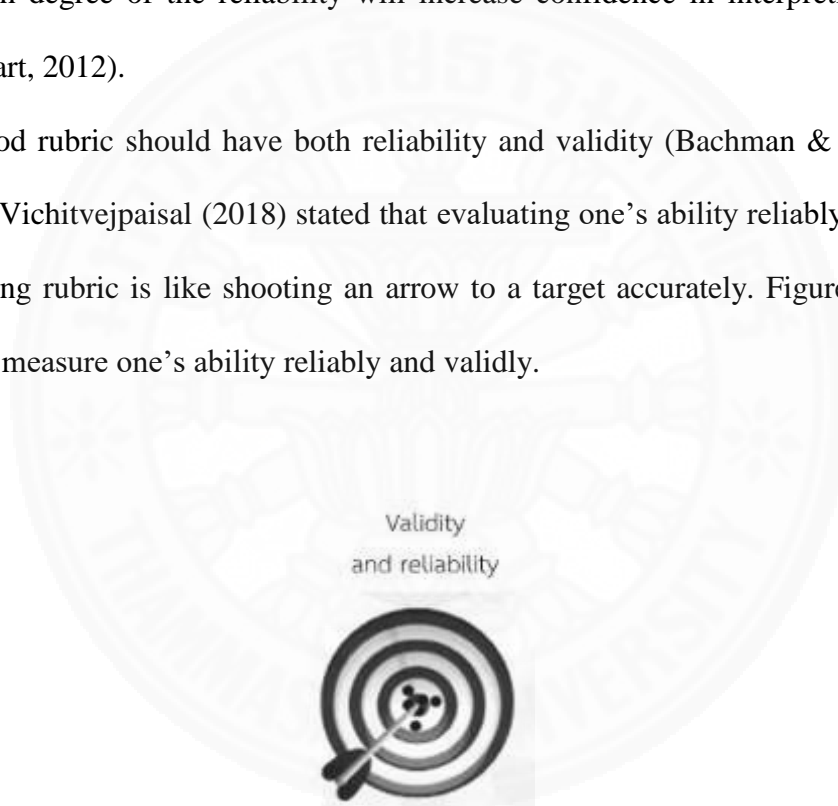


Figure 2.4 Reliability and validity (See Ongiem & Vichitvejpaisal, 2018, p.37)

Four types of inter-rater reliability can be affected by undesirable factors. First, the student-related reliability can be affected by illness, fatigue, a bad day anxiety and other physical or psychological factors that may make an observed score deviate from one's true score (Brown & Abeywickrama, 2010). Second, the inter-rater reliability can be affected by a preconceived bias, a lack of adhering to scoring criteria, inexperience of a rater and a lack of

training. The intra-rater reliability can be affected by unclear scoring criteria, fatigue of a rater, bias toward a particular good and bad student and/or carelessness of a rater (Brown & Abeywickrama, 2010). Third, the test administration reliability can be affected by unreliable conditions in which a test is administered (Brown & Abeywickrama, 2010). For example, in a listening test situation where a teacher plays an audio player as the stimuli, a student sitting next to open windows may not hear the stimuli accurately due to a street noise outside the building (Brown & Abeywickrama, 2010). Such testing condition is unreliable and affects the test administration reliability. Fourth, in a classroom-based assessment, the test reliability can be affected by test items that are ambiguous or have more than one correct answer (Brown & Abeywickrama, 2010). Reliability is a crucial aspect of the rubric, so a test developer should be aware of and try to prevent these undesirable factors when developing or implementing an assessment instrument.

### **2.6.3 Practicality**

A good test should be practical; that is, it should be easy and cheap to construct, administer, score and interpret (Hughes, 2003). According to Brown and Abeywickrama (2010), a practical test has the following qualities: 1) stay within budgetary limit, 2) use an available human resource, 3) have a clear direction for administration, 4) can be completed by students within appropriate time constraints and 5) not require a lot of time and effort in developing a rubric and in a scoring procedure. The practicality is considered as one of the desirable qualities of a test (Hughes, 2003).

### **2.6.4 Assessment of learning versus assessment for learning**

This section discusses the differences between assessment of learning and assessment for learning and suggests how to use the assessment for learning in EFL writing class.

#### **2.6.4.1 The assessment of learning**

In the 20<sup>th</sup> century, a traditional view of testing was affected by a behaviorist learning theory and an objective scientific measurement (Anderson, 1998; Marzano et al., 1993; Shepard, 2000). The traditional paradigm of testing has the following assumptions:

- Knowledge has a single consensus meaning.
- Assessment is viewed as objective, so facts or knowledge should be measured objectively.
- Transfer is limited, so each objective needs to be taught explicitly.
- Learning occurs through associations between a stimulus and a response and through the accumulation of isolated bits of knowledge. A teacher teaches in order to test students' knowledge. Also, he or she uses scoring results to make sure that students have learned and understood the lesson before letting them move on to the next level of learning.
- A teacher is the only person who has a power to make decisions about testing, so students have little or no chance to get involved in testing (Anderson, 1998; Shepard, 2000; Marzano et al., 1993).

The assessment of learning (or summative assessment) comes in the form of a recall, a completion, matching and true-false and multiple choice tests (Marzano et al., 1993). It is used for testing students' skills at the end of a course (Daud, 2012; Shepard, 2000) with the purpose of summarizing "knowledge and skills students have acquired from the instruction" (Katz, 2012, p.67).

#### **2.6.4.2 The assessment for learning**

The assessment for learning (or formative assessment) is the use of "various types of assessment" as "part of the learning process" in order to foster student learning

(Katz, 2012, p.67). The assessment for learning includes dynamic assessment, teacher conferencing, portfolio assessment and self-and-peer assessment (Daud, 2012; Green, 2014). The assessment for learning is influenced by Vygotsky' learning theory (Shepard, 2000). This theory perceives that learning takes place through social interactions and that learning is not a final ended product but an on-going process of development (Shepard, 2000). The learning theory inspires many teachers to change their instruction and assessment by encouraging their students to think critically and do performance-based assessments (Shepard, 2000; Marzano et al., 1993). The assessment for learning can be used as an instructional tool, provides useful, diagnostic feedback to students, helps teachers plan and improve an instruction and helps students be aware of and take more control of their own learning and development (Green, 2014; Stiggins et al., 2004). In conclusion, the assessment of learning (summative assessment) does not promote students' learning, whereas assessment for learning (formative assessment) does (Afflerbach et al., 2015; Neff-Lippman, 2012).

### **2.6.5 How to use the assessment for learning in writing class**

Writing teachers can use assessment to promote student learning by many ways. For example, they can use performance-based assessments by requiring the students to analyze information, apply knowledge and solve problems. In assessing L2 writing, teachers can assess students' ability to utilize a writing process, such as brainstorming, free writing, group-writing, peer-conference and revising. They can also incorporate self- and peer-assessments to encourage the students to learn from each other , learn from the assessment, monitor their own learning and promote independent learning (Brookhart, 2010; Marzano et al., 1993; Shepard, 2000).

### 2.6.6 Rubric

The rubric is a scoring tool. Generally, it contains three crucial components: 1) scoring criteria or domains of the assessment, 2) performance levels or rating scales and 3) specific feedback or evaluative descriptors in each rating scale. An example of an analytic rubric which is adapted from Stevens and Levi (2005) is presented in Table 2.2.

**Table 2.2:** The example of an analytic rubric

Name of the rubric: _____				
Domains of Assessment	Levels of performance			
	Excellent (4 points)	Good (3 points)	Fair (2 points)	Poor (1 point)
Domain of Assessment 1	feedback/ evaluative descriptors	feedback/ evaluative descriptors	feedback/ evaluative descriptors	feedback/ evaluative descriptors
Domain of Assessment 2	feedback/ evaluative descriptors	feedback/ evaluative descriptors	feedback/ evaluative descriptors	feedback/ evaluative descriptors
Domain of Assessment 3	feedback/ evaluative descriptors	feedback/ evaluative descriptors	feedback/ evaluative descriptors	feedback/ evaluative descriptors
Domain of Assessment 4	feedback/ evaluative descriptors	feedback/ evaluative descriptors	feedback/ evaluative descriptors	feedback/ evaluative descriptors

#### 2.6.6.1 Developing a scoring rubric

There are many ways for developing a critical thinking rubric. For instance, according to Reddy (2010), eight steps for developing an analytic rubric are as follows:

- identify learning objectives, scoring method and scoring criteria,
- identify levels of performance,
- develop evaluative descriptors in each performance level,
- obtain feedback about the rubric from others,
- revise the rubric based on the obtained feedback,

- investigate the validity and the reliability of the rubric,
- construct a pilot testing of the rubric, and
- use the results of the pilot study to revise the rubric (Reddy, 2010).

Stevens and Levi (2005) also suggested that teachers should use four

following stages for constructing a rubric:

- Come and discuss class objectives, task objectives, assignment, students' prior knowledge and teachers' previous experience about the assignment.
- Specify what teachers expect to see in students' completed assignment.
- Categorize similar expectations together and identify domains of the assessment clearly.
- Apply the results from the previous stages to design rating scale and evaluative descriptors for each rating scale.

In addition, Daud (2012) proposed a four-step model (See Figure 2.5) for developing a critical thinking rubric. In this model, while a test developer works in one step, he or she can revisit the earlier steps in order to improve the rubric (Daud, 2012).

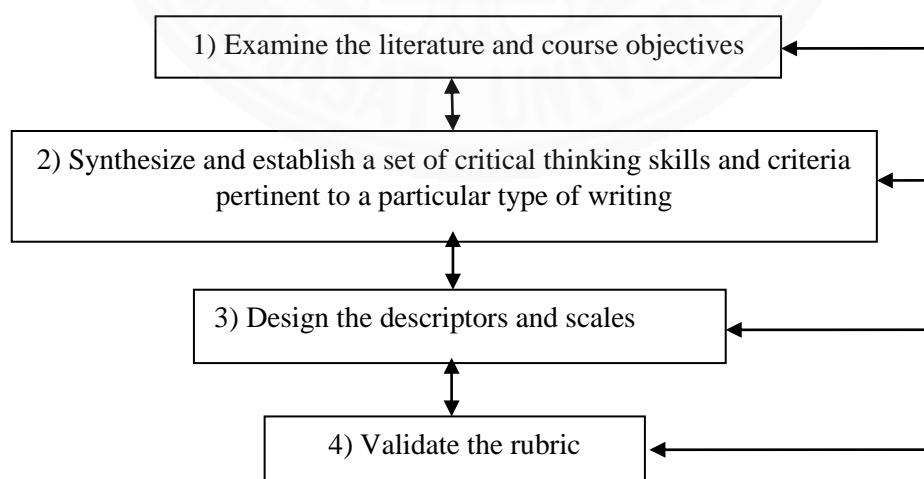


Figure 2.5 Daud's (2012) model for developing a critical thinking rubric

It can be concluded that although there are various models for constructing a rubric, these models include similar crucial stages as follows: 1) review the related literature and/or discussing with other teachers about the assignment and the assessment, 2) identify domains of the assessment, 3) design rating scales and descriptors of each rating scale, 4) examine the validity, reliability and practicality of the rubric and 5) gain feedback from the experts, teachers and/or students to revise the rubric.

## **2.7 Problems of the previous rubrics for assessing critical thinking in students' English argumentative essays**

A rubric for assessing critical thinking in argumentative writings should be clear, valid, reliable and practical, provide the students with useful, diagnostic feedback, and promote student learning and critical thinking. Unfortunately, the previous critical thinking rubrics lack either one or more of these qualities. The next section will describe problems of the previous rubrics. First, it will explain the nature of each type of the critical thinking rubric (i.e., checklist, holistic and analytic). Then, it will describe the previous instruments and discuss their problems.

### **2.7.1 The checklist critical thinking rubric**

#### **2.7.1.1 The nature of the checklist rubrics**

The checklist rubric is one method for evaluating students' writing ability. Two types of a checklist rubric include a simple checklist rubric and a point-allocated checklist rubric (Creighton, 2006). The simple checklist rubric has a short description for each domain of the assessment; however, it has neither levels of performance nor evaluative descriptors in each performance level (Arter & Mctighe, 2001).

Unlike the simple checklist rubric, the point-allocated checklist rubric provides more detailed descriptions about each domain of the assessment (Creighton, 2006). Also, it includes a weighted rating scale; for example, a rater may assess a quality of a

student's essay using a 4-point rating scale (Creighton, 2006). Yet, the point-allocated checklist rubric does not provide evaluative descriptors in each level of performance (Kuhs et al., 2001).

### **2.7.1.2 The previous critical thinking checklist rubrics**

Stapleton (2001a) borrowed some elements of argumentative writing from Toulmin's model and proposed his own checklist rubric for evaluating critical thinking in argumentative essays of Japanese EFL university students. His rubric includes five domains of assessment; namely, 1) an argument (a claim and reasons), 2) evidence, 3) a counter-argument and a refutation, 4) a conclusion (the restated claim) and 5) a fallacy. Likewise, Liu and Stapleton (2014) proposed a simple checklist rubric which includes the following features of critical thinking: claims, data, counter-arguments and rebuttals. The rubric of Stapleton (2001a) and that of Liu and Stapleton (2014) count the frequency of each critical thinking element; however, they do not evaluate students' levels of performance in writing each element of critical thinking.

Liu, Wu and Shieh (2015) also developed a simple checklist rubric for checking the presence of critical thinking elements in argumentative writing of Chinese EFL university students. The investigators used seven higher-order thinking skills from the cognitive dimension of the Revised Bloom's Taxonomy as the domains of critical thinking assessment. Their critical thinking rubric is presented in Table 2.3.

**Table 2.3:** The critical thinking checklist rubric of Liu, Wu and Shieh (2015)

<ol style="list-style-type: none"> <li>1) <u>Remember</u> factual knowledge and/or organize the existence of the debated issues, such as death penalty and same sex marriage</li> <li>2) <u>Understand and explain</u> the concept of the debated topics and opposing viewpoint</li> <li>3) <u>Apply</u> knowledge of the debated issues in real life by using examples to illustrate and support one's own position</li> <li>4) <u>Analyze</u> supporting evidence by differentiating the advantage and disadvantage of one's own and opposing view</li> <li>5) <u>Evaluate</u> or critique the impact of the debated issues on society</li> <li>6) <u>Create</u> (or write) an essay with the components of adequate position statements, reasoning, and supporting evidence/examples</li> </ol> <p style="text-align: right;">(Liu, Wu &amp; Shieh, 2015)</p>
--

For the use of this checklist rubric, students' essays will be analyzed by raters and coded as remember, understand and explain, apply, analyze, evaluate and/or create to identify the critical thinking skills that are present in the essays (Liu, Wu, & Shieh, 2015). Once a level of critical thinking is identified, it is marked as the students can reach that level of critical thinking (Liu, Wu, & Shieh, 2015).

### **2.7.1.3 Problems of the previous critical thinking checklist rubrics**

The rubric of Stapleton (2001) and that of Liu and Stapleton (2014) suffer from two important problems. First, their rubrics merely count the frequency of each critical thinking element, so they can neither evaluate levels of performance of each critical thinking element nor provide the students with diagnostic feedback. According to James and his colleagues (2006), to assess with the purpose of promoting learning, a rubric should have clear domains of the assessment, clear rating scales and clear descriptors in each scale level. Such a complete rubric will provide meaningful score results that help students develop their writing abilities (James et al., 2006).

Second, the previous rubrics overlook some crucial, relevant aspects of critical thinking in argumentative writing. As Moskal and Leydens (2002) stressed, a rubric will achieve both content validity and construct validity if it includes all significant, relevant dimensions of critical thinking. The rubrics of Stapleton (2001) and Liu and Stapleton (2014) do not assess four critical thinking abilities—i.e., 1) writing an introduction to gain readers' attention (which shows the ability to plan, design and create), 2) organizing ideas coherently (which shows the skill of analyzing a structure of an argument), 3) writing a final, relevant, constructive comment in a conclusion (which shows the ability to provide a suggestion, a recommendation or a solution) and 4) using word choices and sentence structures clearly and appropriately. These problems call the validity of the previous rubrics into question.

In addition, the rubric of Liu, Wu and Shieh (2015) merely checks the presence of each critical thinking element in argumentative writing and does not evaluate the quality levels of each critical thinking element. Thus, when receiving score results from this rubric, a student who has a position, and reasons and evidence may feel pleasant to know that his or her essay achieves the ability to create (the highest level of critical thinking abilities). However, he or she will have no ideas of how to further improve the essay because the scoring results do not inform the levels of performance of his/her critical thinking abilities. Thus, it is plausible that a poor or average student cannot make a progress since s/he does not receive diagnostic feedbacks from the rubric. As Condon and Kelly-Riley (2004) stressed, if teachers need to use results from critical thinking assessments as a basis for teaching and promoting students' learning, they need assessments that yield more detailed, not less, information.

To recap, the critical thinking checklist rubrics of the previous researchers (Liu, Wu, & Shieh, 2015; Liu & Stapleton, 2014; Stapleton, 2001a) do not provide sufficient,

diagnostic information to students; therefore, it is inappropriate to use these rubrics with EFL students.

## 2.7.2 The holistic critical thinking rubrics

### 2.7.2.1 The nature of the holistic rubric

The holistic rubric assigns a single score to a text based on the overall impression of the text (Weigle, 2002). Raters will read each text quickly and judge against a rubric right away, so grading students' essays by using the holistic rubric is easy and fast (O'Neill, 2003; Weigle, 2002). According to Matthews (2000), when raters evaluate a large number of students' essays, they often use a holistic rubric because it is practical, easy to use and the most cost effective. An example of a holistic rubric is presented in Table 2.4.

**Table 2.4:** The example of a holistic rubric

Score levels	Domains of assessment and evaluative criteria
Excellent (4 points)	An essay in this level has the following qualities: <ul style="list-style-type: none"> <li>- domain 1 and its detailed descriptions</li> <li>- domain 2 and its detailed descriptions</li> <li>- domain 3 and its detailed descriptions</li> </ul>
Good (3 points)	An essay in this level has the following qualities: <ul style="list-style-type: none"> <li>- domain 1 and its detailed descriptions</li> <li>- domain 2 and its detailed descriptions</li> <li>- domain 3 and its detailed descriptions</li> </ul>
Fair (2 points)	An essay in this level has the following qualities: <ul style="list-style-type: none"> <li>- domain 1 and its detailed descriptions</li> <li>- domain 2 and its detailed descriptions</li> <li>- domain 3 and its detailed descriptions</li> </ul>
Poor (1 point)	An essay in this level lacks the following qualities: <ul style="list-style-type: none"> <li>- domain 1 and its detailed descriptions</li> <li>- domain 2 and its detailed descriptions</li> <li>- domain 3 and its detailed descriptions</li> </ul>

### 2.7.2.2 The previous holistic critical thinking rubrics

The holistic critical thinking rubrics of the previous researchers (e.g., Angeli & Valanides, 2009; Bauer, 2016; Blattner & Frazier, 2002; Facione & Facione, 1996, as cited in Schamber & Mahoney, 2006; Welch, Hieb, & Graham, 2015) assess the following critical thinking elements in English argumentative writings:

- identify a position,
- provide reasons and evidence to support the position,
- state an opposing viewpoint and refute it,
- organize ideas in a logical way,
- use word choices and sentence structures effectively,
- provide a useful suggestion, a recommendation, or a solution,
- use English grammar correctly.

Appendix 2 provides a table summarizing the domains of critical thinking assessment of the previous holistic rubrics, together with the gaps of the rubrics.

### 2.7.2.3 The problems of the previous holistic critical thinking rubrics

The existing holistic rubrics cannot provide an accurate picture of a student's critical thinking ability in argumentative essays. In the existing rubrics, the critical thinking skills were believed to be developed at the same rate and can be graded in a single score (Saxton et al., 2012). However, this assumption is faulty because in reality, a student who is proficient in *giving reasons and evidence to support a claim* may not necessarily be proficient in *using appropriate language* or *organizing ideas coherently* (Saxton et al., 2012). Thus, generalizing a student's critical thinking ability levels to one holistic score represents invalid data (Condon & Kell-Riley, 2004; Haswell, 1991; Saxton et al., 2012). Also, the holistic rubrics cannot be used to promote learning and critical thinking of students. To explain, the

holistic rubrics are a summative assessment (Mearman, 2013) and cannot diagnose problems that EFL students face when writing an argument (Zhao, 2010). For these reasons, the previous critical thinking holistic rubrics are not appropriate tools for use with EFL students.

### **2.7.3 The critical thinking analytic rubrics**

#### **2.7.3.1 The nature of the analytic rubrics**

The analytic rubric provides detailed information about EFL students' writing skills in different aspects, such as ideas/content, organization, vocabulary and grammar (Weigle, 2002). Unlike the checklist and holistic rubrics, the analytic rubric can provide EFL/ESL students with valid, diagnostic information about L2 writing, and many researchers (e.g., Ahour & Mukundan, 2009) recommend using analytic rubrics as a classroom-based assessment. Also, raters who use analytic rubrics can grade essays consistently if they use an anchor paper that shows clearly what essay in each score level looks like (Weigle, 2002).

It is appropriate to use analytic rubric to assess critical thinking in argumentative writing due to following reasons. First, the analytic rubric is a formative assessment because it specifically evaluates each domain of critical thinking and offers detailed, diagnostic feedbacks for each domain (Reddy, 2011). Second, it helps students understand domains of critical thinking assessment clearly and gain detailed, diagnostic feedbacks to improve their critical thinking (Çavdar & Doe; 2012; Marzano et al., 1993; Stevens & Levi, 2005). Third, the analytic rubric can help teachers from different classes maintain standards and consistency in grading and it can also be used a tool for teaching writing to students (Çavdar & Doe, 2012; Marzano et al., 1993; Stevens & Levi, 2005).

#### **2.7.3.2 Problems of the previous critical thinking analytic rubrics**

The existing analytic critical thinking rubrics suffer from two main problems. First, they lack clear rating scales and/or lack clear evaluative descriptors in each scale level.

Second, they ignore some relevant, crucial aspects of critical thinking in argumentative writing. The next section will discuss these problems in details.

### **2.7.3.2.1 The previous rubrics lack clear rating scales and/or lack clear evaluative descriptors in each scale level**

The rubrics of the previous researchers (e.g., Dong, 2015; Finken & Ennis, 1993; Saxton et al., 2016) lack evaluative descriptors that can make a clear distinction between score levels. For instance, in a rubric of Saxton, Belanger and Becker (2016), unclear descriptors in the domain of the assessment, namely Inference, are as follows: “Demonstrates **confident** ability to apply or extend key concepts” and “demonstrates **surprising/insightful** ability to take concepts further” (Score Level 6) vs. “Demonstrates a **clear** ability to apply or extend key concepts” and “demonstrates **strong** ability to take concepts further” (Score Level 5) (p.268, emphasis added). Also, Dong’s (2015) rubric suffers from the same problem. In her rubric, domain of the assessment, namely Depth, contains unclear descriptors as follows: “Implying thoroughness in thinking” (Very Good), “implying depth in thinking” (Good), “not implying enough depth in thinking” (Fair), “not implying depth in thinking” (Poor) and “not implying any depth in thinking” (Very Poor) (Dong, 2015, p.116). In addition, rubrics of Billups (2006), Daud (2012) and McLaughlin and Moore (2012) do not have evaluative descriptors in each rating scale, while Carbone’s (2009) and Osana and Seymour’s (2004) rubrics do not have both rating scales and evaluative descriptors in each scale level. According to many researchers (e.g., Green & Hawkey, 2002; Rakedzon & Baram-Tsabari, 2017; Weigle, 2002), clear rating scale and clear descriptors indicate the validity of the rubric. Thus, lacking these crucial components calls the validity of the previous rubrics into question. Appendix 3 provides a table summarizing the rubrics that suffer from this problem.

### **2.7.3.2.2 The previous rubrics overlook some crucial, relevant aspects of critical thinking in English argumentative essays of EFL students**

The existing rubrics (e.g., Mckitrick, & Barnes, 2012; Brookhart, 2010; Hughes, 2000; Mulnix & Mulnix, 2010) assess related, vital aspects of critical thinking in argumentative writing (e.g., claims, reasons, counter-arguments and conclusion); however, they ignore some relevant, crucial dimensions of critical thinking—i.e., 1) the ability to write an introduction to gain readers' attention, 2) the ability to provide a final, constructive comment in a conclusion and 3) the ability to use word choices and sentence structures that are appropriate and convey meaning clearly. As Green and Hawkey (2002) stated, domains of the assessment are “central to the validity” of a rubric (p.305). Thus, to achieve the validity, a rubric needs to cover all important, related domains of critical thinking assessment (Moskal & Leydens, 2000; Nair, 2011). The previous rubrics do not cover some relevant, crucial aspects of critical thinking, so they lack a construct representation and cannot provide EFL students with diagnostic feedbacks. Appendix 4 presented a table summarizing the criteria of these rubrics.

In conclusion, the previous critical thinking rubrics suffer from many problems. That is, the previous checklist and holistic rubrics cannot provide valid, diagnostic feedbacks to students, while the existing analytic rubrics overlook some relevant, crucial aspects of critical thinking in argumentative writing and contain evaluative descriptors that cannot make a clear distinction between score levels. These threaten the validity of the previous rubrics and make them inappropriate for use with EFL students. Thus, there is clearly a need to develop a valid and reliable critical thinking rubric for use with EFL students.

## **2.8 Chapter summary**

This review examines literature relevant to critical thinking and argumentative writing. The literature reviewed emphasizes the importance of the principles of assessment and discusses the problems of the previous rubrics. The work presented here leads to a need to develop a valid and reliable critical thinking rubric for use with EFL students.



## CHAPTER 3

### RESEARCH METHODOLOGY

Vīmaṃsā: using wise investigation; to diligently apply wise reflection to examine cause and effect within what one is doing and to reflect on, for example, its pros and cons, gains and shortcomings or obstructions. This can be achieved by experimenting, planning and evaluating results, and devising solutions and improvements in order to manage and carry out the activity in hand so as to achieve better results.

(Payutto, 2017, p.38)

#### 3.1. Research design

This study was designed as a mixed methods research which is defined as the class of research where the researcher collects and analyses data, integrates the findings and draws inferences by using both qualitative and quantitative methods in a single study (e.g., Dörnyei, 2007; Johnson & Onwuegbuzie, 2004; Tashakkori & Creswell, 2007). Using both methods enabled the researcher to understand a research problem thoroughly (Creswell, 2014), strengthened the study in terms of data triangulation (Denzin, 1978) and enhanced the validity and the reliability of research findings (Greene, Caracelli & Graham, 1989).

In this study, data were gathered from many sources: (a) the experts' validation survey for the CTER, (b) free interview with the experts and filed notes, (c) argumentative essay samples of EFL students, (d) semi-structure in-depth interview of raters, and (e) open-ended questionnaire and semi-structure in-depth interview of writing teachers. The purposes of this study were to 1) develop the CTER—the analytic rubric for assessing critical thinking in English argumentative essays of EFL university students, 2) examine the validity and the inter-rater reliability of the CTER and 3) investigate the attitudes of the raters and the writing teachers after they use this rubric.

#### 3.2 Research questions

Based on the research purposes above, the present study addresses the following research questions:

(1) Is the CTER a valid and reliable assessment instrument?

(2) What are the attitudes of the raters and the writing teachers after they use the CTER?

### **3.3 Participants and context**

#### **3.3.1 Three experts taking part in the first level of validation of the CTER**

According to Daud's (2012) and Williams' (1999) criterion for choosing experts to validate a critical thinking rubric, the experts need to have in-depth knowledge or expertise in critical thinking and other related fields. Thus, a purposive sampling method was used, so the researcher could obtain the experts that meet this criterion (Cozby, 2001). In the first level of validation of this study, three Thai experts having in-depth knowledge in critical thinking, language assessment, and L2 writing were purposively sampled. The experts came from two leading universities in Bangkok, Thailand. Their pseudonymous names are expert A, expert B and expert C. Expert A was a male university professor having an expertise in language assessment. Expert B was a female university lecturer having an expertise in critical thinking. Expert C was a female university lecturer who had an expertise in L2 writing. The experts were asked to 1) review and critique the first version of the CTER (See Appendix 6) and the detailed descriptions of the six domains (See Appendix 5), 2) initially validate the rubric and 3) share their opinions and comments about revising the rubric.

#### **3.3.2 Five experts taking part in the second level of validation of the CTER**

After revising the rubric based on three experts' comments, the researcher proposed the second version of the CTER (See Appendix 7) and conducted the second level of

validation of the rubric. To do so, the researcher asked experts A, B and C to validate the rubric again and purposively selected the other two Thai experts who had in-depth knowledge about L2 writing. Thus, there were five experts validating the CTER. These experts were asked to review and validate the rubric by using the validation survey (See Appendix 12) and provide constructive comments about the rubric revision. All experts came from three universities in Bangkok, Thailand. Their personal information was pseudonymous to protect their identity. The following section briefly provided their profiles.

### **Expert A**

Expert A was an expert in the field of language assessment. He was a university professor in a public university in Bangkok, Thailand. He earned a Ph.D. in Curriculum and Instruction from a university in USA. He had approximately 20 years of teaching experience. His research expertise lied in the areas of second language reading and writing, reading comprehension strategies, motivation to read, self-perception as readers, language assessment and critical literacy to promote peace.

### **Expert B**

Expert B, an expert in the field of critical thinking, obtained a Ph.D. in Language Education from a university in USA. She worked at a public university in Bangkok, Thailand. She had 13 years of teaching experience. Her research expertise included critical thinking, Teaching English to Speakers of Other Languages (TESOL), Computer Assisted Language Learning (CALL), WebQuests and critical pedagogy.

### **Expert C**

Expert C, an expert in the field of L2 writing, received a Ph.D. in Curriculum and Instruction from a university in USA. She worked at a public university in Bangkok, Thailand and had around 28 years of teaching experience. Her research expertise included reading and process-based writing instruction.

### **Expert D**

As an expert in the field of L2 writing, expert D earned a Ph.D. in Teaching English as a Foreign Language (TEFL) from a university in USA. She was a university professor in the Department of English at a public university, in Bangkok, Thailand. She had more than 35 years of teaching experience. Her research expertise included EFL writing and reading, Teaching English as a Foreign Language (TEFL) and translation.

### **Expert E**

As an expert in the field of teaching and writing, Expert E obtained a doctorate in English from a university in USA. She worked at a public university in Bangkok, Thailand. She had more than 30 years of teaching experience. Her research expertise included TEFL, L2 writing and reading, English vocabulary instruction and translation.

### **3.3.3 Three raters who implement the CTER**

After revising the rubric based on the experts' comments, the researcher proposed the final version of the CTER (See Appendix 6) and asked three Thai experienced raters to use this rubric to grade the first set of writing samples (10-15 argumentative essays) and to share their opinions about using six domains to assess critical thinking in writing of EFL students. This enabled the researcher to investigate the validity and practicality of the rubric. These raters were purposively sampled based on a criterion suggested by Gebril and Plakans (2014) and Qin (2009). That is, the raters needed to have expertise in English language teaching, language assessment and/or argumentation (Gebril & Plakans, 2014; Qin, 2009). The raters' names were pseudonymous to protect their identity. Their profiles were briefly presented in the following section.

### **Rater 1**

Rater 1 was a female lecturer teaching English at a public university in Thailand. She obtained a master's degree in English and was a doctoral student in the English Language Studies Program (Concentration on language assessment) in a university in Thailand. She had experience in teaching and grading English argumentative writing of Thai undergraduates. Her current research interest included self-and peer-assessment, learner autonomy and higher-order thinking.

### **Rater 2**

Rater 2 was a female lecturer teaching English at a university in Thailand. She earned a master's degree in the English Language Studies Program (Concentration on language assessment) from a university in Thailand. She taught English reading and writing to Thai undergraduate students. She had an expertise in language assessment. Her research interest included methodology in English language teaching and language assessment.

### **Rater 3**

Rater 3 was a female doctoral student in Global Study Program in a university in China. She earned master's degree in communication (Concentration on critical discourse analysis) from a university in Thailand. She was interested in using a rubric to foster student learning and critical thinking. Her research interest included (critical) discourse analysis and argumentation.

#### **3.3.4 Three writing teachers who implement the CTER**

To investigate the validity and the inter-rater reliability of the CTER, the researcher purposively chose three Thai experienced writing teachers and requested them to use the final version of the CTER to grade the writing samples (20 argumentative essays) and to share their opinions about the rubric. Their pseudonymous names are writing teacher 1, writing

teacher 2 and writing teacher 3. These teachers came from three different universities in Thailand. Their profiles were briefly provided as follows:

### **Writing Teacher 1**

Writing teacher 1 earned a Ph.D. in English Language Studies (Concentration on Applied Linguistics) from a university in Bangkok, Thailand. She had 10 years of teaching experience. Currently, she worked at a public university in Bangkok and taught reading and writing and English public speaking to Thai university students. Her current research interest included project-based learning, critical thinking and critical reading.

### **Writing Teacher 2**

Writing teacher 2 earned a master's degree in English (Concentration on English Writing) from a university in Thailand. Currently, she was an Assistant Professor in the Department of English at a public university. She had 20 years of teaching experience. Her research expertise lies in the areas of second language writing and translation.

### **Writing teacher 3**

Teacher 3 earned a Ph.D. in Applied Linguistics (Concentration on English Writing) from a university in Thailand. Her research interest included English writing, ESP, English for Academic Purposes and self-and peer assessment. She taught critical thinking and language use, argumentative writing and critical reading to Thai university students. She had 18 years of teaching experience.

## **3.4 Ethic consideration**

Ethical issues need to be addressed if a participant was a human being (Lada, 2015). The participants of this study were humans; therefore, after the proposal defense, the researcher submitted all required ethical documents to the Human Ethics Committees of

Thammasat University. The documents included the research proposal, the CTER, detailed descriptions of each domain of the assessment, the experts' validation survey for the rubric, the writing samples, a consent form and information sheet. The committees examined crucial issues related to justice, risks, benefits, confidentiality and the rights of the participants (Lada, 2015), and they gave the researcher permission to conduct this research.

The researcher sent a letter to request permission from the participants to conduct this research with them. They were informed about the details of this study and the ethical issues before they decided to participate in this study. Also, they were presented with the information sheets and consent forms and were required to read and complete the form in case that they were willing to be the participants. In addition, they were informed that they could withdraw from the study whenever they wanted and that their personal information would be kept confidential in order to protect their identities.

### **3.5 Research instruments**

This study used the following instruments: 1) the experts' validation survey for the CTER, 2) face-to-face, free interview with the experts and the field notes, 3) the argumentative writing samples, 4) the raters' in-depth interview asking about their perceptions about the rubric and 5) the writing teachers' open-ended questionnaire and in-depth interview asking about their attitudes towards the rubric. The following section described the details of these instruments.

#### **3.5.1 Experts' validation survey for the second version of the CTER**

The validation survey (See Appendix 12) were used in the second level of validation of the rubric. The survey allowed experts A, B, C, D and E to validate both content validity and construct validity of the CTER. The survey were developed from two main sources; that is, 1) guidance for validating a critical thinking rubric suggested by many scholars (e.g.,

Daud, 2012; Lada, 2015; Moskal & Leydens, 2002; Williams, 1999) and 2) guidance for validating an analytic rubric suggested by lots of researchers (e.g., Alsina et al., 2017; Castilleja, 2012; Green & Hawkey, 2002; Kaven, 2010; Purpura, 2004; Weigle, 1994, 2002, 2007). The validation survey was examined by experts A, B and C to ascertain its content validity.

The validation survey contained four parts which were designed in the form of five-point Likert Scale ranging from *Strongly Agree* to *Strongly Disagree*. In part 1, the experts were asked to rate how strongly they agreed upon whether the six domains of the assessment are valid and clear indicators for assessing critical thinking in English argumentative essays of EFL university students. In part 2, they were asked to rate how strongly they agreed upon whether the descriptions of each domain are valid and clear. In part 3, they were asked to rate how strongly they agreed upon whether the evaluative descriptors in each level of performance are valid and clear. In part 4, they were asked to rate how strongly they agreed upon whether the evaluative descriptors of each domain can make a clear distinction among the four score levels.

### **3.5.2 Face-to-face, free interview with the experts and field notes**

When receiving the validation results from the experts, the researcher conducted the face-to-face, free interviews with them. As Bolderston (2012) stated, the interview enabled the researcher to probe or clarify important issues that are raised. In this study, the researcher went to receive the validation results from the experts by himself; therefore, he had an opportunity to discuss the rubric with the experts and gain their constructive comments about the rubric through using the interview. Besides, the researcher used field notes as the other instrument for collecting the data. According to Merriam (2015) and Patton (2002), while interviews and written texts can be used to analyze people's thought, observation and field notes can be used to analyze people's behaviors and/or conversations. Thus, during the

interview, the researcher of the current study recorded the experts' verbal comments about the CTER in the personal notes in order to gain detailed descriptions of the conversation.

### **3.5.3 English argumentative writing samples of Thai EFL students**

The researcher used the two sets of the students' argumentative essays as the writing samples. As Rowicki (2008) stated, students' argumentative essays can be used as the writing samples for examining the validity and the inter-rater reliability of a critical thinking rubric.

The first set of the writing samples was 37 second-draft argumentative essays of Thai third-year English-major students from a public university in Bangkok, Thailand. The researcher asked the raters to use the CTER to grade these samples and share their opinions about the rubric. The researcher asked for these samples from one Thai professor who taught English composition to Thai university students. That professor provided the researcher with 37 essays written by Thai students. She taught these students in the course Composition II in the academic year of 2008. At that time, she allowed the students to choose a topic of writing freely, so the topics of the essays varied and could be categorized into the six groups, i.e., technology, ethics, health, education, media and others. The length of their essays was about 2-3 pages long (about 600 words per one text). Students' names and surnames were anonymous to protect their identities.

The second set of the writing samples was 31 argumentative essays of 31 Thai EFL university students who studied English composition in the academic year of 2016. The researcher asked three writing teachers in the study to use the CTER to grade these samples in order to investigate their opinions about the rubric and examine the inter-rater reliability of the scoring instrument. The researcher obtained these samples from two people—i.e., expert C and one native English instructor teaching English composition at a public university in Bangkok, Thailand. To explain, the researcher initially asked for 16 essays from expert C and

then asked for the other 15 essays from the native English instructor. All students' personal information was anonymous to protect their identities. The rationale for using the new set of the writing samples was that the first set of the samples came from only one group of Thai EFL students. Also, these students studied English composition in 2008, so their texts were quite old. Thus, to ensure the validity, the inter-rater reliability and the practicality of the CTER, it was useful to use the more updated writing samples.

### **3.5.4 Interview questions for the raters and the writing teachers**

According to Stake (1995), interview enables the researcher to gain realities from multiple dimensions. The researcher of this study used five semi-structure, in-depth interview questions (See Appendix 13) to investigate the raters' and the writing teachers' attitudes towards the CTER. The interview questions were adapted from Kaven's (2010) and Zhao's (2010) interview questions for raters in the process of validating an analytic rubric. During the interview, the researcher asked the participants some probing questions in order to gain their in-depth opinions about the rubric and to help them clarify their ideas (Merriam, 1998). Conversation during the interview was recorded, transcribed and then translated into English by the researcher.

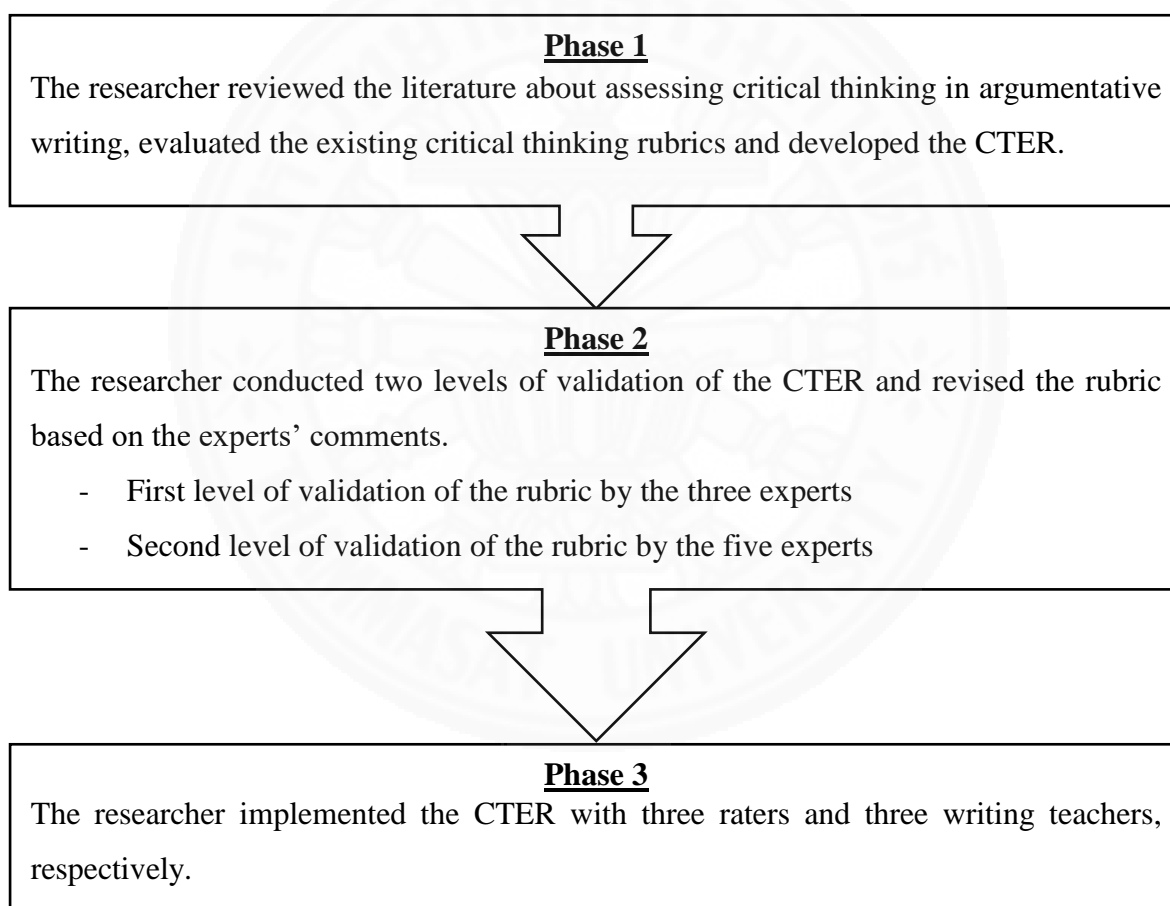
### **3.5.5 Open-ended questionnaire for the writing teachers**

This study also used the open-ended questionnaire (See Appendix 14) to gain the writing teachers' additional comments about the CTER. This questionnaire was adapted from Kaven's (2010) and Zhao's (2010) interview questions for raters in the process of the rubric validation. The researcher of the present study provided the participants with "the flexibility and time to provide answers reflecting their thoughts" about the rubric (Tulgar, Yagiz, & Han, 2017, p. 65). The questionnaire was examined by Experts A, B and C to ensure its content validity. The data from the questionnaire were used to triangulate with the data from the interview to ascertain the consistency of the findings.

### 3.6 Research procedures

#### 3.6.1 The process of developing and validating the CTER

The process of developing and validating the CTER required time, effort and assistance from a number of people. The researcher of this study spent five years constructing the CTER (February of 2015 – February of 2019). The process for developing and validating this rubric was adapted from guidance for developing a critical thinking rubric suggested by many researchers (e.g., Daud, 2012; Reddy, 2011; Stevens & Levi, 2005). Overview of the research phases for developing the CTER were illustrated in Figure 3.1.



**Figure 3.1** Overview of the research phases for developing the CTER

#### **The development of the CTER**

Literature in the realms of English argumentative writing, language assessment, critical thinking theories and critical thinking assessment in argumentative writing were

extensively reviewed in order to analyze the existing critical thinking rubrics and identify aspects of critical thinking, domains of assessment and strengths and weaknesses of the existing rubrics. A table summarizing domains of critical thinking assessment from the previous rubrics was presented in Appendix 4. This table serves as “a comprehensive database” that shows the critical thinking elements that are assessed and overlooked by the previous rubrics (Hodges et al., 2019, p.10).

The researcher developed the CTER from reviewing the theoretical frameworks of critical thinking, argumentative writing, assessment and all of the previous critical thinking rubrics. According to Liu and his colleagues (2014), reviewing the related literatures about critical thinking assessment helps the researcher learn and be able to propose valid and clear operational definitions of critical thinking. Also, as Nair (2011) suggested, a valid rubric assesses exactly what it proposes to assess, cover all crucial, relevant aspects of critical thinking and does not measure unrelated variables. Thus, reviewing the related literatures enables the researcher of the present study to assess what it is proposed to assess (i.e., critical thinking in English argumentative writing) and initially ensure that the CTER has construct representation and construct relevance.

The CTER contains six crucial aspects of critical thinking assessment; namely, 1) the ability to gain readers’ attention in the introduction, 2) the ability to give reasons and evidence to support a claim, 3) the ability to state a counter-argument and a refutation, 4) the ability to write a conclusion, 5) the ability to organize ideas in a logical way and 6) the ability to use word choices and sentence structures effectively. The researcher used four rating scales ranging from Excellent (4 points), Good (3 points), Fair (2 points) to Poor (1 point) because the four levels of performance can make a clear distinction among the score levels (Chan, Inoue & Taylor, 2015; Janssen, Meier & Trace, 2015; Penny, Johnson & Gordon, 2000). Detailed descriptors of the CTER were presented in Appendices 5 through 6.

### **The first level of validation of the CTER**

The researcher asked three experts (experts A, B and C) to review the CTER thoroughly around three weeks and initially validate it by judging the accuracy, clarity and appropriateness of the six domains of the assessment and the evaluative descriptors in each score level. To obtain their opinions about the CTER, the researcher conducted a group, face-to-face, free interview with three experts at one university in Bangkok, Thailand. It took around 1 hour to interview the experts, and the experts' comments about the rubric were written in the field notes. The rubric was subsequently revised based on their comments, and the revisions made the CTER clearer.

### **The second level of validation of the CTER**

The researcher administered the revised version of the CTER to experts A, B, C, D and E and asked them to review the CTER, validate this rubric by using the validation survey and submit the validation results to the researcher. To gain their constructive comments about the CTER, the researcher conducted the individual, face-to-face, free interview with the experts and recorded their comments in the field notes. It took around 30 – 50 minutes to interview each expert. The researcher subsequently revised the rubric based on the experts' comments.

### **Implementing the CTER with the raters**

To ensure the validity and the practicality of the CTER, the researcher asked raters A, B and C to use the rubric to grade 15 essays randomly selected from the first set of the writing samples and share their opinions about using the six domains for assessing critical thinking in argumentative writing of EFL students.

Before having the raters implement the rubric, the researcher conducted a three-stage group training with the raters. This rater training took place at a public university in Bangkok,

Thailand, and it took around two hours to train the raters. The training process was adapted from that suggested by many researchers (e.g., Gebril & Plakan, 2014; Qin, 2009; Rowicki, 2008). Each training stage was briefly explained in the followings.

In the first stage of training, three raters were provided with the CTER and received thorough descriptions about each domain of the assessment. The researcher also provided them with the anchor papers (See Appendix 10) and requested them to read the anchor papers thoroughly, so they could understand what essay in each score level looked like and grade the writing samples consistently. Then, the researcher randomly selected 2 from 32 essays. The essays were graded by the raters and the researcher. During the grading, the raters could discuss with the researcher or with each other. For example, they could ask questions about the domain of the assessment and evaluative descriptors in each score level. This helps the raters be familiar with the rubric and have the same concept of scoring. After the grading, the raters and the researcher discussed, compared and contrasted the scoring results with each other (Chase, 2011; McNamara, 2000). During training, it is hard to get a perfect inter-rater reliability (Douglas, 2010). Therefore, the ratings which had less than a 1-point difference were acceptable, whereas the ratings that had more than a 1-point difference were not acceptable (Huang, 2007; Zhao, 2010; Penny, Johnson & Gordon, 2000). The raters needed to discuss the scoring results until the consensus was reached (Huang, 2007; Zhao, 2010).

In the second stage of training, the researcher randomly selected 3 essays. At this time, the raters and the researcher graded these samples individually. While grading the essays, the raters were not allowed to ask the researcher questions or discuss with each other. After the grading, the scoring results were discussed, compared and contrasted. If there was more than a 1-point difference, and if the raters could not come to an agreement, the researcher of this study would read that essay, listened to the rationale from both raters and then made a decision by himself (Gustafson & Bonchner, 2009; Huang, 2007).

In the last stage of training, the researcher randomly selected 10 essays, and three raters and the researcher graded these samples independently. If the raters felt tired, they could take a break, so the tiredness could not affect their scoring. After the grading, the raters and the researcher discussed the scoring results with each other.

After the group training finished, the raters were asked to review the CTER again at their homes for two days in case that they had additional comments about the rubric. Two days later, the researcher conducted an individual, face-to-face interview with the raters by using the semi-structure interview questions (See Appendix 13) in order to investigate their attitudes towards the rubric.

### **Implementing the CTER with the writing teachers**

To examine the validity and the inter-rater reliability of the CTER, three writing teachers were asked to use the rubric to grade 26 argumentative essays. The researcher selected these 26 essays from the second set of the writing samples by using simple random sampling. At first, the researcher planned to conduct the group training on using the rubric with the teachers. Unfortunately, at that period of time, all teachers had their urgent works and had different schedules, so they were not convenient to come and receive the training together on the same date. The researcher, therefore, trained the teachers individually with the rigorous and well-planned training procedure suggested by Gebril and Plakan (2014), Qin (2009) and Rowicki (2008). The following section described the three-stage training process that was used with each writing teacher.

In the first stage of training, the researcher explained the CTER and the anchor papers to the writing teachers thoroughly. Any questions about the CTER were discussed. After that, the researcher randomly selected 3 out of the 31 essays. These essays were graded by the teachers and the researcher. During the grading, the teachers were not allowed to ask

questions or discuss with the researcher. This helped the researcher ensure that the teachers clearly understood the rubric, had the same concept in scoring and could grade the essays consistently. The ratings that fell within a 1-point difference were acceptable, whereas the ratings that had more than a 1-point difference were not acceptable (Becker, 2016; Huang, 2007; Zhao, 2010). Also, the teachers needed to discuss the scoring results with the researcher until the consensus was reached (Huang, 2007; Zhao, 2010).

In the second stage of training, the researcher randomly selected 3 essays. These essays were graded by the teachers and the researcher. While grading the essays, the teachers could take a break whenever they wanted in order to prevent the tiredness from affecting their scoring results. After the grading, the teachers and the researcher discussed the scoring results with each other to ensure that the scoring results are reliable.

In the final training stage, the researcher randomly selected 20 essays and asked the teachers to grade these essays carefully. It is acceptable to use the 20 essays for examining the inter-rater reliability of a critical thinking or argumentative writing rubric (Rowicki, 2001; Qin, 2009). In this stage, the researcher allowed the writing teachers to grade these essays according to their own schedules and asked them to submit the scoring results via an email within three weeks. Besides, to facilitate the grading and to ensure the consistency in scoring, the researcher provided each teacher with the rating manual that contained thorough descriptions of the CTER (See Appendices 5 and 9), including the anchor papers (See Appendix 10). Also, in case of having questions about the CTER, the writing teachers were allowed to ask the researcher via an email for clarification.

After all teachers submitted the scoring results of the 20 essays samples, the researcher conducted the individual, face-to-face interview with them to gain their opinions about using the six domains for assessing critical thinking in argumentative writing of EFL students. Also, the researcher asked them to review and consider the CTER thoroughly again

at their homes for one week, and if they popped up any useful or interesting ideas in mind about the rubric, they could write their comments in the open-ended questionnaire (See Appendix 14). The data from the questionnaire were used to triangulate with the data from the in-depth interview to ensure the consistency of the findings.

To investigate the inter-rater reliability of the CTER, the scoring results of the 20 argumentative writing samples obtained from three writing teachers and the researcher were analyzed by using two statistical measures, namely Kendall's coefficient of concordance (W) and Intra-class correlation coefficient (ICC). To ensure the credibility of the findings, the researcher implemented two statistical approaches to examine the inter-rater reliability of the rubric. The criteria for judging the inter-rater reliability of the CTER were discussed in the following section.

### **3.7 Data collection and data analysis for each research question**

This section described the data collection techniques in relation to each research question and reported how the data were analyzed. The following data sources were included for triangulation purposes: 1) the experts' validation survey for the CTER, 2) the face-to-face, free interview with the experts and the field notes, 3) the argumentative writing samples, 4) the raters' interview asking about their perceptions about the rubric, and 5) the writing teachers' interview and the questionnaire asking about their attitudes towards the rubric. Table 3.2 summarized the data sources and the data analysis to address each research question in the study.

#### **Research question 1: Is the CTER a valid and reliable assessment instrument?**

This research question was an attempt to investigate whether the CTER achieved the validity and the inter-rater reliability. To answer this research question, the researcher triangulated the data from multiple sources; that is, 1) the quantitative results from the experts' validation survey, 2) the qualitative results from the field notes for describing the

experts' opinions and comments about the rubric, 3) the qualitative results from the raters' interview and the writing teachers' questionnaire and interview asking about their attitudes towards the rubric and 4) the scoring of the 20 writing samples from the writing teachers and the researcher.

To examine whether the CTER is a valid assessment instrument, the researcher conducted the two levels of validation of the rubric. In the first level of validation, the researcher asked experts A, B and C to review and validate the CTER (See Appendix 6) and obtained their comments about the rubric by using the face-to-face, group interview and the field notes. Thematic analysis method was used to analyze the qualitative data. Also, since the development of a valid critical thinking rubric required on-going revisions (Niedringhaus, 2001), the researcher revised the rubric based on the experts' comments.

In the second level of validation of the rubric, the researcher asked experts A, B, C, D and E to review the revised version of the CTER (See Appendix 7) and validate this rubric by using the validation survey (See Appendix 12). The quantitative data from the completed validation survey was performed by using Statistical Packages for the Social Sciences (SPSS) to compute mean ( $\bar{x}$ ) and standard deviation (S.D.). The quantitative data analysis allowed the researcher to gain a confidence level of mean for the six domains of the assessment and the evaluative criteria in each score level. The interpretation of mean, which was presented in Table 3.1, was adapted from Adunyarittigun's (1998, 2015) and Whittenburg's (1995) criteria. That is, for a domain or a descriptor to be valid, the confidence level of mean had to meet or exceed 3.50; a domain or a descriptor not meeting the criteria had to be revised based on the experts' comments (Adunyarittigun, 1998, 2015; Whittenburg, 1995). In Table 3.1,  $\bar{x}$  refers to the average of level of confidence of all experts towards a domain of the assessment or an evaluative descriptor.

**Table 3.1:** Criteria for interpreting the mean score from the experts' validation survey

$\bar{x}$	Interpretation of the confidence level of $\bar{x}$
<b>4.50-5.00</b>	The experts strongly agree that a domain or a descriptor is valid. Thus, such domain or descriptor is retained.
<b>3.50- 4.49</b>	The experts agree that a domain or a descriptor is valid. Thus, such domain or descriptor is retained.
<b>2.50-3.49</b>	The experts are undecided if a domain or a descriptor is valid. Thus, such domain or descriptor is revised based on the experts' comments.
<b>1.50-2.49</b>	The experts disagree that a domain or a descriptor is valid. Thus, such domain or descriptor is revised based on the experts' comments.
<b>1.00-1.49</b>	The experts strongly disagree that a domain or a descriptor is valid. Thus, such domain or descriptor is revised based on the experts' comments.

Besides, in the second level of validation of the rubric, the researcher conducted free, individual interview with five experts in order to receive their constructive comments about the CTER. During the interview, the researcher wrote the experts' verbal comments in the field notes and analyzed the data from the field notes by constructing categories or themes that captured recurring patterns (Merriam, 1998, 2015). Then, the researcher revised the rubric based on the experts' comments and implemented the rubric with the raters and the writing teachers, respectively.

To answer the first research question, the researcher also interviewed two groups of the participants (the raters and the writing teachers) in order to gain their opinions about using the six domains for assessing critical thinking in argumentative writing. The conversation during the interview was audio recorded, transcribed and then translated into English by the researcher. Their opinions were analyzed by using the thematic analysis method (Merriam, 1998, 2015), and the qualitative data were used as evidence for confirming the validity of the CTER.

In sum, to answer whether the CTER is a valid assessment, the researcher triangulated the qualitative data from the experts' free interview and field notes with the quantitative data from the experts' validation survey, including the qualitative data from the raters' interview and the writing teachers' questionnaire and interview. The qualitative data were analyzed to

triangulate with the quantitative data analysis and provide additional evidence on the validity of the rubric (Zhao, 2012).

To investigate the inter-rater reliability of the CTER, the researcher requested the writing teachers to use the CTER to grade the 20 writing samples of Thai EFL college students. Two statistical approaches; namely, 1) Kendall's Coefficient of Concordance (W) and Intraclass Correlation Coefficient (ICC) were used to analyze the reliability among four assessors (i.e., three writing teachers and the researcher himself). The following sections briefly described the two statistical measures.

### **1) Kendall's Coefficient of Concordance (W)**

According to Huck (2012), Kendall's W procedure was suitable for situations where each rater was asked to evaluate things by using a rating scale. The current study used Kendall's W procedure suggested by Schmidt (1997) to investigate the inter-rater reliability of the CTER. The theoretical limits of a data-based coefficient of Kendall's W were -1.00 to +1.00; therefore, the Kendall's W criteria for judging the inter-reliability of the rubric were as follows: 0.1=Very weak agreement, 0.3=Weak agreement, 0.5=Moderate agreement, 0.7=Strong agreement and 0.9=Unusually strong agreement (Schmidt, 1997).

### **2) Intraclass Correlation Coefficient (ICC)**

ICC can be used to analyze both inter-rater and intra-rater reliability of an assessment instrument (Koo & Li, 2016; Huck, 2012). Also, it can be used in a situation where each essay was rated by many raters (Schley & Albertini, 2005). The theoretical limits of the data-based coefficient of ICC were 0 to 1.00, so to confirm the inter-rater reliability of the rubric, the researcher should gain a value that is as close to 1.00 as possible (Huck, 2012). In the present study, the ICC criteria for judging the inter-rater reliability of the rubric are as follows: greater than 0.5 = poor, 0.5 - 0.75 = moderate, 0.75 - 0.90 = good, and less than 0.90 = excellent (Koo & Li, 2016).

In the present research, Kendall's *W* approach was used to investigate the inter-rater reliability of each domain of the assessment and investigate the inter-rater reliability of the total scoring results of the six domains. Meanwhile, ICC approach was used to examine the inter-rater reliability of the total scoring results from the six domains. Thus, ICC was used to triangulate with the Kendall's *W* in order to strengthen the validity of the research findings.

To recap, the investigator of this study conducted an on-going process of developing, validating and revising the CTER, and there were three research phases. Thus, to answer whether the CTER is a valid assessment, the qualitative and quantitative data obtained from the experts participating in the two levels of validation were used to triangulate with the qualitative data from the raters' interview and the writing teachers' the questionnaire and the interview. The illustration of how the revisions increased the validity of the CTER was also presented. Besides, to answer whether the CTER is a reliable assessment, the writing teachers' and the researcher' raw scoring results of the 20 writing samples were analyzed by Kendall's *W* and ICC, and the quantitative data were reported as the evidence of the inter-rater reliability of the CTER.

**Research question 2: What are the attitudes of the raters and the writing teachers after they use the CTER?**

To answer this research question, the researcher conducted an in-depth analysis of the raters' and the writing teachers' attitudes towards the CTER by using the semi-structure in-depth interview (See Appendix 13). In this study, the attitudes were defined as "an evaluative reaction" to a rubric based on the participants' "beliefs or opinions" (Gardner, 1985, p.9). The conversation during the interview was audio recorded, transcribed and then translated into English by the researcher. The open-ended questionnaire (See Appendix 14) was also used to further investigate the writing teachers' perceptions about the CTER. The data from the

raters' interview were used to triangulate with the data from the teachers' the questionnaire and the interview in order to strengthen the credibility of the findings.

The results from the semi-structure interview and the open-ended questionnaire were analyzed by the qualitative analysis method, namely a thematic analysis. According to Maxell (1994), the data analysis searched for themes and was grounded in the data. The researcher was the primary instrument for collecting and analyzing the data by constructing themes that captured the recurring patterns (Merriam, 1998). Khaimukd (1999) metaphorically said that analyzing the qualitative data was like "playing a jigsaw" puzzle in that the player (the researcher) needed to put small pieces together, and the "small piece, when put together," would "create a big picture" (p.66). She (1999) further added that in the thematic analysis, "there was no model provided for putting the pieces together. The player of the categorizing game had to figure out the big picture by him or herself. The big picture, thus, was derived from the player's own interpretation" (p.66).

Following the advices of the qualitative researchers, such as Maxell (1994), Merriam (1998) and Khaimukd (1999), the investigator of this study read each piece of data thoroughly, looked at the themes that reflected the perceptions of the raters and the writing teachers and then noted the themes. Also, while analyzing the data, the researcher saw more themes emerge (Khaimukd, 1999). Some themes were repetitive, whereas some themes could be transformed into new themes (Khaimukd, 1999). Thus, to analyze the data effectively, the researcher listed all emerging themes, noted the properties of each theme and categorized and clustered the themes (Khaimukd, 1999; Thongrin, 2002).

To answer the second research question, the data from the questionnaire were used to triangulate with the data from the interview in order to strengthen the reliability of the findings. Table 3.2 summarized the data sources and data analysis for each research question.

**Table 3.2:** A summary of the data sources and data analysis for each research question

<b>Research Question</b>	<b>Data Source</b>	<b>Data Analysis</b>
<b>(1) Is the CTER a valid and reliable assessment instrument?</b>	<ul style="list-style-type: none"> <li>• The experts' validation survey for the CTER</li> <li>• Free interview with the experts and field notes</li> <li>• The scores obtained from marking 20 writing samples by the writing teachers and the researcher</li> <li>• The raters' interview asking about their opinions about using the six domains for assessing critical thinking in writing</li> <li>• The writing teachers' the questionnaire and the interview asking about their attitudes towards using the six domains for assessing critical thinking in writing</li> </ul>	<ul style="list-style-type: none"> <li>• Statistical analysis of the experts' validation survey of the CTER</li> <li>• Thematic analysis of the data obtained from the interview with the experts and filed notes</li> <li>• Kendall's W and ICC for investigating the inter-rater reliability of the CTER</li> <li>• Thematic analysis of the data obtained from the raters' interview and the writing teachers' the questionnaire and the interview asking about their perceptions about the validity of the CTER</li> </ul>
<b>(2) What are the attitudes of the raters and the writing teachers after they use the CTER?</b>	<ul style="list-style-type: none"> <li>• The raters' and the writing teachers' in-depth interview asking about their attitudes towards the CTER</li> <li>• The writing teachers' open-ended questionnaire asking about their perceptions about the CTER</li> </ul>	<ul style="list-style-type: none"> <li>• Thematic analysis of the questionnaire and the interview</li> </ul>

### 3.8 Chapter summary

This chapter presents a description of the participants, research instruments and research procedures. Data collection and data analysis are also presented.

## CHAPTER 4

### RESULTS AND DISCUSSION

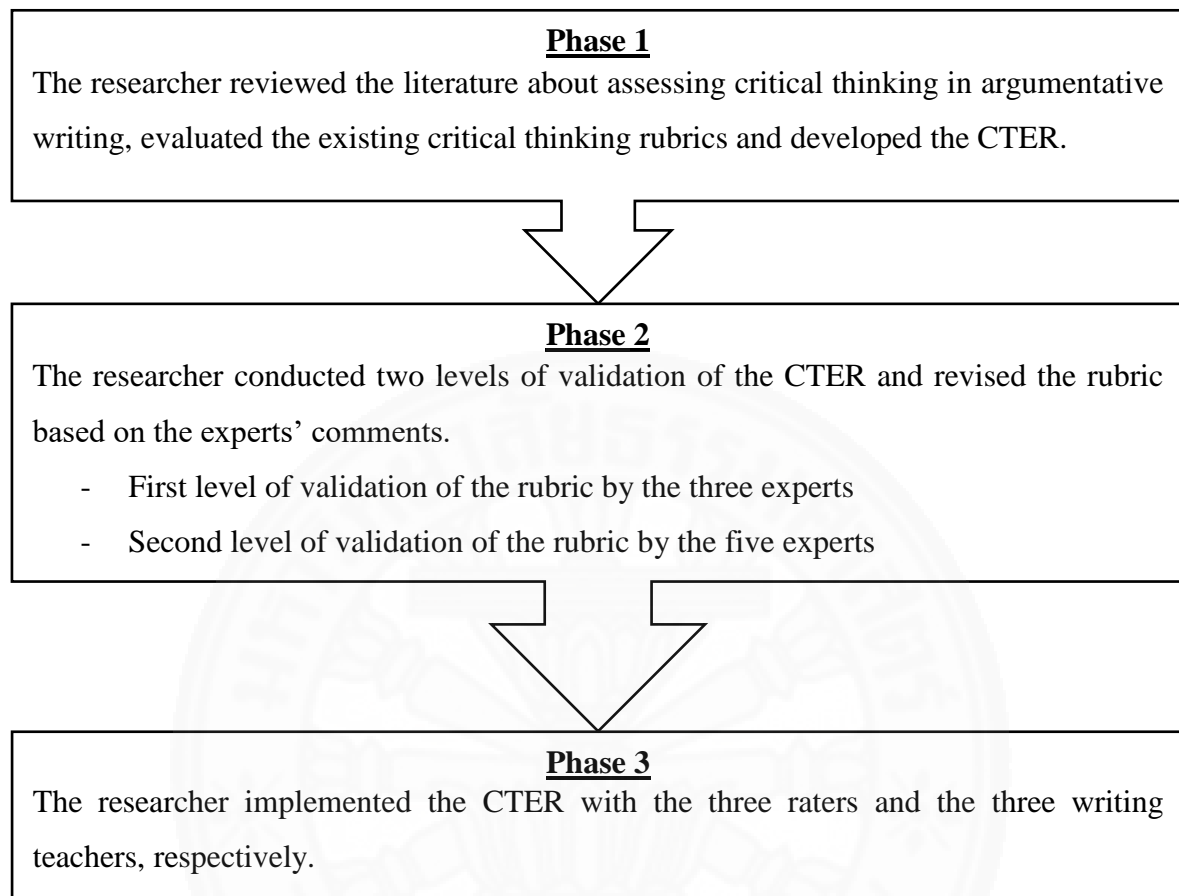
Assessment is a Journey. Developing a meaningful and useful [critical thinking] assessment is a continuing process, not a closed-end project. It does not end at a particular time or after a specific period. It is an ongoing effort that requires constant reevaluation, revision and improvement of procedures.

(Peach, Mukherjee & Hornyak, 2007, p.315)

This study aimed at 1) developing the CTER for assessing critical thinking in English argumentative essays of EFL university students, 2) examining the validity and the inter-rater reliability of the CTER and 3) investigating the attitudes of the raters and the writing teachers after they implement the rubric. Two research questions included: 1) Is the CTER a valid and reliable assessment instrument? and 2) What are the attitudes of the raters and the writing teachers after they use the CTER? The purposes of this chapter are to present and discuss the research findings. Based on the findings and the discussion, the implications for teaching and assessment are provided. Finally, this chapter ends with the limitations of this study and the recommendations for further research.

This study employs a mixed method research design and the ongoing process of developing, validating and revising the CTER. Therefore, there are three research phases. The overview of the research phases is illustrated again in Figure 4.1. In brief, to answer the first research question, the analysis of both qualitative and quantitative data obtained from the experts in the two levels of validation of the CTER was presented, and how the revisions increased the validity of the CTER was illustrated. Also, the analysis of the qualitative data from the raters' interview and the writing teachers' questionnaire and interview regarding the validity of the CTER was demonstrated. Besides, the statistical evidence of the inter-rater reliability of the rubric was reported. To answer the second research question, the analysis of

the qualitative data from the raters' interview and the writing teachers' the questionnaire and the interview was articulated.



**Figure 4.1** Three research phases of the study

#### **4.1 Research Question 1: Is the CTER a valid and reliable assessment instrument?**

The first research question investigated whether the CTER was a valid and reliable rubric. To address this question, the researcher conducted the qualitative analysis of the experts' free interview and field notes obtained from the first level of validation. Also, the researcher performed the analysis of the data from the experts' free interview and field notes, including the experts' validation survey obtained from the second level of validation. The revisions of the evaluative descriptors were also used as the evidence of the validity of the CTER. Besides, the researcher used the analysis of the raters' interview and the writing

teachers' the questionnaire and the interview asking about their opinions regarding the validity of the CTER. To confirm the validity of the CTER, both quantitative and qualitative data from the experts were used to triangulate with the qualitative data from the raters and the writing teachers. Finally, the statistical evidence of the inter-rater reliability of the CTER was uncovered.

#### **4.1.1 Findings from first level of validation of the CTER**

The results of the qualitative data analysis indicated that three experts in the fields of critical thinking, L2 writing and language assessment firmly agree that the CTER includes all crucial, relevant aspects of critical thinking construct in English argumentative essays of EFL university students; that is, 1) the ability to write an introduction to gain readers' attention, 2) the ability to give reasons and evidence to support a claim, 3) the ability to state a counter-argument and a refutation, 4) the ability to write a conclusion, 5) the ability to organize ideas in a logical way and 6) the ability to use word choices and sentence structures effectively. The experts also agree that the evaluative descriptors in each rating scale of the six domains of the assessment are clear, relevant and appropriate for each domain.

Moreover, to enhance the validity of the CTER, the evaluative descriptors are revised based on the experts' comments. For example, subjective adjectives in the second domain of the assessment, such as *very compelling reasons* (in score level 4) and *convincing reasons* (in score level 3), are removed. Besides, clear and objective descriptors, such as *strong, relevant reasons* (in score level 4), are used, and the number of the reasons was quantified. Thus, the descriptors can make a clear distinction among the four score levels. These revisions make the descriptors of the CTER more valid.

#### **4.1.2 Findings from the second level of validation of the CTER**

##### **Validity of the six domains of the critical thinking assessment**

**Table 4.1:** The validation results of the six domains of the critical thinking assessment

<b>The six domains of critical thinking assessment of the CTER</b>	<b>n</b>	<b><math>\bar{x}</math></b>	<b>S.D.</b>
Domain 1: Ability to write an introduction to gain readers' attention	5	5.00	0.00
Domain 2: Ability to give reasons and evidence to support a claim	5	5.00	0.00
Domain 3: Ability to state a counter-argument and a refutation	5	4.80	0.45
Domain 4: Ability to write a conclusion	5	5.00	0.00
Domain 5: Ability to organize ideas in a logical way	5	5.00	0.00
Domain 6: Ability to use word choices and sentence structures effectively	5	5.00	0.00
<b>Overall <math>\bar{x}</math> and S.D.</b>	<b>5</b>	<b>4.96</b>	<b>0.09</b>

As indicated in the table above, five experts in the field of critical thinking, second language writing and language assessment strongly agree that six domains of the assessment are the valid and clear indicators of critical thinking construct in English argumentative essays of EFL university students with a mean ranging from 4.80 (S.D. = 0.45) to 5.00 (S.D. = 0.00) and with an overall mean of 4.96 (S.D. = 0.09). The data affirm the validity of the CTER.

#### **Validity of the descriptions of the six domains of the critical thinking assessment**

**Table 4.2:** The validation results of the descriptions of the six domains of the assessment

<b>The descriptions of the six domains of the assessment</b>	<b>n</b>	<b><math>\bar{x}</math></b>	<b>S.D.</b>
<b>The descriptions of domain 1:</b> Writes opening sentence(s) by using a rhetorical strategy for gaining readers' attention clearly, provides clear, detailed background information and states a claim clearly.	5	4.60	0.55

<b>The description of the six domains of the assessment</b>	<b>n</b>	<b><math>\bar{x}</math></b>	<b>S.D.</b>
<b>The description of domain 2:</b> Gives strong, accurate and relevant reasons with concrete evidence to support the claim. Elaborates the reasons and evidence clearly and thoroughly.	5	4.60	0.55
<b>The description of domain 3:</b> States a counter-argument and refutes it by using strong reasons and concrete evidence. Elaborates the reasons and evidence clearly and thoroughly.	5	4.60	0.55
<b>The description of domain 4:</b> Provides a summary or a restatement of the claim and a final, relevant constructive comment.	5	4.60	0.55
<b>The description of domain 5:</b> Organizes ideas coherently. That is, the essay has clear main parts: an introduction, a body paragraph and a conclusion. Ideas between paragraphs and ideas within each paragraph connect with each other logically. Uses logical connectors to link ideas clearly.	5	4.80	0.45
<b>The description of domain 6:</b> Uses word choices and sentence structures that are varied, appropriate and convey meaning clearly and appropriately. May use a metaphor or a simile to convey ideas clearly.	5	4.60	0.55
<b>Overall <math>\bar{x}</math> and S.D.</b>	<b>5</b>	<b>4.63</b>	<b>0.40</b>

According to Table 4.2, the experts strongly agree that the descriptions of the six domains of the assessment are valid and clear with a mean ranging from 4.60 (S.D. = 0.55) to 4.80 (S.D. = 0.55) and with an overall mean of 4.63 (S.D. = 0.40). The statistical data strongly affirm the validity of the CTER.

#### **Validity of the specific evaluative descriptors in each score level of the CTER**

**Table 4.3:** The validation results of the evaluative descriptors in each score level

Domains	Descriptors in each score level	n	$\bar{x}$	S.D.
<b>Domain 1</b>	Descriptors in the level of Excellent	5	4.60	0.55
	Descriptors in the level of Good	5	4.20	0.45
	Descriptors in the level of Fair	5	3.60	1.14
	Descriptors in the level of Poor	5	4.00	1.73
	<b>Overall <math>\bar{x}</math> and S.D.</b>	<b>5</b>	<b>4.10</b>	<b>0.84</b>
<b>Domain 2</b>	Descriptors in the level of Excellent	5	4.00	1.73
	Descriptors in the level of Good	5	4.20	0.84
	Descriptors in the level of Fair	5	3.80	1.64
	Descriptors in the level of Poor	5	4.00	1.73
	<b>Overall <math>\bar{x}</math> and S.D.</b>	<b>5</b>	<b>4.00</b>	<b>1.45</b>
<b>Domain 3</b>	Descriptors in the level of Excellent	5	3.80	1.64
	Descriptors in the level of Good	5	3.60	0.55
	Descriptors in the level of Fair	3	4.67	0.58
	Descriptors in the level of Poor	4	3.80	0.50
	<b>Overall <math>\bar{x}</math> and S.D.</b>	<b>3-5</b>	<b>3.93</b>	<b>1.12</b>
<b>Domain 4</b>	Descriptors in the level of Excellent	4	4.50	0.58
	Descriptors in the level of Good	4	4.75	0.50
	Descriptors in the level of Fair	4	4.75	0.50
	Descriptors in the level of Poor	5	4.00	1.73
	<b>Overall <math>\bar{x}</math> and S.D.</b>	<b>4-5</b>	<b>3.95</b>	<b>1.70</b>
<b>Domain 5</b>	Descriptors in the level of Excellent	5	4.20	1.30
	Descriptors in the level of Good	3	4.70	0.58
	Descriptors in the level of Fair	3	4.70	0.58
	Descriptors in the level of Poor	5	4.20	1.30
	<b>Overall <math>\bar{x}</math> and S.D.</b>	<b>3-5</b>	<b>4.20</b>	<b>1.30</b>
<b>Domain 6</b>	Descriptors in the level of Excellent	4	4.50	0.58
	Descriptors in the level of Good	4	4.00	1.41
	Descriptors in the level of Fair	4	3.80	0.96
	Descriptors in the level of Poor	4	3.80	0.96
	<b>Overall <math>\bar{x}</math> and S.D.</b>	<b>4</b>	<b>4.00</b>	<b>0.82</b>

As shown in Table 4.3, the experts agree that the specific descriptors in each score level of the six domains possess the validity. That is, for domain 1, the descriptors are valid with an overall mean of 4.10 (S.D. = 0.84). For domain 2, the descriptors are valid with an overall mean of 4.00 (S.D. = 1.45). For domain 3, the descriptors are valid with an overall mean of 3.93 (S.D. = 1.12). For domain 4, the descriptors are valid with an overall mean of 3.95 (S.D. = 1.70). For domain 5, the descriptors are valid with an overall mean of 4.20 (S.D. = 1.30). For domain 6, the descriptors are valid with an overall mean of 4.00 (S.D. = 0.82). The statistical data indicate the validity of the evaluative descriptors of the CTER.

As revealed in the table below, the experts believe that the evaluative descriptors of the six domains of the assessment can make a clear distinction among the four score levels with a mean ranging from 3.60 (S.D. = 1.14) to 4.40 (S.D. = 0.89) and with an overall mean of 4.00 (S.D. = 0.72). The statistical evidence ensures the validity of the CTER.

**Table 4.4:** The validation of the descriptors' ability to make a clear distinction among the four score levels

<b>The six statements about the descriptors</b>	<b>n</b>	<b><math>\bar{x}</math></b>	<b>S.D.</b>
1.The evaluative descriptions of domain 1 can make a clear distinction among the four score levels.	5	4.00	0.71
2.The evaluative descriptions of domain 2 can make a clear distinction among the four score levels.	5	4.20	0.84
3.The evaluative descriptions of domain 3 can make a clear distinction among the four score levels.	5	3.80	0.84
4.The evaluative descriptions of domain 4 can make a clear distinction among the four score levels.	5	4.40	0.89
5.The evaluative descriptions of domain 5 can make a clear distinction among the four score levels.	5	4.00	1.00
6.The evaluative descriptions of domain 6 can make a clear distinction among the four score levels.	5	3.60	1.14
<b>Overall <math>\bar{x}</math> and S.D.</b>	<b>5</b>	<b>4.00</b>	<b>0.72</b>

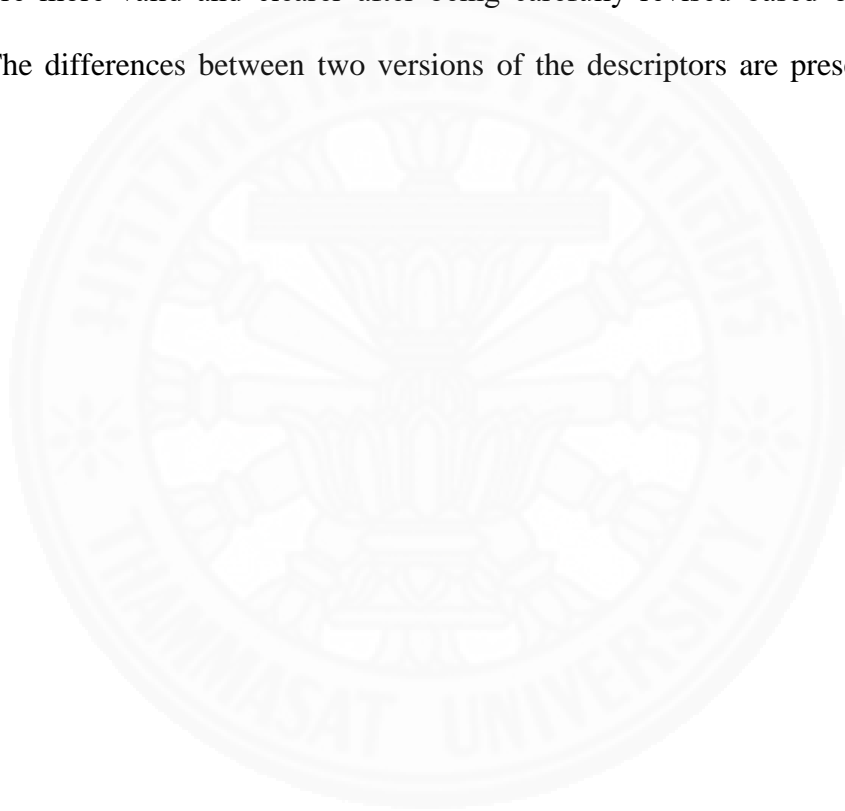
In sum, based on the experts' opinions, the six domains of the assessment are the valid and clear indicators of critical thinking in English argumentative essays of EFL university students. The evaluative descriptors of the six domains are also valid and can make a clear distinction among the four score levels. The quantitative data firmly ensure that the CTER is a valid assessment instrument.

The revisions of the descriptors based on the experts' comments also significantly enhance the validity of the CTER. There are four aspects of the revisions. First, ambiguous or subjective descriptors are removed. Second, existing terminologies are elaborated clearly in order to enable the writing teachers to have the same and clear concept of scoring in mind when they grade the essays. Third, more detailed evaluative descriptors are added to make the descriptors more valid and clearer. Finally, the number of the crucial elements of English argumentative writing was quantified; therefore, the descriptors can make a clear distinction among the four score levels. The next section will manifest the validity of the evaluative descriptors, and the differences between the two versions of the descriptors are illustrated.

### **1. The descriptors of the first domain of the critical thinking assessment (the ability to write introduction to gain readers' attention)**

The descriptors of domain 1 are more valid and clearer because of the following reasons. First, subjective descriptors in the score level of Excellent are revised to make the descriptors more objective, more concrete and clearer. The revised descriptors are as follows: *Writes an opening sentence by using a rhetorical strategy for gaining readers' attention clearly, writes clear, detailed background information and writes a claim clearly. The introduction is clear and well-elaborated.* Second, the descriptors, namely *the rhetorical strategies for gaining readers' attention*, are clearly elaborated, and the crucial elements of the rhetorical strategies are provided in the rating manual clearly. That is, the rhetorical strategies include: 1) using a relevant, interesting quotation, 2) using a metaphor or a simile,

3) using a rhetorical question, 4) using research results, 5) confessing something related to a writer's life and a topic, 6) stating a striking problem, 7) stating an opposing view or a debate pertaining to a topic, and 8) stating a crucial fact that is related to the topic. Third, the descriptors in the score level of Fair are clearly explained as follows: *May write an opening sentence and background information without a clear claim.* Finally, the descriptors in the score level of Poor are revised to make the descriptors concise and concrete as follows: *Write an unclear and irrelevant introduction.* Obviously, the descriptors of the first domain of the assessment are more valid and clearer after being carefully revised based on the experts' comments. The differences between two versions of the descriptors are presented in Table 4.5.



**Table 4.5:** The second and the final versions of the descriptors of domain 1

Four score levels	Domain 1: the ability to write an introduction to gain readers' attention	
	The second version of the descriptors of domain 1	The final version of the descriptors of domain 1
Excellent (4)	-The introduction contains 3 elements: 1) interesting opening sentence(s), 2) relevant background information and 3) a clear claim. Background information links opening sentence(s) to the claim clearly. The introduction is outstanding and successfully captures readers' interest.	-Writes an opening sentence by using a rhetorical strategy for gaining readers' attention clearly, writes clear, detailed background information and writes a claim clearly. The introduction is clear and well-elaborated.
Good (3)	-The introduction contains 3 elements: 1) good opening sentence(s), 2) relevant background information and 3) a claim.	-Writes a relevant opening sentence, detailed background information and a clear claim. -May only state the claim and briefly provide reasons.
Fair (2)	-The introduction contains 2 elements: 1) opening sentence(s) and 2) relevant background information. No claim.	-May write an opening sentence and background information without a clear claim. -May write an opening sentence and a claim without detailed background information. -OR writes a claim only.
Poor (1)	-No introduction.	-Writes an unclear and irrelevant introduction.

## **2. The descriptors of the second domain of the critical thinking assessment (the ability to give reasons and evidence to support a claim)**

The descriptors of domain 2 are more valid, more comprehensive and clearer because of three reasons. First, the crucial descriptors are clearly explained, and their detailed definitions are stated in the rating manual clearly. For example, a reason is defined as logical statements or explanations that support a claim, while evidence includes a logical elaboration, first or second-hand experience, research findings, personal observations, opinions of experts in the field, interviews, comparisons and contrasts, anecdotes and personal narratives.

Second, more evaluative descriptors are added, so this domain of the assessment can capture all possibilities of the responses validly and clearly. Examples of such descriptors in the level of Good are as follows:

- *Gives at least 2 strong, relevant reasons with relevant, concrete evidence. Reasons and evidence are convincing and non-fallacious, although few minor details may need more elaborations.*
- *May give many reasons, and good reasons far outnumber poor reasons.*

Third, clear detailed descriptors are employed, and the number of the reasons is quantified. Therefore, the descriptors can make a clear distinction among the four score levels. Examples of such descriptors include the followings:

- *Gives at least 3 strong, relevant reasons with relevant, accurate, concrete evidence... (Level Excellent),*
- *Gives at least 2 strong, relevant reasons with relevant, concrete evidence... (Level Good),*
- *May give 1 strong, relevant reason with relevant evidence... (Level Fair), and*
- *Reasons and evidence may be inaccurate, irrelevant, fallacious, too weak or too limited... (Level Poor)*

Obviously, these revisions increase the validity of the descriptors. The differences between two versions of the descriptors of domain 2 are presented in Table 4.6.

**Table 4.6:** the second and the final versions of the descriptors of domain 2

<b>Domain 2: The ability to give reasons and evidence to support a claim</b>		
<b>Four score levels</b>	<b>The second version of the descriptors of domain 2</b>	<b>The final version of the descriptors of domain 2</b>
Excellent (4)	-Gives at least 2 strong, relevant reasons with concrete evidence to support claim. Reasons and evidence are clearly and convincingly elaborated.	-Gives at least 3 strong, relevant reasons with relevant, accurate, concrete evidence. May use many kinds of solid and credible evidence, so the arguments are very strong. All reasons and evidence are non-fallacious and clearly, thoroughly and convincingly explained.
Good (3)	-Gives at least 2 relevant reasons with concrete evidence to support claim. However, the evidence is not elaborated.	-Gives at least 2 strong, relevant reasons with relevant, concrete evidence. Reasons and evidence are convincing and non-fallacious, although few minor details may need more elaborations. -May give many reasons, and good reasons far outnumber poor reasons.
Fair (2)	-Gives at least 2 relevant reasons without evidence.	-May give 1 strong, relevant reason with relevant evidence. Yet, the evidence may not be concrete or clearly explained. -May give 2-3 relevant reasons without evidence. -May give many relevant reasons, but they are still unclear or need more evidence. -May give many reasons which have different levels of quality. -OR the number of good reasons equals the number of poor reasons.
Poor (1)	-No relevant reasons and no evidence.	-Reasons and evidence may be inaccurate, irrelevant, fallacious, too weak or too limited. -May give many reasons, but poor reasons far outnumber good reasons. -OR reasons and evidence are absent.

### **3. The descriptors of the third domain of the critical thinking assessment (the ability to state a counter-argument and a refutation)**

The descriptors of domain 3 are more valid and can make a clear distinction among the four score levels because clear detailed qualitative descriptors are employed and because the number of the counter-arguments and refutations is clearly quantified. Examples of such descriptors in the four score levels of this domain include the followings:

- *States at least 2 counter-arguments and refutes them by using convincing reason(s) with relevant, accurate, concrete evidence (Level Excellent),*
- *States at least 1 counter-argument and refutes it by using relevant reason(s) with relevant evidence (Level Good),*
- *States 1 counter-argument and refutes it by using relevant reason(s) with relevant evidence. Yet, the evidence may not be concrete or clearly explained.... (Level Fair), and*
- *May state counter-argument(s) and refutation(s) inaccurately, unclearly, incompletely, irrelevantly and/or fallaciously (Level Poor).*

Obviously, after being carefully revised based on the experts' comments, the descriptors of domain 3 are more comprehensive, more objective and much clearer. Thus, the revisions significantly increase the validity of the descriptors of the CTER. The differences between the two versions of the descriptors of this domain are presented in Table 4.7.

**Table 4.7:** The second and the final versions of the descriptors of domain 3

<b>Domain 3: the ability to state a counter-argument and a refutation</b>		
<b>Four score levels</b>	<b>The second version of the descriptors of domain 3</b>	<b>The final version of the descriptors of domain 3</b>
Excellent (4)	-States a strong counter-argument and refutes it with much stronger reasons and concrete evidence. Reasons and evidence are clearly and convincingly elaborated.	-States at least 2 counter-arguments and refutes them by using convincing reason(s) with relevant, accurate, concrete evidence AND explains all reasons and evidence clearly and thoroughly.
Good (3)	-States a strong counter-argument, and attacks it with much stronger reasons and evidence. However, the evidence is not elaborated.	-States at least 1 counter-argument and refutes it by using more convincing reason(s) with relevant, concrete evidence AND explains all reasons and evidence clearly.
Fair (2)	-States a general counter-argument, and refutes it by using general reasons without evidence.	-States 1 counter-argument and refutes it by using relevant reason(s) with relevant evidence. Yet, the evidence may not be concrete or clearly explained. -May state 2 counter-arguments, but can refute only one.
Poor (1)	-No counter-argument and no refutation  -OR states a strong counter-argument but cannot attack it. This makes a student's own argument less effective and weak.	-May state counter-argument(s) and refutation(s) inaccurately, unclearly, incompletely, irrelevantly and/or fallaciously. -May state strong counter-arguments, but cannot refute them. -OR a counter-argument and a refutation are absent.

#### **4. The descriptors of the fourth domain of the critical thinking assessment (the ability to write a conclusion)**

The descriptors of domain 4 are more valid and clearer because of four reasons. First, the descriptors in the level of Excellent are revised to make the descriptors more concrete and clearer as follows: *Gives a summary or a restatement of a claim AND writes a final, relevant,*

*constructive comment*. Second, the descriptors, namely *constructive comment*, are defined clearly as a suggestion, a recommendation for someone to take an action or a solution to a problem. Third, the descriptors in the score level of Fair are revised to make the descriptors more valid and clearer as follows: *Gives a summary or a restatement of a claim without a final, relevant comment (or vice versa)*. Fourth, the descriptors in the score level of Poor, namely *no conclusion*, are revised to make the descriptors more concrete and more comprehensive as the followings: *May write a conclusion unclearly, irrelevantly or illogically. OR a conclusion is absent*. These revisions make the evaluative descriptors clearer and thus enhance the validity of the CTER. The differences between the two versions of the descriptors of the fourth domain are presented in Table 4.8.

**Table 4.8:** the second and the final versions of the descriptors of domain 4

<b>Domain 4: the ability to write a conclusion</b>		
<b>Four score levels</b>	<b>The second version of the descriptors of domain 4</b>	<b>The final version of the descriptors of domain 4</b>
Excellent (4)	The essay has two elements: 1) a summary AND 2) a final, constructive comment.	-Gives a summary or a restatement of a claim AND writes a final, relevant, constructive comment. <b>Note:</b> Constructive comment intends to help or improve something. It can be a useful suggestion, a recommendation (for someone to take action) or a solution.
Good (3)	The essay has two elements: 1) a summary AND 2) a final, general comment. OR, may have only a final, constructive comment.	-Gives a summary or a restatement of a claim AND writes a final, relevant comment -OR only writes a final, relevant, constructive comment.
Fair (2)	-The essay has a summary WITHOUT a final comment.	-Gives a summary or a restatement of a claim without a final, relevant comment (or vice versa).
Poor (1)	No conclusion.	-May write a conclusion unclearly, irrelevantly or illogically. -OR a conclusion is absent.

## **5. The descriptors of the fifth domain of the critical thinking assessment (the ability to organize ideas in a logical way)**

The descriptors of domain 5 are more valid and clearer due to the following reasons. First, ambiguous descriptors, such as *ideas can be followed very easily* (in the level of Excellent) and *ideas can be followed easily* (in the level of Good), are removed. Second, the original descriptors in the level of Good are revised from *the essay is mostly coherent and organized...* to *...the essay has a clear organization. Although unclear connectors between ideas occasionally appear, they do not cause misunderstanding*. Obviously, such revisions make the descriptors more concrete and clearer. Finally, the descriptors in the score level of Fair are changed, and this made the descriptors more comprehensive and clearer as follows: *some parts of the essay are organized, but some parts are poorly organized or have connectors that obscure the connection of ideas and cause misunderstanding*. Undeniably, these revisions make the descriptors of the CTER more valid and clearer. The differences between the two versions of the descriptors of the fifth domain are presented in Table 4.9.

**Table 4.9:** The second and the final versions of the descriptors of domain 5

<b>Domain 5: the ability to organize ideas in a logical way</b>		
<b>Four score levels</b>	<b>The second version of the descriptors of domain 5</b>	<b>The final version of the descriptors of domain 5</b>
Excellent (4)	The essay is well-organized. That is, the essay has clear main parts: an introduction, a body paragraph and a conclusion. Ideas within each paragraph AND between paragraphs connect with each other logically. Ideas flow smoothly and can be followed very easily.	-Ideas are carefully and very well organized. Ideas within each paragraph and between paragraphs flow smoothly and are very easy to follow. All connectors show a clear and logical connection between ideas.
Good (3)	The essay is mostly coherent and organized. That is, the essay has an introduction, a body paragraph and a conclusion, and ideas in each part mostly connect with each other. Ideas can be followed easily.	-Overall, the essay has a clear organization. Although unclear connectors between ideas occasionally appear, they do not cause misunderstanding.
Fair (2)	Some parts of the essay are more coherent and organized than the other parts. The essay can be followed with some difficulties.	-Some parts of the essay are organized, but some parts are poorly organized or have connectors that obscure the connection of ideas and cause misunderstanding.
Poor (1)	-Ideas lack organization. The connectors do not show a logical connection between ideas and cause misunderstanding.	-Ideas lack organization. The connectors do not show a logical connection between ideas and cause misunderstanding.

#### **6. The descriptors of the sixth domain of the critical thinking assessment (the ability to use words and sentence structures effectively)**

The descriptors of domain 6 are more valid and clearer due to the following reasons. First, subjective and unclear descriptors, such as *word choices are outstanding and enrich an argument* (in the level of Excellent) and *word choices help enrich an argument* (in the level of Good), are removed. Second, clear, detailed, qualitative descriptors are employed, and the

frequency of the use of word choices and sentence structures is identified. Therefore, the descriptors can make a clear distinction among the four score levels. Examples of such descriptors are as follows:

- ***all** word choices and **all** sentence structures convey meanings clearly...* (Level Excellent),
- ***most** word choices and **most** sentence structures are **appropriate** and **clear**...* (Level Good),
- *word choices and sentence structures convey meanings to some extent. **Some inappropriate** and **unclear** word choices and/or sentences structures do appear and cause misunderstanding.* (Level Fair), and
- ***all** or **almost all** of word choices and sentence structures are unclear, inappropriate or distort meaning dramatically, causing misunderstanding* (Level Poor).

Due to the revisions, the evaluative descriptors of the sixth domain of the assessment are more comprehensive and clearer, and this increases the validity of the CTER. The differences between the two versions of the descriptors are presented in Table 4.10.

**Table 4.10:** the second and the final versions of the descriptors of domain 6

<b>Domain 6: The ability to use word choices and sentence structures</b>		
<b>Four core levels</b>	<b>The second version of the descriptors of domain 6</b>	<b>The final version of the descriptors of domain 6</b>
Excellent (4)	-Uses word choices and sentence structures that are varied, appropriate and clear throughout the essay. May use metaphor or simile to convey ideas clearly and appropriately. Word choices are outstanding and enrich an argument.	-Uses various, appropriate word choices and sentence structures throughout the essay. All word choices and all sentence structures convey meanings clearly. <u>May</u> use metaphor or simile to convey ideas clearly and appropriately.
Good (3)	-Word choices and sentence structures are mostly appropriate, and clear. Word choices enrich an argument.	-Most word choices and most sentence structures are appropriate and clear. Although unclear or inappropriate word choices occasionally appear, they do not cause misunderstanding.
Fair (2)	Word choices and sentence structures are quite clear and demonstrate some variety	-Word choices and sentence structures convey meanings to some extent. Some inappropriate and unclear words choices and/or sentence structures do appear and cause misunderstanding.
Poor (1)	Word choices and sentence structures are not varied and not clear.	-All or almost all of word choices and sentence structures are unclear, inappropriate or distort meaning dramatically, causing misunderstanding.

All in all, the revisions of the descriptors under the experts' supervision make the CTER more valid. That is, the descriptors in each score level are more objective, concrete and comprehensive and can evaluate all possibilities of the responses clearly. Besides, all terminologies are clearly explained, and their clear, detailed definitions are stated in the rubric or in the rating manual clearly (See Appendices 5 and 9). Also, the use of both quantitative and qualitative descriptors makes a clear distinction among the four score levels.

In addition to the revisions of the descriptors, the analysis of the qualitative data from the raters' interviews and the writing teachers' questionnaire and interview also demonstrates that the CTER is valid because it covers all crucial aspects of critical thinking in argumentative writing and contains clear detailed evaluative descriptors. The results of this study resonate with those of Alsina and her colleagues (2017) which found that the clarity and appropriateness of the domains of assessment and the evaluative descriptors indicate the validity of the assessment instrument. The excerpts from the participants in the present study are indicative:

*The six domains of the assessment are clear. The rubric can assess critical thinking in argumentative writing. Without the rubric, there are no criteria that help teachers grade the essays validly. (Rater A, Interview)*

*I think the scoring criteria are clear and can help me assign a score to a text accurately, so for me, the rubric works. (Rater B, Interview)*

*I like the rubric because there are clear detailed descriptors in each score level of the six domains of the assessment. This is the outstanding feature of the rubric. I think the rubric can assess students' critical thinking abilities. (Rater C, Interview)*

*Your rubric includes all crucial elements of critical thinking in argumentative writing. The rubric is good because it can help teachers assess critical thinking, whereas other rubrics cannot. (Writing Teacher 1, Interview)*

*The rubric is a good tool for assessing critical thinking because students have to state a claim and support it with reasons and evidence. Also, a counter-argument and a refutation are needed to clarify their ideas. Therefore, they need to practice thinking critically and organizing ideas logically. (Writing teacher 2, Questionnaire)*

*I like the rubric because it can be used for assessing critical thinking in argumentative essays. I agree upon the six domains of the assessment. These domains are clearly defined already. (Writing Teacher 2, Interview)*

*I think your rubric is ok. The defined criteria are ok. The criteria are important because they limit the scope of grading for raters. I think your criteria are clear and appropriate already.* (Writing Teacher 3, Interview)

Contrary to the critical thinking rubrics of the previous studies (e.g., Brookhart, 2010; Hodges et al., 2019; Liu & Stapleton, 2018; McLendon, 2008; Metcalfe, 2007; Nejmaoui, 2019; Pei et al., 2017; Shehab & Nussbaum, 2015; Stapleton & Wu, 2015; Wright, 2016), the CTER embraces all relevant, clear and crucial aspects of critical thinking in English argumentative essays of EFL university students ( $\bar{x} = 4.96$ , S.D. = 0.09). Also, the current study found that the analytic critical thinking rubric is valid since it provides users with numerical scoring results and detailed evaluative information. Thus, the findings of this study contradict those from the previous studies (e.g., Angeli & Valanides, 2009; Bauer, 2016; Liu, Wu & Shieh, 2015; Stapleton, 2001a; Welch, Hieb, & Graham, 2015) which proposed checklist or holistic critical thinking rubrics—assessment tools that do not provide the students with clear, detailed, diagnostic feedbacks. Also, unlike the analytic rubrics of the previous researchers (e.g., Daud, 2012; Dong, 2015; McLaughlin & Moore, 2012; Saxton et al., 2012), the CTER has the evaluative descriptors that can make a clear distinction among the score levels.

Three important factors account for the validity of the CTER. First, this rubric has clear operational definitions of critical thinking construct—i.e., the six critical thinking abilities in English argumentative writing. These domains include the ability to 1) write an introduction to gain readers' attention, 2) give reasons and evidence to support a claim, 3) state a counter-argument and a refutation, 4) write a conclusion, 5) organize ideas logically and 6) use word choices and sentence structures effectively. These domains are developed from extensively reviewing and examining the related literature about critical thinking

theories, argumentative writing, language assessment and existing rubrics for assessing critical thinking in argumentative writing.

Second, the six domains of the CTER reflect critical thinking skills of the existing critical thinking taxonomies. Such critical thinking skills of the taxonomies include planning how to write an argument, analyzing, synthesizing and evaluating information from various perspectives, solving problems, developing and defending a position, reasoning constructively, categorizing ideas, using language that is clear and appropriate for an audience or an academic context and so on (Anderson & Krathwohl, 2001; Bloom et al., 1956; Ennis, 1993; Facione, 1990, 2018; Larsson, 2017; Levenrenz, 2014; Munzenmaier & Rubin, 2013; Paul & Elder, 2008). The CTER includes six important abilities. Expressing these six abilities in writing effectively requires the use of critical thinking skills, such as analyzing, synthesizing, evaluating and reasoning. Thus, the six abilities of the CTER manifest the crucial skills of the critical thinking taxonomies. This factor contributes to the validity of the CTER (Hodges et al., 2019; Messick, 1996b; Puthikanon, 2009).

Third, the rubric has clear rating scales and descriptors that are specific, concrete, sufficient, relevant and appropriate for each domain of the assessment (Alsina et al., 2017; Cargas et al., 2017; Farhardy, 2012; Messick, 1995; Weigle, 2002; Zhao, 2010). The CTER has been validated thoroughly by the experts and has been revised carefully and continuously by the researcher under the experts' supervision (Carter et al., 2017; Lallmamode, Daud, & Kassim, 2016; Peach et al., 2007; Polston, 2014; Williams, 1999). This makes the descriptors of the rubric illustrate a clear distinction among the score levels (Hodges et al., 2019; Janssen et al., 2015). When teachers use the rubric to evaluate students' essays, they can clearly identify different levels of critical thinking in writing of each student. This also makes this rubric be able to capture and assess "all response possibilities" appearing in EFL students' argumentative essays (Green & Hawkey, 2012, p. 305; Knoch, 2007, 2009).

### 4.1.3 The inter-rater reliability of the CTER

The first research question also examined the inter-rater reliability of the CTER. To answer this research question, the scoring results of 20 writing samples from the four scorers (i.e., three writing teachers and the researcher) were analyzed by using two statistical approaches: Kendall's coefficient of concordance (Kendall's W) and Intra-class correlation coefficient (ICC). Kendall's W was used to investigate the inter-rater reliability of each domain of the assessment and investigate the total scoring results of the six domains. ICC was used to examine the inter-rater reliability of the total scoring results of the six domains. The data from Kendall's W statistical method were used to triangulate with those from ICC statistical approach in order to ensure the consistency of the findings.

**Table 4.11:** the results of the inter-rater reliability of the CTER via Kendall's W

<b>Critical thinking domains</b>	<b>Kendall's W values</b>
Domain 1	0.5*
Domain 2	0.3
Domain 3	0.4*
Domain 4	0.5*
Domain 5	0.4*
Domain 6	0.5*
<b>Total</b>	<b>0.5*</b>

\*P < 0.05

According to Table 4.11, the inter-rater reliability of the five domains of the assessment (i.e., domains 1, 3, 4, 5 and 6) is moderate (Kendall's W = 0.4 - 0.5) and is statistically significant at the 0.05 level, whereas the inter-rater reliability of domain 2 is weak (Kendall's W = 0.3) and is not statistically significant at the 0.05 level. Table 4.11 also displays that the inter-rater reliability of the total scoring results of the six domains of the assessment is moderate (Kendall's W = 0.4 - 0.5) and is statistically significant at the 0.05 level. Thus, based on the data, the CTER achieves the overall inter-rater reliability.

**Table 4.12:** The results of the inter-rater reliability of the CTER via ICC

Intraclass Correlation Coefficient	F test
0.70	3.223*

\*p&lt;0.01

Similarly, Table 4.12 reveals that the inter-rater reliability of the total scoring results from the six domains of the assessment is moderate with an ICC of .70 and is statistically significant at the 0.01 level. The data suggest that the CTER achieves the moderate degree of the inter-rater reliability.

Domains of the assessment 1, 3, 4, 5 and 6 of the CTER and the total scoring results from the six domains achieve the acceptable degree of the inter-rater reliability (Kendall's  $W = 0.4 - 0.5$ ; ICC = 0.70) because of three important factors.

First, the clarity of the CTER reduces the scorers' subjectivity and promotes consistency in scoring. To explain, the writing teachers may have had different perceptions about a good or bad writing and different criteria in mind when grading an essay (Lim, 2009; Plakans & Gebril, 2017; Quintieri, 2005; Simpkins, 2016; Wright, 2016; Weigle, 1994). For example, teacher A is very strict with grammatical accuracy and grades it rigorously, while other teachers are not strict with it and grade it leniently (Quintieri, 2005). Teacher B marks an essay holistically, whereas teacher C prefers the use of an emotional appeal in a persuasive writing and focuses on language use over other aspects of writing. Undeniably, human raters have subjectivity which causes them to grade the essay differently or grade the same aspect of writing too harshly or too leniently than other raters do (e.g., Lim, 2009; Quintieri, 2005). The subjectivity from having different perceptions about writing makes the scoring results unreliable. The CTER has clear operational definitions of critical thinking and has specific, concrete detailed descriptors that can make a clear distinction among the four score levels—

especially “in the middle levels” where the rating is “the most difficult” (Janssen et al., 2015, p.64). Thus, using the CTER helps teachers develop clear and same perception about the assessment, and this helps reduce their personal bias or subjectivity and makes them mark an essay consistently (Ben-Simon & Bennett, 2007; Chan, Inoue, & Taylor, 2015; Douglas, 2010; Gebril & Plakans, 2014; Green, 2014; Harvin, 2016; Kang, 2015; Knoch, 2007, 2009, 2011; Lim, 2009; Martens, 2018; Matthews, 2000; Mearman, 2013; New, 2005; Rezaei and Lovorn, 2010; Riesser, 1994; Rogers, 2005; Ross, 2006; Stellmack et al., 2009).

Second, the individual training on using the CTER helped the writing teachers grade the essays consistently. In the training, each writing teacher discussed the six domains of the assessment and the evaluative criteria in the rubric with the researcher, practiced using the CTER to grade some writing samples, compared, contrasted and discussed the scoring results with the investigator and received constructive comments about their grading. Thus, the training is a key factor making the writing teachers understand the rubric clearly and grade the writing samples consistently (Gebril & Plakan, 2014; Jossen & Svingby, 2007; Qin, 2009; Rowicki, 2008).

Third, using the anchor papers during training and during grading the essays promotes the rater consistency. The anchor papers exemplify clear and concrete responses at each level of performance identified in the CTER; therefore, reading, analyzing and discussing the anchor papers with the researcher enable the three writing teachers to have clear and same understanding about the scoring criteria and about how to mark the essays through the CTER (Jossen & Svingby, 2007; Kang, 2015). Also, the writing teachers can check their scoring results against the anchor papers to ensure that “their judgments are consistent” with those in the anchor papers (Green, 2014, p.155). In sum, the anchor papers serve as a benchmark that

represents each score level of the CTER, so using the anchor papers makes three writing teachers grade the essays through the CTER consistently.

The second domain of the CTER has the inter-rater reliability that is weak (Kendall's  $W = 0.3$ ) and is not statistically significant at the 0.05 level because the three writing teachers interpreted the descriptors of this domain differently. To illustrate, at that period of time, all teachers had urgent works and were unavailable to come and obtain the training together. To facilitate the teachers, the researcher provided them with an individual training on how to use the rubric, not a group training; that is, he trained three writing teachers in different sessions at a different period of time, based on their convenience. Thus, even though the writing teachers discussed the rubric with the researcher, they did not have the opportunity to discuss the scoring criteria with each other. The second domain of the CTER contains many detailed descriptors which require careful consideration and interpretation from the writing teachers. Examples of such descriptors include "*reasons and evidence may be .... fallacious, too weak or too limited.*" Therefore, without the group training, the writing teachers might interpret these descriptors differently and mark the essays into a different direction. As Green (2014) pointed out, although the rubric has "complete and clearly written" descriptors, without the group training, the scorers might interpret some descriptors differently (p.155). Hence, to make the second domain more reliable, all scorers need to be trained together. For example, they need to 1) discuss the scoring criteria together, 2) grade the essays individually, 3) compare, contrast and discuss the scoring results with each other and 4) justify their grading by referring to the descriptors in order to reach an agreement (e.g., McNamara, 2000; Weigle, 2002).

#### **4.1.4 Summary of the results about the validity and the inter-rater reliability of the CTER**

The CTER contains the six valid and clear indicators of critical thinking in argumentative essays of EFL university students ( $\bar{x} = 4.96$ , S.D. = 0.09). The evaluative descriptors are valid ( $\bar{x} = 4.63$ , S.D. = 0.40) and can make a clear distinction among the four score levels ( $\bar{x} = 4.00$ , S.D. = 0.72). Also, the revisions of the evaluative descriptors based on the experts' comments enhance the validity of the CTER. Besides, the raters and the teachers express the unanimous agreement that the rubric is valid because it contains all clear, relevant and important elements of critical thinking in argumentative writing. In addition, the total scoring results from the six domains of the rubric achieves the acceptable level of the inter-rater reliability (Kendall's  $W = 0.4 - 0.5$ ; ICC = .70). For these reasons, it can be concluded that the CTER is a valid and reliable critical thinking assessment.

#### **4.2 Research Question 2: What are the attitudes of the raters and the writing teachers after they use the CTER?**

To respond to this research question, the analyses of the attitudes of the three raters and the three writing teachers about the CTER are reported. The data are drawn from the raters' interviews and the teachers' questionnaire and interview. To provide more in-depth perspectives about the CTER, the data from the raters are used to triangulate with the data from the teachers in order to ensure the validity of the findings. The results from both sources of data will be presented and discussed together as the evidence of this research question.

The researcher of the present study found three emerging themes that show the participants' attitudes towards the CTER. First, the CTER can be used to promote learning and critical thinking of EFL students. Second, the CTER can be used to help teachers plan and improve teaching and provide an appropriate instruction to best serve the students' needs. Third, the CTER should be further improved, especially in an aspect of being time-consuming.

## **1. Teachers can use the CTER to promote learning and critical thinking of EFL students**

Based on the questionnaires and the interviews with the participants, the first emerging theme from the data analysis indicate that the CTER can be used to foster learning and critical thinking of EFL students by two ways as follows:

### **a) Teachers can use the CTER as a tool for teaching EFL students to express six critical thinking abilities in English argumentative essays effectively**

The participants corroborate that teachers can use the CTER as an instructional tool for helping EFL students understand six domains of critical thinking and include these crucial elements in their argumentative essays skillfully. To explain, six domains of the CTER are the rhetorical and cognitive strategies for English argumentation. Also, the rubric provides “descriptive and specific” feedback in each score level of the six domains (Katz, 2012, p.72). Therefore, this rubric serves as an “instructional illuminator” (Popham, 1997, p.75) and a “writing development model” (Bailey et al., 2015, p.525) that help the students learn how to write an interesting introduction, how to construct a strong argument, how to arrange ideas clearly and so on. Thus, if teachers neither teach critical thinking nor define its definitions in a rubric clearly, the students will not understand what critical thinking is and cannot write an argument effectively (McKinley, 2013; Mehta & Al-Mahrouqi, 2014; Xu, Huang, & You, 2016; Yeh, 2001). Yet, if teachers use the CTER to teach the crucial elements in argumentative writing, e.g., claims, reasons, counter-arguments and rebuttals, and demonstrate how these elements function and relate to each other, the students will be able to internalize “the rhetorical organization of an argument” and will have “an argument schema” which helps them think critically and write an argument efficiently (Reznitskaya et al., 2009, p.219).

*The rubric can be used for assessment-for-learning purposes. You can use the rubric as a guideline and detailed information about argumentative writing for teachers, scorers and students. (Rater A, interview)*

*Your rubric is good because it has clear, detailed information about critical thinking abilities in argumentative writing. I can teach my students by having them read and discuss the rubric, so they see the structure of argument and have clearer picture in minds about how to write an argument. ... Most students can write better. Some use a rhetorical strategy suggested in the rubric. One student uses Jane Austin's quote as a rhetorical strategy for gaining readers' attention in the introduction. It works. Her writing is much more interesting than it was before. I felt so surprised. (Rater B, Interview)*

*I think your rubric can foster students' critical thinking because it helps them learn the pattern and elements of argumentative essays. (Rater C, Interview)*

*You can use the rubric to develop students' critical thinking abilities. For example, you can teach students the definitions of evidence, such as using research findings, and using one's own observation or a personal experience, so they will know the definitions clearly and can provide evidence that is concrete and clear. (Writing teacher 1, Interview)*

*The rubric will enable the students to find reasons, think more critically and write an argument more logically. (Writing teacher 2, Questionnaire)*

*We can use the rubric as an instructional tool or an example of a good writing. ...Students will have a clear picture in minds about how to think critically and write an argument convincingly. (Writing teacher 3, Interview)*

The findings in this study correspond with those by Bailey and her colleagues (2015) which show that an assessment can be used for teaching critical writing to the students effectively. Also, as Ka-kan-dee and Kaur (2015) stated, in a Thai context, many teachers still lack knowledge and effective strategies for teaching English argumentative writing, so many

Thai students are not aware of the pattern and features of argumentative writing, write a claim unclearly, give no concrete evidence and lack a good organization of ideas. Therefore, the CTER can be a tool that can be used for teaching critical writing to the students.

**b) Teachers can encourage EFL students to use the CTER to do a self-and peer-assessment, so they become aware of their own learning, strengths and weaknesses and areas for improvement.**

The participants believe that doing the self-and peer-assessment by using the CTER helps EFL students realize the importance of demonstrating critical thinking in argumentative writing, be aware of their strengths and weaknesses of critical thinking and learn how to improve their own and their peers' critical thinking abilities. Thus, teachers need to provide the students with effective trainings on how to use the rubric, so they can understand the evaluative criteria and do the self-and peer-assessment more effectively (Cheng & Warren, 2005). The trainings include explaining crucial elements of critical thinking in argumentative essays to the students and demonstrating how to use the CTER to grade the essays. Teachers need to go through the scoring with the students in order to gain feedbacks about the scoring from the students and provide them with constructive comments about their self-and peer-assessment (Ross, 2006). Also, teachers can have the students read and analyze the anchor papers, so they can understand how to mark the essays through the rubric and evaluate their levels of performance more accurately (Bowman, 2017). This helps the students do the assessment by adhering to the evaluative criteria and avoiding over-estimating or under-estimating their own and their peers' abilities (Anderson, 2012). In conclusion, doing the self-and peer assessment by using the rubric promotes learning and critical thinking; that is, the rubric helps the students be aware of their own learning and be aware of thinking and writing an argument effectively (Bowman, 2017; Green, 2014; Light & Micari, 2013; Ornstein & Lasley, 2000; Widodo, 2012).

*I think the students can use the rubric to do a self-and peer-assessment, so they will be aware of their own learning and can use the constructive feedback from the rubric to improve their writing by themselves. (Rater A, Interview)*

*Student can use the rubric to do the self-and peer-assessment. For instance, when they evaluate their ability to write an introduction, they may know that they do not use any rhetorical strategy for gaining readers' attention at all. So, they will try to find and use such strategy. ... Also, they will know that an essay should have an introduction, a body, and a conclusion. ...They will know what a good argumentative essay looks like. (Rater C, Interview)*

*The rubric can help students develop their critical thinking abilities because the students can use it to do a self-assessment. They will know the components of argumentative writing. They will know whether and to what extent they can write the essays effectively and will know what their weaknesses are. The rubric can be a pathway for students to know that they need to improve and find out how to improve it. (Writing teacher 1, Questionnaire)*

Doing the self-and peer-assessment by using the CTER yields many other benefits. First, it promotes autonomous learning because the students have an opportunity to critique and improve their own abilities (Hedge, 2000; Hyland & Hyland, 2006). Second, it stimulates critical thinking. As Daud (2012) stated, an analytic rubric promotes critical thinking more effectively than a checklist because through using the analytic rubric, the students need to review the scoring criteria thoroughly and evaluate the levels of quality of each domain of the assessment judiciously. Also, the CTER helps the students challenge their peers to think more critically about writing, provide peers with constructive comments and apply new, useful ideas from peers into their own writing (Scheuer et al., 2014; Thongrin, 2002). Third, the CTER contains clear, detailed, diagnostic information which is the heart of “effective formative assessment” (Katz, 2012, p.72), so it helps the students revise their own and their peers' essays strategically. Tangkiengsirisin (2014) found that writing teachers did not

provide EFL students with clear, thorough, useful feedbacks, so when revising the essays, many students were not aware of crucial skills, such as writing an introduction strategically, writing a claim clearly, writing a final, constructive comment and using connectors to link ideas logically. Therefore, involving in the self- and peer-assessment and using useful feedbacks from the CTER to revise the essays can greatly help the students learn and improve their writing and critical thinking abilities (Bowman, 2017; Ennis, 1993; Stobart, 2012; Wang et al., 2017; Wilson & Post, 2019).

The findings of the study support the claim made by Reddy and Andrade (2010) that to improve students' English writing, just providing them with a rubric is not enough. Teachers need to encourage them to use the rubric to do the self- and peer-assessment and make revisions, so they can identify crucial elements of writing, evaluate their own and their peers' abilities meaningfully, gain feedbacks about their own strengths and weaknesses in writing and learn how to write the essays effectively themselves (Reddy & Andrade, 2010).

## **2. The CTER can help teachers plan and improve teaching and provide appropriate instructions that best serve the students' needs.**

The participants believe that the CTER provides teachers with scoring results that can identify levels of critical thinking abilities of learners, including their strengths and weakness and how to make improvement, so teachers can use these to plan and improve teaching or learning activities that are appropriate for the needs and levels of performance of the students. For example, teachers can use the CTER to assess students' critical thinking before, during and after they write essays in order to diagnose their abilities and be aware of their on-going development (Hyland, 2003; Saxton et al., 2002). If the students in the class have different levels of critical thinking abilities in writing, teachers can have them do group activities, such as collaborative writing and peer discussion, so more proficient students and less proficient ones can interact with and learn from each other (Katz, 2012). If most students have average

levels of critical thinking, teachers can use “a developmental approach” (Zhao, 2010, p.141) by teaching critical thinking skills appropriate for the students’ levels of performance at hand. For instance, teachers can teach them to use personal opinions or logical explanations as evidence to support a claim. When the students master these techniques, teachers can teach other kinds of evidence, such as research results and opinions of experts in the field, and challenge them to use various types of evidence. Also, teachers can use a teacher conference to give feedbacks that match with each student (Douglas, 2010). For instance, they can meet a student individually, ask him or her to elaborate unclear points in an essay, discuss how to use the CTER to improve the essay and provide the student with additional, constructive comments. If most students are proficient learners, teachers can teach them all critical thinking abilities identified in the CTER and challenge them to make progress in order to reach the score level of Excellent in all domains of the assessment. As Beck and his colleagues (2018) stated, “to improve writing instruction, teachers must be able to identify students’ challenge with writing and specify instruction that will address those challenges” (p.68). Undeniably, the CTER can help teachers collect and scrutinize the assessment data and improve and design teaching that 1) meets the students’ need (Coombe, Troudi, & Al-Hakly, 2012), 2) encourages weak students to write more productively and 3) challenges proficient ones to show their best critical thinking abilities (Farhady, 2012).

*I think we can use the scoring results from the rubric to improve teaching and students’ learning (Rater A, interview)*

*Clear, detailed evaluative descriptors in each domain of the assessment can help teachers know which critical thinking skills the students are still lacking and which skill they should fulfill in order to maximize their critical thinking abilities in argumentative writing. (Rater C, interview)*

*The score results identify students’ strengths and weaknesses in argumentative writing and can be used as a guideline for teachers to plan an instruction. This can*

*help the teachers know what they should teach in order to help the students master the target skills. (Writing teacher 1, Questionnaire)*

*The score results indicate levels of students' critical thinking abilities. When considering the score levels in each domain of the assessment, instructors will know strengths and weaknesses of the students and can plan how to improve teaching and foster students' critical thinking performances. (Writing teacher 2, Questionnaire)*

### **3. The CTER should be further improved, especially in the aspect of being time-consuming**

The other emerging theme indicates that the CTER has a limitation that the researcher should consider for further improving the rubric. That is, the participants thought that the CTER contains many detailed evaluative descriptors; therefore, using this rubric to grade a lot of students' argumentative essays requires a large amount of time and efforts from teachers. The following excerpts indicated the point.

*The rubric contains many detailed evaluative descriptors, so it is quite time consuming to use it to grade the students' essays. (Rater A, Interview)*

*Marking the essays by using the rubric takes time. Sometimes, while grading the essays, I cannot recognize the criteria, so I need to reread the rating manual and review the six domains of the assessments and the evaluative descriptors. For example, when I see the term 'rhetorical strategy', I need to review its definitions. Therefore, apart from your rubric, I need to have one paper and write detailed descriptions of the 'rhetorical strategy' in the paper so that I can remember them. (Writing teacher 1, Interview)*

*I spend around 5 – 6 minutes grading one essay through using the rubric, so if there are a large number of students, teachers may not have that much time to grade all of their essays by using the rubric. ... If there are just 10-20 students, then it is ok. However, if there are more than 20 students, it will be time-consuming. (Writing teacher 2, Interview)*

*It is time consuming to use the rubric to grade the essays, so it may not be appropriate to use the rubric to grade tons of students. (Writing teacher 3, Questionnaire)*

*I think the rubric is suitable for use in a small-scale assessment project. If it's a large-scale assessment project which you need to grade around 40-50 students, then grading will be very time-consuming. 20 students are ok. If there are tons of students, using this rubric may not be appropriate. (Writing teacher 3, Interview)*

This aspect of the study is in alignment with previous studies (e.g., Farhady, 2012; Han & Huang, 2017) which found that an analytic rubric may be impractical if it requires lots of time and energy from raters in scoring the essays. However, teachers can solve this problem by highlighting important descriptors in the rubric with color, such as highlighting the number of reasons and evidence in each score level, so the teachers can see the evaluative descriptors clearly and grade the essays faster. Also, the researchers can conduct a frequent, comprehensive and careful rater training, so the teachers are more familiar with the scoring criteria and can use the rubric to grade the essays faster (Chinda, 2014). Also, the validity is the most important aspect of the rubric; therefore, if grading the essays via the rubric is not too burdensome, teachers should use the rubric and ignore the practicality (Farhady, 2012; Knoch, 2009). Chinda (2013) argued that although it is time-consuming to use the analytic rubric, teachers should utilize it since it can greatly promote learning of EFL students, particularly Thai learners. In sum, it is worth using the CTER because its benefits are much greater than its limitation (Daud, 2012).

### **Summary of the raters' and the writing teachers' attitudes towards the CTER**

The previous section reports on the participants' opinions about the CTER. The qualitative data are collected from the raters' interview and the teachers' the questionnaire and the interview. The results show that although the CTER is time-consuming to use, it has the potential to promote teaching, learning, and critical thinking of EFL students. That is, the

CTER is a pedagogically useful tool which helps EFL students understand how to write and express critical thinking in argumentative writing effectively. Also, the students can use the CTER to do the self-and peer-assessment, so they can learn their own strengths and weaknesses about their critical thinking abilities. In addition, the CTER can help teachers plan and improve their own instructions in order to best serve the students' need.

### **4.3 Implications of the study**

In consideration of the findings, this study proposes three areas of pedagogical implications for EFL teachers.

First, teachers can provide EFL college students with an argumentative-genre based instruction effectively by using the CTER. To explain, the results of this study show that the CTER is a rubric that is valid and reliable and has clear, comprehensive descriptions about critical thinking features in argumentative writing. Therefore, teachers can use this rubric as a supplementary tool for teaching argumentative writing in order to promote thinking and writing skills of the students. For example, teachers can elucidate the crucial elements of critical thinking, such as rhetorical strategies, claims, reasons, evidence, counter-arguments, refutations, constructive comments and effective language use in English argumentation. This will help the students see a clearer picture about how to write an effective argument and be able to include these crucial features in their argumentative essays more effectively. Using the CTER in teaching writing can develop EFL students, especially Thai learners, into skillful writers because they will be able to 1) understand “the structural components or principles” of English “arguments” (Hanscomb, 2017; Newell et al., 2011, p.295), 2) realize that “Thai and English compositions have different styles and organisations” (Dhanarattigannon, 1995, p.70), 3) practice thinking and writing in English clearly (Liu, 2018; Qoura & Zahran, 2018), 4) improve communicative competence in English writing (Matsuda & Hammill, 2014; Purpura, 2014), 5) develop creative ideas in writing (Butterworth & Thwaites, 2013;

Kaewnuch, 2008; Thongrin, 2002, 2009), 6) express voice or opinion in writing effectively (Alagozlu, 2007; Barnawi, 2011; Matsuda & Matsuda, 2010; Mauk & Metz, 2019; Stapleton, 2002; Zhao, 2010, 2017), 7) be aware of audience by answering all potential questions in minds of readers (Brown, Glasswell & Harland, 2004; Malcolm, 2011; Winkler & McCuen, 1997) and 8) demonstrate knowledge of English rhetorical convention in argumentative writing (Tapinta, 2006).

Second, teachers can have EFL students use the CTER as an initial tool to assess and improve their own and their peers' argumentative writing abilities in each stage of a writing process. The CTER has clear domains of the assessment, clear rating scales and clear evaluative descriptors in each scale level; therefore, the students can use this rubric to assess their own and their peers' writing abilities in the process-based writing activities, such as planning, drafting, collaborative writing, self-evaluating, peer-conference and revising. According to Yu (2008), "the writing process is a recursive, dynamic and exploratory process in which writers search for their ideas, evaluate them and develop their arguments" (p.iii). Thus, incorporating the CTER in doing these writing activities, the students will be able to think, share and write their ideas critically. Also, they will be able to evaluate their arguments effectively, obtain useful, diagnostic feedbacks from the rubric, learn their own and their peers' strengths and weaknesses in writing, make a self-improvement and provide peers with constructive comments about writing (Alhosani, 2008; Morris, 2019; Newell et al., 2011; O'Donnell, 2015; Pan, 2008). In sum, the CTER contains "elements and evaluative criteria for argumentation" (Haria, 2010, p.80), so the students can use this rubric as a "guide" to assess and improve their writing effectively (Martens, 2018, p.31).

Third, teachers can use the CTER to foster critical thinking of EFL students. To explain, the CTER illustrates the six crucial domains of the critical thinking assessment in argumentative texts, including specific evaluative descriptors in each domain of the

assessment; therefore, the rubric provides the students with useful detailed information about not only writing an argument but also evaluating the argument. Hence, the students can use such information from the CTER as their knowledge and strategy that aid in writing, reading, analyzing and critiquing each aspect of an argumentative text effectively. For example, through using the CTER in reading a writer's argumentative text, EFL students can: 1) scrutinize, identify and evaluate important elements in the writer's argumentative text, such as a stance, rhetorical strategies, reasons, evidence, counter-arguments and a conclusion, 2) examine a fallacy in the writer's argument, 3) state their own position after reading the text, 4) challenge the writer to think more critically about his or her writing and 5) provide the writer with constructive comments about his or her writing abilities (Hanscomb, 2017; Haria & Midgette, 2014; Huh, 2016; Lada, 2015; Leist et al., 2012; Mayweg-Paus, Thiebach, & Jucks, 2016; Ploysangwal, 2018; Wilson, 2016). This will greatly develop analytical and critical thinking abilities of EFL learners.

#### **4.4 Limitations of the study**

First, the study had a necessity to conduct training each writing teacher how to implement the rubric individually. This is very likely to affect the inter-rater reliability. At that time, three writing teachers had urgent works and could not come and obtain a group training together. If all teachers had received the group training and had discussed the CTER and their scoring results altogether, they could have marked the essays through the rubric more consistently. Thus, to improve the quality of the scoring and increase the inter-rater reliability of the CTER, future researchers should provide the evaluators with the effective group training on using the rubric and have them discuss their scoring results in the training session thoroughly, so they were aware of their strengths, weaknesses and/or problems in the scoring, improve their grading and mark the essays through the CTER more reliably.

Second, the writing samples of this study are limited in terms of the number and representation of the population—i.e., EFL university students. To illustrate, in examining the inter-rater reliability of the CTER, this study employed the 20 writing samples which are a small sample size. Also, these writing samples come from only Thai EFL learners. Therefore, the findings may not be generalizable to other groups of the population in different contexts or different countries. Thus, to investigate the inter-rater reliability of the rubric, future studies should include more writing samples that have a larger sample size and represents other groups of EFL students.

#### **4.5 Recommendations for future research**

First, future researchers could explore how EFL students utilize the CTER in their writing process and to what extent they can improve their critical thinking and their ability to write argumentative essays after teachers integrate the CTER into their writing instruction. To investigate this reality, the researchers might use various research instruments, including a checklist (See Appendix 10) in order to gain complete data about how the students write argumentative essays, how they use the CTER to plan, design and improve their writing, what type of rhetorical strategies they use in writing introductions, what type of evidence they use to support their reasons, what are their perceptions about using the CTER and so on. This would yield a clear understanding of how the CTER promotes positive washback in a writing class or fosters critical thinking and writing of EFL college students.

Second, future researchers could investigate construct validity of the CTER by using factor analysis in order to strongly ensure the validity the rubric and have solid empirical evidence to support the findings. The current study used only the analysis of means and standard deviations, so future researchers could use more advanced statistical approaches, such as factor analysis, in investigating the construct validity of the CTER. As Ockey (2014) stated, constructs--or abilities--are factors or latent variables. The CTER identifies the six

crucial abilities as the operational definitions of the critical thinking construct in English argumentative essays of EFL students. Therefore, performing the factor analysis of this rubric could enable future researchers to “determine how many factors or constructs underlie a set of test scores” and “to what extent” and in what way “these factors are related to each other” (Ockey, 2014, p.1125). In sum, there might be a need to conduct empirical research to investigate construct validity of the CTER, and advanced statistical approaches, such as factor analysis, might be needed.

Third, future researchers should use both think-aloud protocol and follow-up interview to investigate the writing teachers’ opinions while they are using the CTER to mark the essays. As Han and Huang (2017) stated, using only interviews may provide incomplete information about scoring of the evaluators because they may not fully remember what went on while grading the essays. The think-aloud protocol requires the scorers to verbalize their thoughts out loud while grading the essays through the CTER. Therefore, using both think-aloud protocol and follow-up interview will enable future researchers to thoroughly investigate the rating and decision making process of the scorers while they are marking the essays through the CTER (East, 2009; Han & Huang, 2017; Hodges, 2019; Lim, 2009; Quintieri, 2005). This will also enable future researchers to learn problems in scoring of the writing teachers and use such information to improve the rubric and the rater training.

Fourth, to ensure the reliability of the CTER, it is interesting to investigate the intra-rater reliability by examining whether the rubric is still reliable when being used to mark the same essays by the same evaluator on two or more different scoring occasions. Having high intra-rater reliability is important because this indicates a self-consistency of the evaluator in scoring the essays by using the rubric (Weigle, 1994) and increases the confidence in interpreting the scoring results (Stobart, 2012). East (2009) suggested that there should be a gap around one month between the two scoring sessions, so the rater cannot accurately

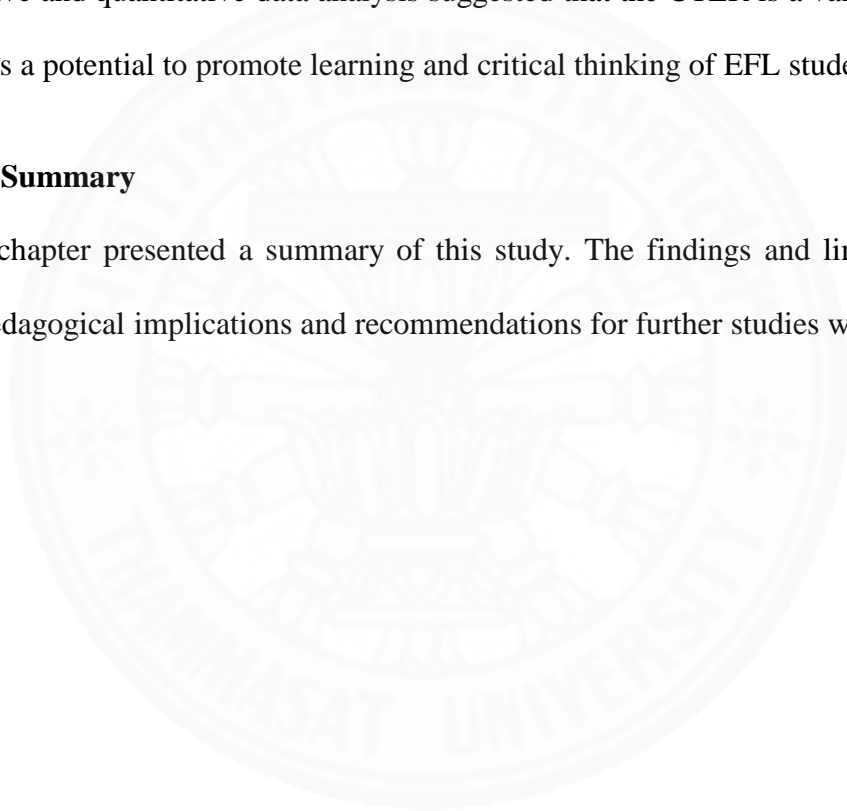
remember the score results in the first scoring occasions. Also, the researchers should investigate the intra-rater reliability of the rubric more than once in order to ensure the credibility of the findings. The intra-rater reliability of the rubric can be investigated by using statistical approaches, such as percent agreement and Cohen's Kappa (Ergai, 2013).

#### **4.6 Conclusions of the study**

This research aimed to develop and validate the CTER—an analytic rubric for assessing critical thinking in English argumentative essays of EFL students. Findings from both qualitative and quantitative data analysis suggested that the CTER is a valid and reliable rubric and has a potential to promote learning and critical thinking of EFL students.

#### **4.7 Chapter Summary**

This chapter presented a summary of this study. The findings and limitations were discussed. Pedagogical implications and recommendations for further studies were presented.



## REFERENCES

- Adler-Kassner, L., & Estrem, H. (2005). Critical thinking, reading, and writing: A view from the field. In S. Harrington, K. Rhodes, R.O. Fischer, & R. Malenczyk (Eds.), *The outcomes book: Debate and consensus after the WPA outcomes statement* (pp. 60 - 71). Logan, UT: Utah State University Press.
- Adunyarittigun, D. (1998). *The effects of reciprocal teaching procedure on Thai EFL students' reading performance and self-perception as readers* (Unpublished doctoral dissertation). The University of Maryland College Park.
- Adunyarittigun, D. (2015). Developing a scale to measure reader self-perception for EFL students. *PASAA*, 50, 1-30.
- Afflerbach, P., Cho, B., & Kim, J. (2015). Conceptualizing and assessing higher-order thinking in reading. *Theory into Practice*, 54, 203-212.
- Ahour, T., & Mukundan, J. (2009). Analytic assessment of writing: Diagnosing areas of strength and weakness in the writing of TESL undergraduate students. *Iranian Journal of Language Studies*, 3(2), 195-208.
- Akbari, R. (2012). Validity in language testing. In C. Coombe, P. Davidson, B. O'Sullivan & S. Stoyloff (Eds.), *The Cambridge guide to second language assessment* (pp.30-36). Cambridge: Cambridge University Press.
- Alagozlu, N. (2007). Critical Thinking and voice in EFL writing. *The Asian EFL Journal Quarterly*, 9(3), 118-136.
- Alhosani, N. M. (2008). *Utilizing the writing process approach with English as a second language writers: A case study of five grade ESL Arab students* (Unpublished doctoral dissertation). Kansas State University.

- Alsina, Á., Ayllón, S., Colomer, J., Fernández-Peña, R., Fullana, J., Pallisera, M., Pérez Burriel, M., Serra, L. (2017). Improving and evaluating reflective narratives: A rubric for higher education students. *Teaching and Teacher Education*, 63, 148-158.
- Al-Musaad, H. (2001). *A measurement of the critical thinking ability of college students with learning disabilities* (Unpublished doctoral dissertation). The Pennsylvania State University.
- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Addison Wesley Longman
- Anderson, N. J. (2012). Student involvement in assessment: Healthy self-assessment and effective peer assessment. In C. Coombe, P. Davidson, B. O'Sullivan & S. Stoyhoff (Eds.), *The Cambridge guide to second language assessment* (pp.187-197). Cambridge: Cambridge University Press.
- Andrade, H. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57, 13-18.
- Angeli, C., & Valanides, N. (2009). Instructional effects on critical thinking: Performance on ill- defined issues. *Learning and Instruction*, 19, 322-334.
- Angelo, T. A. (1995). Classroom assessment for critical thinking. *Teaching of Psychology*, 22(1), 6-7.
- Annis, D. B. (1974). *Techniques of critical reasoning*. Ohio: Charles E. Merrill Publishing Company.
- Arter, J., & McTighe, J. (2001). *Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance*. New York: Corwin Press.
- Association of American Colleges and Universities. (2012). *Critical Thinking VALUE Rubric*. Retrieved from <https://www.aacu.org/value/rubrics/critical-thinking>

- Atkinson, D. (1997). A critical approach to critical thinking in TESOL. *TESOL Quarterly*, 31(1), 71-94.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Bailey, A., Zanchetta, M., Velasco, D., Pon, G., & Hassan, A. (2015). Building a scholar in writing (BSW): A model for developing students' critical writing skills. *Nurse Education in Practice*, 15, 524-529.
- Barnawi, O. (2011). Finding a place for critical thinking and self-voice in college English as a foreign language writing classrooms. *English Language Teaching*, 4(2), 190-197.
- Barnet, S., & Bedau, H. (1993). *Critical thinking, reading, and writing: A brief guide to argument*. New York, NY: St. Martin's Press, Inc.
- Bauer, J. A. (2016). *A new approach: Closing the writing gap by using reliable assessment to guide and evaluate cross-curricular argumentative writing* (Master's thesis). University of Wisconsin.
- Bean, J. C. (2011). *Engaging ideas: The professor's guide to integrating writing, critical thinking, and active learning in the classroom* (2nd Ed.). San Francisco, CA: The Jossey-Bass higher and adult education series.
- Beck, S. W., Llosa, L., Black, K., Anderson, A. T.G. (2018). From assessing to teaching writing: What teachers prioritize. *Assessing Writing*, 37, 68-77.
- Becker, A. (2016). Student-generated scoring rubrics: Examining their formative value for improving ESL students' writing performance. *Assessing Writing*, 29, 15-24.
- Ben-Simon, A., & Bennett, R. E. (2007). Toward more substantively meaningful automated essay scoring. *The Journal of Technology, Learning, and Assessment*, 6(1), 1-46.
- Bhumibol Adulyadej, His Majesty King. (1997). *The story of Mahajanaka*. Bangkok: Amarin Printing and Publishing Company Limited.

- Bierman, A. K., & Assali, R. N. (1996). *The critical thinking handbook*. New Jersey: Prentice-Hall, Inc.
- Billups, N. Z. (2006). *An analysis of the impact of America's choice program on critical thinking and writing skills of selected fifth grade students* (Unpublished doctoral dissertation). Fielding Graduate University.
- Blattner, N. H., & Frazier, C. L. (2002). Developing a performance-based assessment of students' critical thinking skills. *Assessing Writing*, 8, 47-67.
- Bloom, B. S., Engelhart, M.D., Furst, E.J., Hill, W.H., & Krathwohl, D.R. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook 1: Cognitive domain*. New York: David McKay.
- Bolderston, A. (2012). Conducting a research interview. *Journal of Medical Imaging and Radiation Sciences*, 43, 66-76.
- Bowman, J. E. (2017). *Engaging students in the assessment process: A quantitative study of peer- and self-assessment* (Unpublished doctoral dissertation). Notre Dame of Maryland University
- Breland, H. M. (1983). *The Direct assessment of writing skills: A measurement review*. College Board Report No. 83-6. (ETS RR No. 83-32.). NY: College Entrance Examination Board.
- Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. Alexandria, VA: ASCD.
- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices* (2nd ed.). White Plain, NY: Pearson Education, Inc.
- Brown, G., T. L., Glasswell, K., & Harland, D. (2004). Accuracy in the scoring of writing: Studies of reliability and validity using a New Zealand writing assessment system. *Assessing Writing*, 9, 105-121.

- Bunsom, T., Vungthong, S., & Singhasiri, W. (2011). Developing engineering students' critical thinking skills through reading short stories. *Journal of English Studies, 6*, 118-140.
- Burkhalter, N. (1993). *How persuasive writing aids critical thinking*. Paper presented at the Annual Meeting of the Speech Communication Association, Florida, U.S.A.
- Burkhalter, N. (1995). A Vygotsky-based curriculum for teaching persuasive writing in the elementary grades. *Language Arts, 72*(3). 192-199.
- Butterworth, J., & Thwaites, G. (2013). *Thinking skills: Critical thinking and problem solving* (2nd ed.). Cambridge: Cambridge University Press.
- CAAP Program Management. (2012). *ACT CAAP technical handbook 2011–2012*. Iowa City, IA: Author. Retrieved from <http://www.act.org/caap/pdf/CAAPTechnicalHandbook.pdf>
- Carbone, P. M. (2009). *Investigating a critical writing pedagogy: Implication for classroom practice* (Unpublished doctoral dissertation). University of California.
- Cargas, S., Williams, S., & Rosenberg, M. (2017). An approach to teaching critical thinking across disciplines using performance tasks with a common rubric. *Thinking skills and creativity, 26*, 24-37.
- Carter, A., Creedy, D. K., & Sidebotham, M. (2017). Critical thinking evaluation in reflective writing: Developing and testing of Cater Assessment of critical thinking in Midwifery (Reflection). *Midwifery, 54*, 73-80.
- Carwie, L. (2009). *The effect of computer-supported argument mapping on the critical thinking skills of undergraduate nursing students* (Unpublished doctoral dissertation). University of South Alabama.

- Castilleja, M. F. P. (2012). *Developing a process to create and validate an instrument assessing student attainment of competencies at an intercultural university in Mexico* (Unpublished doctoral dissertation). Florida International University.
- Çavdar, G., & Doe, S. (2012). Learning through writing: Teaching critical thinking skills in writing assignments. *American Political Sciences Association*, 45(2), 298-306.
- Celuch, K., & Slama, M. (2010). Teaching critical thinking skills for the 21st century: An advertising principles case study. *Journal of Education for Business*, 74(3), 134-139.
- Chan, N., Ho, I., T., & Ku, K., Y.L. (2011). Epistemic beliefs and critical thinking of Chinese students. *Learning and Individual Differences*, 21, 67-77.
- Chan, S., Inoue, C., & Taylor, L. (2015). Developing rubrics to assess the reading-into writing skills: A case study. *Assessing Writing*, 26, 20-37.
- Chapelle, C. A., & Brindley, G. (2010). Assessment. In N. Schmitt (Ed.), *An introduction to applied linguistics* (2nd ed., pp.247-267). London: Hodder Education.
- Chase, B. J. (2011). *An analysis of the argumentative writings of academically underprepared college students* (Unpublished doctoral dissertation). Columbia University.
- Chatfield, T. (2018). *Critical thinking: Your guide to effective argument, successful analysis and independent study*. London: Sage Publications Ltd.
- Chen, M. M. (2017). *Writing and critical thinking: A case study of community college students in English courses* (Unpublished doctoral dissertation). Claremont Graduate University.
- Cheng, F. (2003). *The effects of rhetorical specification in writing assignments on EFL (English as a foreign language) writing* (Unpublished doctoral dissertation). Boston University.

- Cheng, W., & Warren, M. (2005). Peer assessment of language proficiency. *Language Testing*, 22(1), 93-121.
- Chinda, B. (2013). Consideration in performance-based language assessment: Rating scales and rater training. *PASAA*, 46, 141-159.
- Chinda, B. (2014). Teachers' reactions towards performance-based language assessment. *PASAA*, 48, 57-88.
- Chowske, R. D. (2013). *Discovering the common grading preferences, patterns, and policies of those who assess writing* (Unpublished doctoral dissertation). Long Island University.
- 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.
- Connor, U. (1990). Linguistic/rhetorical measures for international persuasive student writing. *Research in Teaching of English*, 24(1), 67-87.
- Connor, U. (1996). *Contrastive rhetoric*. New York: Cambridge University Press.
- Connor, U., & Lauer, J. (1985). Understanding persuasive essay writing: Linguistic/rhetorical approach. *Text-Interdisciplinary Journal for the Study of Discourse*, 5(4), 309-326.
- Coombe, C., Troudi, S., & Al-Hamly, M. (2012). Foreign and second language teacher assessment literacy: Issues, challenge, and recommendations. In C. Coombe, P. Davidson, B. O'Sullivan & S. Stoyloff (Eds.), *The Cambridge guide to second language assessment* (pp. 20-29). Cambridge: Cambridge University Press.
- Cottrell, S. (2011). *Critical thinking skills: Developing effective analysis and argument*. Basingstoke: Palgrave MacMillan.
- Cozby, P. C. (2001). *Methods in behavioral research* (10th ed.). Boston: McGraw-Hill.

- Creighton, S. D. (2006). *Examining alternative scoring rubrics on a statewide test: The impact of different scoring methods on science and social studies performance assessments* (Unpublished doctoral dissertation). University of South Carolina.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: SAGE Publications.
- Crusius, T. W., & Channell, C. E. (2000). *The aims of argument: A brief rhetoric*. Mountain View, CA: Mayfield Publishing Company.
- Daud, N. S. B. M. (2012). *Developing critical thinking skills in tertiary academic writing through the use of an instructional rubric for peer evaluation*. (Doctoral Dissertation). The University of Canterbury.
- Davidson, B. W. (1998). Comments on Dwight Atkinson's "A critical approach to critical thinking in TESOL": A case for critical thinking in the English language classroom. *TESOL Quarterly*, 32(1), 119-123.
- Davidson, B. W., & Dunham, R. L. (1996). *Assessing EFL Student Progress in Critical Thinking*. Paper presented at the Annual International Conference of the Japan Association for Language Teaching, Nagoya, Japan.
- Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods* (2nd ed.). New York: McGraw-Hill.
- Dhanarattigannon, J. (1995). Journal writing: Language activity and teaching/learning evaluation. *Humanities Journal*, 3(1), 69-80.
- Dhanarattigannon, J. (2008). *Thai college students' response to nontraditional writing instruction in a Thai university* (Unpublished doctoral dissertation). University of Florida.
- Dickinson, P., & Adams, J. (2017). Values in evaluation – The use of rubrics. *Evaluation and Program Planning*, 65, 113–116.

- Dong, Y. (2015). *Critical thinking in second language writing: Concept, theory, and pedagogy* (Unpublished doctoral dissertation). The University of British Columbia.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. New York: Oxford University Press.
- Douglas, D. (2010). *Language testing* (1st ed.). London: Hodder Education.
- East, M. (2009). Evaluating the reliability of a detailed analytic scoring rubric for foreign language writing. *Assessing Writing, 14*, 88-115.
- ETS [Educational Testing Service] (2010). *ETS Proficiency Profile user's guide*. Princeton, NJ: Author.
- ETS [Educational Testing Service] (2015). *HEIghten™ critical thinking test at a glance*. Retrieved from [https://www.ets.org/s/heighten/pdf/heighten\\_critical\\_thinking\\_test\\_at\\_a\\_glance\\_acc.pdf](https://www.ets.org/s/heighten/pdf/heighten_critical_thinking_test_at_a_glance_acc.pdf)
- ETS [Educational Testing Service] (2019). *The GRE® Analytical Writing measure*. Retrieved from [https://www.ets.org/gre/revised\\_general/prepare/analytical\\_writing/score\\_level\\_descriptions](https://www.ets.org/gre/revised_general/prepare/analytical_writing/score_level_descriptions)
- Eggington, W. G. (2015). When everything's right, but it's still wrong: Cultural influences on written discourse. In N. W. Evans, N. J. Anderson, & W. G. Eggington (Eds.), *ESL readers and writers in higher education: Understanding challenges, providing support* (pp. 198-208). New York, NY: Routledge.
- Ennis, R. H. (1993). Critical thinking assessment. *College of Education, 32*(3),179-186.
- Ennis, R. H., Millman, J., & Tomko, T. N. (1985). *Cornell critical thinking test level X and level Z manual*. 3rd. ed. Pacific grove, CA: Midwest publications.

- Ennis, R. H., & Weir, E. (1985). *The Ennis–Weir Critical Thinking Essay Test*. Pacific Grove, CA: Midwest Publications.
- Epstein, R. L. (1999). *Critical thinking*. Belmont, CA: Wadsworth.
- Ergai, A. O. (2013). *Assessment of the human factors analysis and classification system (HFACS): Intra-rater and inter-rater reliability* (Unpublished doctoral dissertation). Clemson University.
- Erwin, T. D., & Sebrell, K. W. (2003). Assessment of critical thinking: ETS's tasks in critical thinking. *The Journal of General Education*, 52(1). 50-70.
- Eva, W. (2018). *Critical thinking in intensive language programs for international students in U.S. universities* (Unpublished doctoral dissertation). University of Pittsburgh.
- Facione, P. A. (1990a). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. Millbrae, CA: California Academic Press.
- Facione, P. A. (1990b). *The California Critical Thinking Skills Test-college level. Technical report #2. Factors predictive of CT skills*. Millbrae, CA: California Academic Press.
- Facione, P. A. (2018). Critical thinking: What it is and why it counts. *Insight Assessment*, 1-28.
- Facione, P. A., & Facione, N. C. (1996). *CT in GE: Measuring thinking in college*. Workshop at the meeting of the American Association of Higher Education, Washington, DC.
- Farhady, H. (2012). Principles of language assessment. In C. Coombe, P. Davidson, B. O'Sullivan & S. Stoyloff (Eds.), *The Cambridge guide to second language assessment* (pp.37-46). Cambridge: Cambridge University Press.
- Fiken, M., & Ennis, R. H. (1993). *Illinois critical thinking essay test*. Retrieved from <http://evolkov.net/critic.think/tests/IIICTEssayTestFinken-Ennis12-1993LowR.pdf>
- Fisher, A. (2011). *Critical Thinking: An Introduction* (2nd ed.). New York, NY: Cambridge University Press.

- Fleming, D. (2014). Rhetoric and argumentation. In G. Tate, A. R. Taggart, K. Schick, & H. B. Hessler (Eds.), *A guide to composition pedagogies* (2nd ed., pp. 248-265). Oxford: Oxford University Press.
- Folse, K. S., Muchmore-Vokoun, A., & Solomon E., V. (2002). *Great Essays: An Introduction to writing essays* (2nd ed.). Boston, MA: Patricia A. Coryell.
- Foresman, G. A., Fosl, P. S., & Watson, J. C. (2017). *The critical thinking toolkit*. West Sussex: John Wiley & Sons, Inc.
- Franklin, D., Weinberg, J., & Reifler, J. (2014). Teaching writing and critical thinking in large political science classes. *Journal of Political Science Education*, 10(2), 155-165.
- Fulan, L. (2014). *Investigating critical thinking in the argumentative writing of English majors at a mainland Chinese university: Implications for policy changes* (Unpublished doctoral dissertation). The Hong Kong Institute of Education.
- Fulkerson, R. (1996). The Toulmin model of argument. In B. Emmel, & P. Resch (Eds.), *Argument revisited, argument refined: Negotiating meaning in the composition classroom* (pp. 45-72). Thousand Oaks: Sage.
- Gardner, R. C. (1985). *Social psychology and language learning: The role of attitudes and motivation*. London: Edward Arnold.
- Gebriel, A., & Plakans, L. (2014). Assembling validity evidence for assessing academic writing: Rater reactions to integrated tasks. *Assessing Writing*, 21, 56-73.
- Goodfellow, L. L. T. (1999). *A self-assessment of critical thinking behaviors in respiratory care* (Unpublished doctoral dissertation). The University of Georgia.
- Grabe, W., & Kaplan, R. B. (1996). *Theory and practice of writing: An applied linguistic perspective*. London: Longman.
- Green, A. (2014). *Exploring language assessment and testing: Language in action*. New York, NY: Routledge.

- Green, A., & Hawkey, R. (2012). Marking assessments: Rating scales and rubrics. In C. Coombe, P. Davidson, B. O'Sullivan & S. Stoyhoff (Eds.), *The Cambridge guide to second language assessment* (pp.299-306). Cambridge: Cambridge University Press.
- Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11(3), 255-274.
- Gronlund, N. E. (1998). *Assessment of student achievement* (6th ed.). Boston, MA: Allyn & Bacon.
- Gustafson, M., & Bochner, J. (2009). Assessing critical thinking skills in students with limited English proficiency. *Assessment Update*, 21 (4), 8–10.
- Hager, P., & Kaye, M. (1992). Critical thinking in teacher education: A process-oriented research agenda. *Australian Journal of Teacher Education*, 17(2), 26-33.
- Halpern, D. (1993). Assessing the effectiveness of critical thinking instruction. *The Journal of General Education*, 50(4), 270-286.
- Halpern, D. (1994). *Changing college classroom: New teaching and learning strategies for an increasing complex world*. San Francisco: Jossey-Bass.
- Halsted, S. E. W. (1998). *Facilitating creative and critical thinking in middle school science* (Unpublished doctoral dissertation). Pacific Lutheran University.
- Hammer, S. (2017). An examination of the instruction provided in Australian essay guides for students' development of a critical viewpoint. *Assessment & Evaluation in Higher Education*, 42(7), 1069-1081.
- Han, T., & Huang, J. (2017). Examining the impact of scoring methods on the instructional EFL writing assessment: A Turkish perspective. *PASAA*, 53, 112-147.
- Hanscomb, S. (2017). *Critical thinking: The basics*. New York, NY: Routledge.

- Haria, P. D. (2010). *The effects of teaching a genre-specific reading comprehension strategy on struggling fifth grade students' ability to summarize and analyze argumentative texts* (Unpublished doctoral dissertation). The University of Delaware.
- Haria P. D., & Midgette, E. (2014). A genre-specific reading comprehension strategy to enhance struggling fifth-grade readers' ability to critically analyze argumentative text. *Reading & Writing Quarterly*, 30(4), 297-327.
- Harvin, D. (2016). *Do "superstar" CEOs impair auditors' independence and professional skepticism?* (Unpublished doctoral dissertation). Florida Atlantic University.
- Haswell, R. H. (1991). *Gaining ground in college writing: Tales of development and interpretation*. Dallas: Southern Methodist University Press.
- Hedge, T. (2000). *Teaching and learning in the language classroom: A guide to current ideas about the theory and practice of English language teaching*. New York: Oxford University Press.
- Hillocks, G., Jr. (2010). Teaching argument for critical thinking and writing: An introduction. *English Journal*, 99(6), 24-32.
- Hirschberg, S. (1996). *Strategies of argument* (2nd ed.). Boston: Allyn & Bacon.
- Ho, V. L. (2011). *Non-native argumentative writing by Vietnamese learners of English: A contrastive study* (Unpublished doctoral dissertation). Georgetown University.
- Hodges, T. S., Wright, K. L., Winda, S. A., Matthews, S. D., Zimmer, W. K., & McTigue, E. (2019). Developing and examining validity evidence for the Writing Rubric to Inform Teacher Educators (WRITE). *Assessing Writing*, 40, 1-13.
- Hollihan, T. A., & Baaske, K. T. (2005). *Arguments and arguing: The products and process of human decision making* (2nd ed.). Long Grove, IL: Waveland Press, Inc.
- Hua, Z. (2014). *Exploring intercultural communication: Language in action*. New York, NY: Routledge.

- Huang, J. (2007). *Examining the fairness of rating ESL students' writing on large-scale assessments* (Unpublished doctoral dissertation). Queen's University.
- Huck, S. W. (2012). *Reading statistics and research* (6th ed.). Boston, MA: Pearson.
- Hughes, A. (2003). *Testing for language teachers* (2nd ed.). New York, NY: Cambridge University Press.
- Hughes, C. (2000). *A Comparative study of teaching critical thinking through persuasive writing to average, gifted and students with learning disabilities* (Unpublished doctoral dissertation). The College of William and Mary.
- Huh, S. (2016). Instructional model of critical literacy in an EFL context: Balancing conventional and critical literacy. *Critical Inquiry and language Studies*, 13(3), 210-235.
- Hyland, K. (2016). Methods and methodologies in second language writing research. *System*, 59, 116-125.
- Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, 39, 77-95.
- Inch, E. S., & Warnick, B. (2002). *Critical thinking and communication: The use of reason in argument* (4th ed.). Boston, MA: Allyn and Bacon.
- Ismail, S. (2010). *Arabic and English persuasive writing of Arabs from a contrastive rhetoric perspective* (Unpublished doctoral dissertation). Indiana University of Pennsylvania.
- James, M., Black, P., Carmichael, P., Connor, C., Dudley, P., Fox, A., Frost, L., Hornor, L., MacBeath, J., McCormick, R., Marshall, B., Pedder, D., Procter, R., Swaffield, S., & Wiliam, D. (2006). *Learning how to learn: Tools for schools*. London, Routledge.
- Janssen, G., Meier, V., & Trace, J. (2015). Building a better rubric: Mixed methods rubric revision. *Assessing Writing*, 26, 51-66.

- Jantrasakul, P. (2004). *School reform and critical thinking: Policy and practice in Thai EFL classrooms* (Unpublished doctoral dissertation). Indiana University.
- Jeffery, J. V. (2009). Constructs of writing proficiency in US state and national writing assessments: Exploring variability. *Assessing Writing*, 14(1), 3-24.
- Johnson, B., & Onwuegbuzie, A. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14-26.
- Jones, E. A., Dougherty, B. C, Fantaske, P., & Hoffman, S. (1997). *Identifying college graduates' essential skills in reading & problem solving: Perspectives of faculty, employers & policymakers*. University Park, PA: U.S. Department of Education/OERI.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2, 130-144.
- Kaewnuch, S. (2008). *Teaching agency and power as social: Creating transformative subjects in the clashes of modernity and postmodernity in Thai EFL writing classrooms* (Unpublished doctoral dissertation). Illinois State University.
- Kalchayanant, T. (2009). *Writing essays*. Bangkok: Thammasat University Press.
- Kang, R. J. L. (2015). *Assessment matters: Negotiating The New York State English Regents exam* (Unpublished doctoral dissertation). Columbia University.
- Kaplan, R. B. (1996). Cultural thought patterns in inter-cultural education. *Language learning*, 16(1-2), 11-25.
- Katz, A. (2012). Linking assessment with instructional aims and learning. In C. Coombe, P. Davidson, B. O'Sullivan & S. Stoyhoff (Eds.), *The Cambridge guide to second language assessment* (pp.66-73). Cambridge: Cambridge University Press.

- Kaven, J. (2013). *The development of a valid and reliable general analytic rubric for a college level public-speaking course* (Unpublished doctoral dissertation). California State University.
- Ka-kan-dee, M., & Kaur, Sarjit. (2015). Teaching strategies used by Thai EFL lecturers to teach argumentative writing. *Procedia – Social and Behavioral Sciences*, 208, 143-156.
- Khaimukd, B. T. (1999). *Points of connection: A qualitative study of dialogue journals in an EFL college writing classroom in Thailand* (Unpublished doctoral dissertation). Indiana University of Pennsylvania.
- Khopitak, A. (2015). *Metadiscourse in newspaper opinion articles: A comparative study between American and Thai English-language newspapers* (Master's thesis). Thammasat University.
- Kirby, G. R., & Goodpaster, J. R. (2007). *Thinking: An interdisciplinary approach to critical and creative thought* (4th ed.). New Jersey: Pearson Education, Inc.
- Knoch, U. (2007). *Diagnostic writing assessment: The development and validation of a rating scale* (Unpublished doctoral dissertation). The University of Auckland.
- Knoch, U. (2009). Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing*, 26(2), 275-304.
- Knoch, U. (2011). Rating scales for diagnostic assessment of writing: What should they look like and where should the criteria come from? *Assessing Writing*, 16, 81-96.
- Kobayashi, H. (1984). Rhetorical patterns in English and Japanese. *TESOL Quarterly*, 18(4), 737-738.
- Koo, T. K., & Li, BPS. M. L. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15, 155-163.

- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice, 4*(41), 212-218.
- Kubota, R. (1998). An investigation of L1-L2 transfer in writing among Japanese university students: Implications for contrastive rhetoric. *Journal of Second Language Writing, 7*(1), 69-100.
- Ku, K. (2009). Assessing students' critical thinking performance: Urging for measurements using multi-response format. *Thinking Skills and Creativity, 4*, 70-76.
- Ku, K., Lai, E., & Hau, K. (2014). Epistemological beliefs and the effect of authority on argument-counterargument integration: An experiment. *Thinking Skills and Creativity, 13*, 67-79.
- Kuhs, T. M., Johnson, R. L., Agruso, S. A., & Monrad, D. M. (2001). *Put to the test: Tools and techniques for classroom assessment*. Portsmouth, NH: Heinemann.
- Lada, S. (2015). *An exploration of the use of project-based language learning to develop critical reading skills of Thai university students* (Unpublished doctoral dissertation). Thammasat University.
- Lallmamode, S. P., Daud, N. M., & Kassim, N. L. A. (2016). Development and initial argument-based validation of a scoring rubric used in the assessment of L2 electronic portfolios. *Assessing Writing, 30*, 44-62.
- Lannon, J. M. (2009). *The writing process* (3rd ed.). New York: Pearson Education, Inc.
- Larsson, K. (2017). Understanding and teaching critical thinking—A new approach. *International Journal of Educational Research, 84*, 32-42.
- Lau, J. Y. F. (2011). *An introduction to critical thinking and creativity: Think more, think better*. New Jersey: John Wiley & Sons, Inc.

- Leist, C. W., Woolwine, M. A., & Bays, C. Y. (2012). The effect of using a critical thinking a scoring rubric to assess undergraduate students' reading skills. *Journal of College Reading and Learning, 43*(1), 31-58.
- Lerner, P. J. (2000). *Graduate college critical thinking: Implications for argumentativeness, verbal aggression and organizational leadership* (Unpublished doctoral dissertation). University of Oklahoma
- Leverenz, C. S., (2014). Design thinking and the wicked problem of teaching writing. *Computer and Composition, 33*, 1-12.
- Light, G., & Micari, M. (2013). *Making six principles for effective college teaching scientists*. Cambridge: Harvard University Press.
- Lim, G. S. (2009). *Prompt and rater effects in second language writing performance assessment* (Unpublished doctoral dissertation). The University of Michigan.
- Liu, F., & Stapleton, P. (2014). Counterargumentation and the cultivation of critical thinking in argumentative writing: Investigating washback from a high-stakes test. *System, 45*, 117-128.
- Liu, F., & Stapleton, P. (2018). Connecting writing assessment with critical thinking: An exploratory study of alternative rhetorical functions and objects of enquiry in writing prompts. *Assessing Writing, 38*, 10-20.
- Liu, L. (2006). *Understanding rhetorical traditions, rethinking writing pedagogies: A cross-national study of written argumentation instruction in secondary school textbooks* (Unpublished doctoral dissertation). Purdue University.
- Liu, O. L., Frankel, L., & Roohr, K. C. (2014). Assessing critical thinking in higher education: Current state and directions for next-generation assessment. *ETS Research Report Series, 2014*(1), 1-23.

- Liu, O. L., Mao, L., Frankel, L., & Xu, J. (2016). Assessing critical thinking in higher education: The HEIghten™ approach and preliminary validity evidence. *Assessment & Evaluation in Higher Education, 41*(5), 677-694.
- Liu, O. L., Shaw, A., Gu, L., Li, G., Hu, S., Yu, N., Ma, L., Xu, C., Guo, F., Su, Q., Kardanovaj, E., Chirikov, I., Shi, J., Shi, Z., Wang, H., & Loyalka, P. (2018). Assessing college critical thinking: Preliminary results from the Chinese HEIghten® critical thinking assessment. *Higher Education Research & Development, 37*(5), 999-1014.
- Liu, P. E., Wu, W. V., & Shieh, R. (2015). Enhancing EFL students' critical thinking and writing: An asynchronous debate instructional design. *English Teaching & Learning, 39*(3), 33-59.
- Malcolm, B. (2011). Say What?: The use of voice in teaching creative writing. *Journal of Letters, Special Issue, 40*(2), 104-125.
- MamourChoul, T. K. (2010). *Developing critical thinking skills through integrative teaching of reading and writing in the L2 writing instruction* (Unpublished doctoral dissertation). University of Newcastle Upon Tyne.
- Manalo, E., & Sheppard, C. (2016). How might language affect critical thinking performance? *Thinking Skills and Creativity, 21*, 41-49.
- MAP Critical Thinking Steering Committee. (1984). *MAP critical thinking booklet*. Pittsburgh Public Schools.
- Martens, K. S. R. (2018). How program evaluators use and learn to use rubrics to make evaluative reasoning explicit. *Evaluation and Program Planning, 69*, 25-32.
- Marzano, R. J., Pickering, D., & McTighe, J. (1993). *Assessing student outcomes: Performance assessment using the dimensions of learning model*. Alexandria, Virginia: Association for Supervision and Curriculum Development.

- Mateus, S. (2019). Affective rhetoric: What it is and why it matters. In L. Zhang, & C. Clark (Eds.), *Affective, emotion, and rhetorical persuasion in mass media communication* (pp. 69-80). New York, NY: Routledge.
- Matsuda, A., & Matsuda, P. (2010). World Englishes and the teaching of writing. *TESOL Quarterly*, 44(2), 369-374.
- Matsuda, P. K., & Hammill, M. J. (2014). Second language writing pedagogy. In G. Tate, A. R. Taggart, K. Schick, & H. B. Hessler (Eds.), *A guide to composition pedagogies* (2nd ed., pp. 266-282). Oxford: Oxford University Press.
- Matthews, D. H. (2000). *Anchor paper selection procedures and their impact on direct writing assessment: A case in point* (Unpublished doctoral dissertation). Tennessee State University
- Mauk, J., & Metz, J. (2019). *The composition of everyday life* (concise 6th ed). Boston, MA: Cengage Learning, Inc.
- Maybin, J., & Swann, J. (Eds.) (2006). *The art of English: Everyday creativity*. New York: Palgrave Macmillan.
- Mayer, R. E. (2002). Rote versus meaningful learning. *Theory into practice*, 41(4), 226-232.
- Mayfield, M. (1994). *Thinking for yourself: Developing critical thinking skills through reading and writing* (3rd ed.). Belmont, CA: Wadsworth.
- Mayweg-Paus, E., Thiebach, M., & Jucks, R. (2016). Let me critically question this! Insights from a training study on the role of questioning on argumentative discourse. *International Journal of Educational Research*, 79, 195-210.
- McKinley, J. (2013). Displaying critical thinking in EFL academic writing: A discussion of Japanese to English contrastive rhetoric. *RELC Journal*, 44(2), 195-208.
- Mckitrick, S. A., & Barnes, S. M. (2012). Assessment of critical thinking: An evolutionary approach. *Journal of Assessment and Institutional Effectiveness*, 2(1), 1-29.

- McLaughlin, F., & Moore, M. (2012). Integrating critical thinking into the assessment of college writing. *TETYC*, 40(2), 145-162.
- McLendon, N. C. G. (2008). *The effects of teaching critical thinking and reading comprehension strategies on students' writing in developmental English in a community college* (Unpublished doctoral dissertation). Auburn University.
- McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.
- Mearman, K. A. (2013). *A statistical estimate of the validity and reliability of a rubric by Connecticut's state education resource center to evaluate the quality of individualized education programs for students with disabilities* (Unpublished doctoral dissertation). Andrews University.
- Mehta, S. R., & Al-Mahrouqi, R. (2014). Can critical thinking be taught? Linking critical thinking and writing in an EFL context. *RELC Journal*, 46(1), 1-14.
- Nejmaoui, N. (2019). Improving EFL learners' critical thinking skills in argumentative writing. *English Language Teaching*, 12(1), 98-109.
- Merriam, S. B. (1998). *Qualitative research and case study applications in education*. San Francisco: Jossey-Bass Publishers.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749. Retrieved from [http://people.tamu.edu/~w-arthur/611/Journals/Messick%20\(1995\)%20AP.pdf](http://people.tamu.edu/~w-arthur/611/Journals/Messick%20(1995)%20AP.pdf)
- Messick, S. (1996a). Validity and washback in language testing. *Language Testing*, 13(3), 241-256.
- Messick, S. (1996b). Validity of psychological assessment. *American Psychologist*, 50, 741-749.

- Metcalfe, K. C. (2007). *Assessing a feasible effective critical thinking protocol for community college* (Unpublished doctoral dissertation). Capella University.
- Morozov, A. (2011). Student attitudes toward the assessment criteria in writing-intensive college courses. *Assessing Writing, 16*, 6–31.
- Morris, J. A., Miller, B. W., Anderson, R. C., Nguyen-Jahiel, K. T., Lin, T., Scott, T., Zhang, J., Sun, J., Ma, S. (2019). Instructional discourse and argumentative writing. *International Journal of Educational Research, 90*, 234–247
- Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research & Evaluation, 7*(10), 1-10.
- Moss, P. A., & Koziol, S. M. (1991). Investigating the validity of a locally developed critical thinking test. *Educational Measurement: Issues and Practice, 10*(3), 17-22.
- Mulnix, J. W., & Mulnix, M. J. (2010). Using a writing portfolio project to teach critical thinking skills. *Teaching Philosophy, 33*(1), 27-54.
- Munzenmaier, C., & Rubin, N. (2013). Bloom's Taxonomy: What's Old Is New Again. *The eLearning Guide Research*. Retrieved from <http://onlineteachered.mit.edu/edc-pakistan/files/best-practices/session-2/Pre-Session-Munzenmaier-Rubin-2013.pdf>
- Nair, G. G. (2011). *Preliminary psychometric characteristics of the critical thinking self-assessment scale* (Unpublished doctoral dissertation). University of Saskatchewan.
- Neely, M. E. (2009). *Students thinking, students writing: Exploring undergraduates' epistemological beliefs and rhetorical writing* (Unpublished doctoral dissertation). The University of Texas at Austin.
- Neff-Lippman, J. (2012). Assessing writing. In I. L. Clark (Ed.), *Concepts in composition: Theory and practice in the teaching of writing* (2nd ed., pp. 145-167). New York: Routledge.

- New, D. K. (2005). *Observer bias influences in the Marschak Interaction Method rating system* (Unpublished doctoral dissertation). Alliant International University.
- Newell, G. E., Beach, R., Smith, J., VanDerHeide, J., Kuhn, D., & Andriessen, J. (2011). Teaching and learning argumentative reading and writing: A review of research. *Reading Research Quarterly, 46*(3), 273- 304.
- Niedringhaus, L. K. (2001). Using student writing assignments to assess critical thinking skills: A holistic approach. *Holistic Nursing Practice, 15*(3), 9-17.
- Norris, S. P., & Ennis, R. H. (1989). *Evaluating critical thinking*. Pacific Grove, CA: Midwest Publications.
- Nussbaum, E. M., & Schraw, G. (2007). Promoting argument–counterargument integration in students’ writing. *The Journal of Experimental Education, 76*(1), 59-92.
- Ockey, G. J. (2014). Exploratory factor analysis and structural equation modeling. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 1224–1244). NJ: John Wiley & Sons.
- Oda, M. (2008). Thinking critically about critical thinking in TESOL: East vs. West. *The Journal of Asia TEFL, 5*(1), 145-173.
- O’Donnell, M. (2015). Science writing, wikis, and collaborative learning. In J. Dougherty, & T. O’Donnell (Eds.), *Web writing: Why and how for liberal arts teaching and learning* (pp. 47-54). Ann Arbor: University of Michigan Press.
- Ongiem, A., & Vichitvejpaisal, P. (2018). Validation of the tests. *Thai J Anesthesiol, 44*(1), 36-42.
- O’Neill, P. (2003). Moving beyond holistic scoring through validity inquiry. *Journal of Writing Assessment, 1*, 47-65.
- Ornstein, A. C., & Lasley, T. J. II. *Strategies for effective teaching* (3rd ed.). Boston: McGraw-Hill Companies, Inc.

- Osana, P. H., & Seymour, J. S. (2004). Critical thinking in preservice teachers: A rubric for evaluating argumentation and statistical reasoning. *Educational Research and Evaluation, 10*(4-6), 473-498.
- Oshima, A., & Hogue, A. (1999). *Writing academic English* (3rd ed.). NY: Longman.
- Oshima, A., & Hogue, A. (2006). *Writing academic English* (4th ed.). NY: Pearson Education, Inc.
- Pan, Y. (2008). A genre-based approach for teaching writing to low-proficiency EFL college students. *PASAA, 42*, 79-95.
- Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Thousand Oaks, California: Sage Publications.
- Paul, R. (1990). *Critical thinking: What every person needs to survive in a rapidly changing world*. Rohnert Park, CA: Center for Critical Thinking and Moral Critique, Sonoma State University.
- Paul, R., & Elder, L. (2001). *Critical thinking: Tools for taking charge of your learning and your life*. Upper Saddle River, NJ: Prentice Hall.
- Paul, R., & Elder, L. (2006). *The miniature guide to critical thinking: Concepts and tools*. Dillon Beach, CA: Foundation for Critical Thinking.
- Paul, R., & Elder, L. (2008). *The thinker's guide to the nature and functions of critical & creative thinking*. Dillon Beach, CA: Foundation for Critical Thinking.
- Pawlowski, D. R. (1997). *Challenging students to think: Making critical thinking and writing to the "basic" course*. Paper presented at the annual meeting of the National communication association, Chicago, U.S.A.
- Payutto, P. A. (2010). *The language and the truth*. Bangkok: Chandpen Publishing.
- Payutto, P. A. (2015). *The Pali cannon: What a Buddhist must know*. Bangkok: Chandpen Publishing.

- Payutto, P. A. (2017). *A constitution for living: Buddhist principles for a fruitful and harmonious life*. Bangkok: Phlidhamm Publishing.
- Peach, B. E., Mukherjee, A., & Hornyak, M. (2007). Assessing critical thinking: A college's journey and lessons learned. *Journal of Education for Business*, 82(6), 313-320.
- Pei, Z., Zheng, C., Zhang, M., & Liu, F. (2017). Critical thinking and argumentative writing: Inspecting the association among EFL learners in China. *English Language Teaching*, 10(10), 31-42.
- Peirce, K. P. (2006). *Alternatives to argumentation: Implications for intercultural rhetoric* (Unpublished doctoral dissertation). The University of Arizona.
- Penner, I. S. (2010). *Comparison of effects of cognitive level and quality writing assessment (CLAQWA) rubric on freshman college student writing*. (Doctoral Dissertation). Liberty University.
- Plakans, L., & Gebril, A. (2017). An assessment perspective on argumentation in writing. *Journal of Second Language Writing*, 36, 85-86.
- Ploysangwal, W. (2018). An assessment of critical thinking skills of Thai undergraduate students in private Thai universities in Bangkok through an analytical and critical reading test. *Journal Humanities and Social Sciences*, 38(3), 75-91.
- Polston, K. (2014). *Self- and peer-assessment of product creativity with a rubric in a collaborative environment: A research study with undergraduate textile and apparel designers*. (Unpublished doctoral dissertation). North California State University.
- Pookcharoen, S. (2009). *Metacognitive online reading strategies among Thai EFL university students*. (Unpublished doctoral dissertation). Indiana University.
- Popham, W. J. (1997). What's wrong—and what's right—with rubrics. *Educational Leadership*, 55(2), 72-75.

- Preiss, D. D., Castillo, J. C., Flotts, P., & Martín, E. S. (2013). Assessment of argumentative writing and critical thinking in higher education: Educational correlates and gender differences. *Learning and Individual Difference, 28*, 193-203.
- Prommak, S. (2016). Thai EFL university teachers' conceptions of critical thinking. *TESOL Journal, 29*(1), 49-72.
- Pu, S., & Evans, M. (2019). Critical thinking in the context of Chinese postgraduate students' thesis writing: A positioning theory perspective. *Language, Culture and Curriculum, 32*(1), 50-62.
- Purpura, J. E. (2004). *Assessing grammar*. Cambridge: Cambridge University Press.
- Purpura, J. E. (2014). Assessing grammar. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp.100-124). Oxford: John Wiley & Sons, Inc.
- Puthikanon, N. (2009). *Examining critical thinking and language use through the use of WebQuests in an EFL reading class* (Unpublished doctoral dissertation). Indiana University.
- Qin, J. (2009). *The analysis of Toulmin elements and use of sources in Chinese university EFL argumentative writing* (Unpublished doctoral dissertation). Northern Arizona University.
- Qoura, A. A., & Zahran, F. A. (2018). The effect of the 6+1 trait writing model on ESP university students critical thinking and writing achievement. *English Language Teaching, 11*(9), 68-79.
- Quintieri, M. (2005). *Measuring rater consistency: An investigation into the effects of two testing instruments on raters' scores*. (Master's thesis). Concordia University.
- Raimes, A. (1983). *Techniques in teaching writing*. New York, NY: Oxford University Press.
- Rainbolt, G. W., & Dwyer, S. L. (2012). *Critical thinking: The art of argument*. Boston, MA: Wadsworth.

- Rakedzon, T., & Baram-Tsabari, A. (2017). To make a long story short: A rubric for assessing graduate students' academic and popular science writing skills. *Assessing Writing*, 32,28-42.
- Ramanathan, V., & Atkinson, D. (1999). Individualism, academic writing, and ESL writers. *Journal of Second Language Writing*, 8(1), 45-75.
- Ramanathan, V., & Kaplan, R. B. (1996a). Audience and voice in current L1 composition texts: Some implications for ESL student writers. *Journal of Second Language Writing*, 5(1), 21-34.
- Ramanathan, V., & Kaplan, R. B. (1996b). Some problematic "Channels" in the teaching of critical thinking in current L1 composition textbooks: Implications for L2 student-writers. *Issues in applied linguistics*, 7(2), 225-249.
- Rancourt, A. M. (2012). Assessing academic/intellectual skills in Keense State college's integrative studies program. *Journal of Assessment and Instructional Effectiveness*, 2(2), 211-262.
- Reddy, M. Y. (2011). Design and development of rubrics to improve assessment outcomes: A pilot study in a Master's level business program in India. *Quality Assurance in Education*, 19(1), 84-104.
- Reddy, M. Y., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435-448.
- Rezaei, A. R., & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing*, 15, 18-39.
- Reznitskaya, A., Kuo, L., Glina, M., & Anderson, R. C. (2009). Measuring argumentative reasoning: What's behind the numbers? *Learning and Individual Differences*, 19(2), 219-224.

- Riesser, S. T. (1994). *Examination of reliability and validity of the Performance Assessment of Science Skills (PASS) instruments, alternative assessment instruments of science process skills* (Unpublished doctoral dissertation). Ball State University
- Rogers, E. P. (2005). *Scale construction and halo effect in secondary student ratings of teacher performance*. (Unpublished doctoral dissertation). Brigham Young University
- Rosenwasser, D., & Stephen, J. (2012). *Writing analytically* (6th ed.). Boston, MA: Wadsworth.
- Ross, J. A. (2006). The reliability, validity, and utilization of self-assessment. *Practical Assessment, Research & Evaluation*, 11(10), 1-13.
- Rottenberg, A. T. (1994). *The structure of argument*. New York, NY: St. Martin's Press.
- Rowicki, M. A. (2001). *A study of the relationship between reflective writing and critical thinking in seventh grade integrated science students* (Unpublished doctoral dissertation). Auburn University.
- Ruggiero, V. R. (1984). *The art of thinking: A guide to critical and creative thought*. New York, NY: Harper & Row, Publishers, Inc.
- Saxton, E., Belanger, S., & Becker, W. (2012). The critical thinking analytic rubric (CTAR): Investigating intra-rater and inter-rater reliability of a scoring mechanism for critical thinking performance assessments. *Assessing Writing*, 17, 251-270.
- Schamber, J. F., & Mahoney, S. L. (2006). Assessing and improving the quality of group critical thinking exhibited in the final projects of collaborative learning groups. *The Journal of General Education*, 55(2), 103-137.
- Scheuer, O., McLaren, B., Weinberger, A., & Niebuhr, S. (2014). Promoting critical, elaborative discussions through a collaboration script and argument diagrams. *Instructional Science*, 42(2), 127-157.

- Schley, S., & Albertini, J. (2015). Assessing the writing of deaf college students: Reevaluating a direct assessment of writing. *Journal of Deaf Studies and Deaf Education, 10*(1), 96-105.
- Schmidt, R. C. (1997). Managing Delphi surveys using nonparametric statistical techniques. *Decision Sciences, 28*(3), 763-774.
- Scriven, M., & Paul, R. (1987). *Defining critical Thinking*. Paper presented at the 8th Annual International Conference on Critical Thinking and Education Reform, Summer 1987. Retrieved from <https://www.criticalthinking.org/pages/defining-critical-thinking/766>
- Shavelson, R. J. (2010). *Measuring college learning responsibly: Accountability in a new era*. Stanford: Stanford University Press.
- Shehab, H. M., & Nussbaum, E. M. (2015). Cognitive load of critical thinking strategies. *Learning and Instruction, 35*, 51-61.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4-14.
- Silva, E. (2008). *Measuring skills for the 21st century*. Washington, DC: Education sector.
- Simpkins, K. (2016). *Inter-rater reliability among principals using the instructional domain of the Tennessee educator acceleration model*. (Doctoral Dissertation). Tennessee State University.
- Sonoma State University. (1996). *ICAT Critical Thinking Essay Test*. Rohnert Park, CA: Sonoma State University, Center for Critical Thinking and Moral Critique.
- Sondel, H. B. (2009). *The effects of curricular programs on aspects of critical thinking as applied to writing* (Unpublished doctoral dissertation). University of Virginia.
- Stapleton, P. (2001a). Assessing critical thinking in the writing of Japanese university students: Insights about assumptions and content familiarity. *Written Communication, 18*(4), 506-548.

- Stapleton, P. (2001b). *Assessing critical thinking in the writing of Japanese university students: Insights about assumptions, content familiarity and biology* (Unpublished doctoral dissertation). University of Leicester.
- Stapleton, P. (2002). Critical thinking in Japanese L2 writing: Rethinking tired constructs. *ELT Journal*, 56(3), 250-257.
- Stapleton, P., & Wu, Y. (2015). Assessing the quality of arguments in students' persuasive writing: A case study analyzing the relationship between surface structure and substance. *Journal of English for Academic Purposes*, 7, 12–23.
- Stein, B., & Haynes, A. (2011). Engaging faculty in the assessment and improvement of students' critical thinking using the CAT. *Change: The Magazine of Higher Learning*, 43(2), 44-49.
- Stellmack, M. A., Konheim-Kalkstein, Y. L., Manor, J. E., Massey, A. R., & Schmitz, J. A. P. (2009). An assessment of reliability and validity of a rubric for grading APA style introductions. *Teaching of Psychology*, 36(2), 102-107.
- Sternberg, J. R. (1985). *Critical thinking: Its nature, measurement, and improvement*. Washington, DC: National Institute of Education.
- Stevens, D. D., & Levi, A. (2005). *Introduction to rubric: An assessment tool to save grading time, convey effective feedback, and promote student learning* (1st ed.). Virginia: Stylus Publishing, LLC.
- 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.
- Stobart, G. (2012). Validity in formative assessment. In J. Gardner (Ed.), *Assessment and learning: An introduction* (pp. 233-242). London: Sage.

- Suda-jitapa, M. (2008). *Systematizing EAP materials development: Design, evaluation and revision in a Thai undergraduate reading course* (Unpublished doctoral dissertation). University of Edinburgh.
- Tangkiengsirisin, S. (2014). Links between teacher written feedback and student revision. *LEARN Journal, Special Issue*, 50-60.
- Tapinta, P. (2006). *Exploring Thai EFL university students' awareness of their knowledge, use, and control of strategies in reading and writing* (Unpublished doctoral dissertation). University of North Texas.
- Tashakkori, A., & Creswell, J. W. (2007). The new era of mixed methods [Editorial]. *Journal of Mixed Methods Research*, 1(1), 3-7.
- Taube, K. T. (1997). Critical thinking ability and disposition as factors of performance on a written critical thinking test. *Journal of General Education*, 46(2), 129-164.
- Thomas, I. (2009). Critical thinking, transformative learning, sustainable education, and problem-based learning in universities. *Journal of Transformative Education*, 7(3), 245-264.
- Thongrin, S. (2002). *E-mail peer response in collectivist Thai culture: Task, social, and cultural dimensions* (Unpublished doctoral dissertation). Indiana University of Pennsylvania.
- Thongrin, S. (2009). Collaboration for liberation: Using TIMET to enhance EFL students' writing skills. *Journal of English Studies*, 4, 31-93.
- Tittle, P. (2011). *Critical thinking: An appeal to reason*. New York, NY: Routledge.
- Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- Toulmin, S., Rieke, R., & Janik, A. (1979). *An introduction to reasoning*. New York: Macmillan Publishing Co., Inc.

- Tulgar, A. T., Yagiz, O., & Han, T. (2017). Bridging the Gap between Linguistic and Pragmatic Assessment in Turkish Contexts. *PASAA*, 54, 58-81.
- University of Cambridge. (1996). *Critical thinking*. Cambridge, UK: University of Cambridge, Local Examinations Syndicate.
- US. Department of Education, National Center for Education Statistics. (2000). *The NPEC sourcebook on assessment, Volume 1: definitions and assessment methods for critical thinking, problem solving, and writing*. NCES 2000--172 prepared by T. Dary Erwin for the Council of the National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development. Washington, DC: U.S. Government Printing Office.
- Uysal, H. H. (2006). *Tracing culture behind writing: Rhetorical patterns and bidirectional transfer in L1 and L2 essays of Turkish writers in relation to educational context* (Unpublished doctoral dissertation). The University of Iowa.
- Vallis, G. L. (2010). *Reason to write: Applying critical thinking to academic writing*. North Carolina: Kona Publishing and Media Group.
- VanderMey, R., Meyer, V., Rys, J. V., & Sebranek, P. (2012). *The college writer: A guide to thinking, writing, and researching* (4th ed.). Boston, MA: Wadsworth, Cengage Learning.
- Verlinden, J. (2005). *Critical thinking and everyday argument*. Boston, MA: Wadsworth, Cengage Learning.
- Wade, C. (1995). Using writing to develop and assess critical thinking. *Teaching of Psychology*, 22(1), 24-28.
- Walton, D. (2013). *Methods of argumentation*. New York, NY: Cambridge University Press.
- Wang, A. Y. (2012). Exploring the relationship of creative thinking to reading and writing. *Thinking Skills and Creativity*, 7, 38-47.

- Wang, X., Hwang, G.-J., Liang, Z.-Y., & Wang, H.-Y. (2017). Enhancing students' computer programming performances, critical thinking awareness and attitudes towards programming: An online peer-assessment attempt. *Educational Technology & Society*, 20 (4), 58–68.
- Watson, G., & Glaser, E. M. (1980). *Watson–Glaser Critical Thinking Appraisal, forms A and B manual*. San Antonio, TX: The Psychological Corporation.
- Watson, G., & Glaser, E. M. (2008a). *Watson–Glaser Critical Thinking Appraisal, forms A and B manual*. Upper Saddle River, NJ: Pearson Education.
- Watson, G., & Glaser, E. M. (2008b). *Watson–Glaser Critical Thinking Appraisal short form manual*. Pearson Education: Upper Saddle River, NJ.
- Watson, G., & Glaser, E. M. (2010). *Watson–Glaser II Critical Thinking Appraisal: Technical manual and user's guide*. San Antonio, TX: NCS Pearson.
- West, T. L. (1994). *The effects on argumentation instruction on critical thinking skills*. (Doctoral Dissertation). Southern Illinois University at Carbondale.
- Weigle, S. C. (1994). *Effects of training on raters of English as a second language compositions: Quantitative and qualitative approaches*. (Unpublished doctoral dissertation). University of California.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Weigle, S. C. (2007). Teaching writing teachers about assessment. *Journal of Second Language Writing*, 16, 194-209.
- Weigle, S. C., & Friginal, E. (2015). Linguistic dimensions of impromptu test essays compared with successful student disciplinary writing: Effects of language background, topic, and L2 proficiency. *Journal of English for Academic Purposes*, 18, 25-39.

- Weiwei, Y., Lu, X., & Weigle, S. C. (2015). Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing, 28*, 53-67.
- Welch, K. C., Hieb, J., & Graham, J. (2015). A systematic approach to teaching critical thinking skills to electrical and computer engineering undergraduates. *American Journal of Engineering Education, 6*(2), 113-123.
- Whittenburg, C. (1995). *The development and validation of a model for disciplined-thinking in a post-anesthesia care unit* (Unpublished doctoral dissertation). Texas Tech University.
- Widodo, H. P. (2012). Pedagogical tasks for sharing EFL college writers' critical thinking and self voicing. *International Journal of Innovation in English Language Teaching, 1*(1), 87-99.
- Williams, J. M., & Colomb, G. G. (2007). *The craft of argument* (3rd ed.). New York: Pearson Education, Inc.
- Williams, R. L. (1999). Operational definitions and assessment of higher-order cognitive constructs. *Educational Psychology Review, 11*(4), 411-427.
- Wilson, E., & Post, J. (2019). Toward critical engagement: Affect and action in student interactions with instructor feedback. In A. R. Gere (Ed.), *Developing writers: A longitudinal study* (pp.29-55). Ann Arbor: University of Michigan Press.
- Wilson, K. (2016). Critical reading, critical thinking: Delicate scaffolding in English for academic purposes (EAP). *Thinking Skills and Creativity, 22*, 256-265.
- Winkler, A. C., & McCuen, J. R. (1997). *Writing talk: Paragraphs and short essays with readings*. New Jersey: Prentice-Hall, Inc.
- Wright, K. E. L. (2016). *Methods and measurement for using writing to transform knowledge in science classes* (Unpublished doctoral dissertation). Texas A&M University.

- Wu, S., & Rubin, D. L. (2000). Evaluating the impact of collectivism and individualism on argumentative writing by Chinese and North American college students. *Research in the Teaching of English, 35*(2), 148-178.
- Wyngaard, S., & Gehrke, R. (1996). Responding to audience: Using rubrics to teach and assess writing. *The English Journal, 85*(6), 67-70.
- Xu, M., Huang, C., & You, X. (2016). Reasoning patterns of undergraduate thesis in translation studies: An intercultural rhetoric studies. *English for Specific Purposes, 41*, 68-81.
- Yang, W. (2014). *Mapping the relationships among the cognitive complexity of independent writing tasks, L2 writing quality, and complexity, accuracy and fluency of l2 writing* (Unpublished doctoral dissertation). Georgia State University.
- Yeh, S. S. (1998a). Empowering education: Teaching argumentative writing to cultural minority middle-school students. *Research in the Teaching of English, 33*(1), 49-83.
- Yeh, S.S. (1998b). Validation of a scheme for assessing argumentative writing of middle school students. *Assessing Writing, 5*(1), 123-150.
- Yeh, S.S. (2001). Tests worth teaching to: Constructing state-mandated tests that emphasize critical thinking. *Educational Research, 30*(9), 12-17.
- Young, L.W. L. (1994). *Crosstalk and culture in Sino-American communication*. New York: Cambridge University Press.
- Yu, H. J. (2008). *Interactional structure in the writing process: A comparison of three ESL writing classes* (Unpublished doctoral dissertation). Arizona State University.
- Yusri. (2018). The effects of problem solving, project-based learning, linguistic intelligence and critical thinking on the students' report writing. *Advances in Language and Literary Studies, 9*(6), 21-26.

- Zahner, D. (2013). *Reliability and validity of CLA+*. New York, NY: Council for Aid to Education. Retrieved from [http://cae.org/images/uploads/pdf/Reliability\\_and\\_Validity\\_of\\_CLA\\_Plus.pdf](http://cae.org/images/uploads/pdf/Reliability_and_Validity_of_CLA_Plus.pdf)
- Zhao, C. G. (2010). *The role of voice in high-stakes second language writing assessment* (Unpublished doctoral dissertation). New York University.
- Zhao, C. G. (2012). Measuring authorial voice strength in L2 argumentative writing: The development and validation of an analytic rubric. *Language Testing*, 30(2), 201-230.
- Zhao, C. G. (2017). Voice in timed L2 argumentative essay writing. *Assessing Writing*, 31, 73–83.
- Zimmerbaum, K. Z. (2014). *Effects of critical questioning on argument* (Master's thesis). Caldwell College.



**Appendices**

## APPENDIX 1

### A table summarizing the previous multiple choice critical thinking tests

Critical thinking Tests	Developer	Test format	Length	Test items	Detailed information about the tests
1. California Critical Thinking Skills Test (CCTST)	Insight Assessment (California Academic Press)	Online or pencil/paper	45 minutes	34 items	The CCTST evaluates a test taker's skills of:  (a) analysis,  (b) evaluation,  (c) inference,  (d) deduction,  (e) induction and  (f) overall reasoning (Facione, 1990b)
2. Collegiate Assessment of Academic Proficiency (CAAP) Critical Thinking	ACT	Paper/ pencil	40 minutes	32 items	The CAAP Critical Thinking measures a test taker's skills in 1) analyzing elements of an argument, and 2) evaluating an argument (CAAP Program Management, 2012)
3. Cornell Critical Thinking Test (CCTT)	The Critical Thinking Co.	Computer based (using the software)  or paper/pencil	50 minutes	Level X: 71 items	Level X is developed for a test taker in Grades 5 to 12. It measures the skills of:  (a) induction,  (b) deduction,  (c) credibility, and  (d) identification of assumptions  (Ennis, Millman, & Tomko, 1985)

Critical thinking Tests	Developer	Test format	Length	Test items	Details information about the tests
4. ETS Proficiency Profile (EPP) Critical Thinking	ETS	Online and paper/pencil	40 minutes	27 items	<p>The EPP measures a test taker's ability to:</p> <ul style="list-style-type: none"> <li>(1) distinguish between rhetoric and argumentation in a piece of nonfiction prose;</li> <li>(2) recognize assumptions and hypothesis;</li> <li>(3) infer and interpret a relationship between variables; and</li> <li>(4) draw valid conclusions based on information presented (ETS, 2010)</li> </ul>
5. Watson-Glaser Critical Thinking Appraisal tool (WGCTA)	Pearson	Online and paper/pencil	40-60 minutes (Forms A and B)	80 items	<p>The WGCTA consists of five tests:</p> <ul style="list-style-type: none"> <li>(a) inference,</li> <li>(b) recognition of assumptions,</li> <li>(c) deduction,</li> <li>(d) interpretation, and</li> <li>(e) evaluation of arguments.</li> </ul> <p>(Watson &amp; Glaser, 1980, 2008a, 2008b, 2010)</p>

## APPENDIX 2

### A table summarizing the domains of assessment of the previous holistic critical thinking rubric

The holistic critical thinking rubrics of the previous studies	Element of critical thinking assessment in argumentative writing						
	Position	Reasons and evidence	Counter-argument and refutation	Organization	Word choices and sentence structures	A summary and a final constructive comment	Others
Collegiate Learning Assessment (CLA) (Shavelson, 2010)	✓	✓	✓	✓	×	×	-
The GRE <sup>®</sup> Analytical Writing rubric (ETS, 2019)	✓	✓	✓	✓	×	×	Grammar
Scholastic Aptitude Test (SAT) by College Board (2017)	✓	✓	×	✓	✓	✓	Grammar
Angeli and Valanides (2009)	✓	✓	✓	×	×	×	-
HCTSR (Facinon & Facinon, 1996)	✓	✓	✓	×	×	✓	-
Welch, Hieb and Graham (2015)	✓	✓	×	×	×	✓	-
Portland State University Studies program rubric (see Stevens & Levi, 2005)	✓	✓	✓	×	×	×	Make an ethical judgment
Bauer (2016)	✓	✓	✓	✓	✓	×	Grammar
Blattner & Frazier (2002)	✓	✓	✓	✓	×	×	Uses and/or cites source material

**Note:** × = absent  
✓ = present

### APPENDIX 3

#### A table summarizing the previous analytic critical thinking rubrics which contain unclear evaluative descriptors

Analytic CT rubrics of the previous researchers	CT domains of assessment	Scoring	Evaluative descriptors that cannot make a clear distinction between different levels of critical thinking performance of the students
Carbone (2009)	<ul style="list-style-type: none"> <li>▪ Write a thesis statement</li> <li>▪ Provide the first reason</li> <li>▪ Provide additional supporting reasons</li> <li>▪ State opposing viewpoint(s)</li> <li>▪ Use appropriate language</li> <li>▪ Use correct grammar</li> <li>▪ Demonstrate a rhetorical square in writing (i.e., an audience, a purpose, a text and an author, all of which work together to make a text convincing)</li> </ul>	<ul style="list-style-type: none"> <li>▪ 5 points for each domain of the assessment</li> </ul>	<ul style="list-style-type: none"> <li>▪ No levels of performance</li> <li>▪ No evaluative descriptors for each performance level</li> </ul>
Daud (2012)	<ul style="list-style-type: none"> <li>▪ Identify a clear stance</li> <li>▪ Explain difficult terms and ideas</li> <li>▪ Break down the issues into parts</li> <li>▪ Support an argument</li> <li>▪ Use reliable source or literature</li> <li>▪ Organize ideas coherently</li> <li>▪ Integrate other people's opinions</li> <li>▪ Use an appropriate academic writing style</li> <li>▪ Use correct English grammar</li> </ul>	6 rating scales  1-2 points= emerging;  3-4 points= developing;  5-6 points= mastering	No evaluative descriptors in each rating scale
McLaughlin and Moore (2012)	<ul style="list-style-type: none"> <li>▪ Focus (organization)</li> <li>▪ Logic (evidence, conclusion, and alternative viewpoints)</li> <li>▪ Content (details and reason)</li> <li>▪ Style (word choices and sentence structures)</li> <li>▪ Correctness (grammar)</li> <li>▪ Research (outside source information and citation)</li> </ul>	<ul style="list-style-type: none"> <li>▪ 4 rating scales (i.e., <i>superior</i>, <i>skilled</i>, <i>adequate</i>, <i>inadequate</i>)</li> <li>▪ No numerical scores</li> </ul>	No evaluative descriptors in each scale level

Analytic CT rubrics of the previous researchers	CT domains of assessment	Scoring	Evaluative descriptors that cannot make a clear distinction between different levels of critical thinking performance of the students
Billups (2006)	<ul style="list-style-type: none"> <li>▪ Ability to write ideas clearly.</li> <li>▪ Ability to develop a writing style</li> <li>▪ Ability to use grammar correctly</li> <li>▪ Ability to transmit information or knowledge</li> <li>▪ Ability to understand what they read</li> <li>▪ Ability to use general concepts in content areas</li> <li>▪ Ability to use appropriate vocabulary</li> </ul>	4 rating scales:  4=Excellent  3=Good  2=Fair  1=Poor	No evaluative descriptors for each level of the performance
Dong (2015)	Nine intellectual standards from Paul and Elder's Critical Thinking theory were used as CT domains of assessment: <ul style="list-style-type: none"> <li>▪ Clarity (use clear word choices and sentence structures)</li> <li>▪ Accuracy (use correct information)</li> <li>▪ Precision (use sufficient, specific examples and explanations)</li> <li>▪ Relevance (use relevant information)</li> <li>▪ Depth (demonstrate a thoroughness in thinking)</li> <li>▪ Breath (state an opposing view)</li> <li>▪ Logic (give reasons and evidence)</li> <li>▪ Significance (demonstrate a substantiality in meaning)</li> <li>▪ Fairness (have an ethical appropriateness of views, reasons, evidence and conclusion) (Dong, 2015)</li> </ul>	<ul style="list-style-type: none"> <li>▪ 5 rating scales (<i>Very good, Good, Fair, Poor</i> and <i>Very Poor</i>)</li> <li>▪ No numerical scores assigned for each rating scale</li> </ul>	Three domains; namely, Depth, Significance and Fairness, contain unclear descriptors. <p><b>Domain of <u>Depth</u>:</b></p> <p><u>Level Very Good</u> "Implying thoroughness in thinking; presenting full understanding of the complexities"</p> <p><u>Level Good</u> "Implying depth in thinking; presenting an understanding of the complexities"</p> <p><u>Level Fair</u> "Not implying enough depth in thinking; presenting a basic understanding of the complexities"</p> <p><u>Level Poor</u> "Not implying depth in thinking; not presenting an understanding of the complexities"</p> <p><u>Level Very Poor</u> "Not implying any depth in thinking; not presenting any basic understanding of the complexities" (Dong, 2015, p.116)</p>

Analytic CT rubrics of the previous researchers	CT domains of assessment	Scoring	Evaluative descriptors that cannot make a clear distinction between different levels of critical thinking performance of the students
Saxton and his colleagues (2012)	<p>Six CT constructs from Facione's (1990) theory are used as the domains of the assessment:</p> <ul style="list-style-type: none"> <li>▪ Interpretation (identify all major viewpoints, interpret evidence and use the key concepts and terminology correctly)</li> <li>▪ Analysis (analyze and present an evaluation of similarities and differences of the views, identify claims and argument or assumption in argument and organize the information)</li> <li>▪ Evaluation (identify the arguments from multiple perspectives and analyze and evaluate all major alternative views)</li> <li>▪ Inference (apply or extend key concepts to make predictions, draw inferences and analyze implications)</li> <li>▪ Explanation (integrates key sources to support a conclusion and provide reasons with evidence)</li> <li>▪ Disposition (consider a context, show consistent views, give reasons and evidence appropriately and be aware of alternative views)</li> </ul>	<ul style="list-style-type: none"> <li>▪ 6 Rating scales (1 - 6 points)</li> </ul>	<p>Domain of Inference</p> <p><u>Score Level 6</u>  “Demonstrates <b>confident</b> ability to apply or extend key concepts”  “Demonstrates <b>surprising/insightful</b> ability to take concepts”</p> <p><u>Score Level 5</u>  “Demonstrates a <b>clear</b> ability to apply or extend key concepts”  “Demonstrates <b>strong</b> ability to take concepts”</p> <p><u>Score Level 3</u>  “Demonstrates a <b>shaky</b> ability to apply or extend key concepts”  “Demonstrates an <b>uneven</b> ability to take concepts”</p> <p><u>Score Level 2</u>  “Demonstrates <b>inadequate</b> ability to apply or extend key concepts”  “Demonstrates a <b>superficial</b> ability to take concepts”</p> <p>(Saxton et al., 2016, p.268, emphasis added)</p>

Analytic CT rubrics of the previous researchers	CT domains of assessment	Scoring	Evaluative descriptors that cannot make a clear distinction between different levels of critical thinking performance of the students
American College Test (ACT) (College Board, 2016)	<ul style="list-style-type: none"> <li>▪ Ideas and analysis (ability to state a claim and generate an argument)</li> <li>▪ Development and support (ability to give reasons and evidence to support a claim and address an opposing view)</li> <li>▪ Organization (ability to organize ideas in a logical way and use transitional words to link ideas)</li> <li>▪ Language use (ability to use skillful word choices, varied sentence structures and correct grammar)</li> </ul>	<p>6 rating scales:</p> <p>6 points = effective skill</p> <p>5 points = well-developed skill</p> <p>4 points = adequate skill</p> <p>3 points = some developing skill</p> <p>2 points = weak or inconsistent skill</p> <p>1 point = little or no skill</p>	<p>Domain of Ideas and analysis:</p> <p><u>Score Level 6</u></p> <p>The response exhibits a <b>skillful</b> organizational strategy. The response is unified by a controlling idea or purpose, and a <b>logical progression of ideas increases the effectiveness of the writer’s argument</b>. Transitions between and within paragraphs <b>strengthen</b> the relationships among ideas.</p> <p><u>Score Level 5</u></p> <p>The response exhibits a <b>productive</b> organizational strategy. The response is mostly unified by a controlling idea or purpose, and a <b>logical consequencing of ideas contributes to the effectiveness of the argument</b>. Transitions between and within paragraphs <b>consistently clarify</b> the relationships among ideas.</p> <p>(College Board, 2016, emphasis) added)</p>

Analytic CT rubrics of the previous researchers	CT domains of assessment	Scoring	Evaluative descriptors that cannot make a clear distinction between different levels of critical thinking performance of the students
Osana and Seymour (2004)	<ul style="list-style-type: none"> <li>▪ Ability to distinguish between evidence quality and evidence type</li> <li>▪ Ability to take evidence other than opinion into account before making a judgment about an issue</li> <li>▪ Ability to use counter-positive evidence to argue a position</li> <li>▪ Ability to consider different perspectives of an argument</li> <li>▪ Ability to describe the types of evidence to support a position</li> <li>▪ Ability to recognize that research findings are more credible as evidence than personal opinion</li> </ul>	No numerical scores assigned for each rating scale	The rubric illustrates responses from test takers as examples of answers for each domain of the assessment. However, there are no rating scales and no evaluative descriptors in each scale level.

Analytic CT rubrics of the previous researchers	CT domains of assessment	Scoring	Evaluative descriptors that cannot make a clear distinction between different levels of critical thinking performance of the students
<p>The Cognitive Level and Quality Writing (CLAQWA) (See Penner, 2010)</p>	<ul style="list-style-type: none"> <li>▪ <i>Assignment Parameters</i> (assignment requirement, main idea, audience and purpose)</li> <li>▪ <i>Organization and Development</i> (opening sentences, coherence devices, paragraph construction and closing sentences)</li> <li>▪ <i>Organization and Development</i> (reasoning, quality of details, and quantity of details)</li> <li>▪ <i>Language</i> (word choices, sentence comprehensibility, sentence construction and point of view)</li> <li>▪ <i>Observation of Standard Edited English</i> (English grammar and mechanics)</li> </ul>	<p>1-5 rating scales</p>	<p>Many domains of assessment (e.g., audience, coherence devices, and reasoning) contain unclear descriptors.</p> <p><b>Domain of Audience:</b>  <u>Score Level 1</u>          “The writer shifts between multiple and/or inappropriate audiences because of a lack of reader awareness”  <u>Score Level 2</u>          “The writer shows a lack of reader awareness by addressing one or more inappropriate audiences”</p> <p><b>Domain of Coherence devices:</b>  <u>Score level 1</u>          “Coherence devices are absent or inappropriate”  <u>Score level 2</u>          “Coherence devices are attempted but are ineffective”</p> <p><b>Domain of Reasoning:</b>  <u>Score level 1</u>          “The ideas are illogical and appear to reflect the writer’s stream of consciousness”  <u>Score level 2</u>          “The attempt at a progression of ideas is unsuccessful due to errors in logic, such as absolutes or contradictions”</p>

Analytic CT rubrics of the previous researchers	CT domains of assessment	Scoring	Evaluative descriptors that cannot make a clear distinction between different levels of critical thinking performance of the students
Illinois Critical Thinking Essay Test (Finken & Ennis, 1993)	<ul style="list-style-type: none"> <li>▪ <i>Focus</i> (the degree to which main ideas are clear and maintained)</li> <li>▪ <i>Supporting reasons</i> (the degree to which supporting reasons and evidence are clear and believable and are from credible sources)</li> <li>▪ <i>Reasoning</i> (the degree to which (1) a conclusion supported by reasons/evidence, (2) an alternative view is addressed, and (3) an argument clear)</li> <li>▪ <i>Organization</i> (the degree to which ideas flow smoothly and are clear and connected)</li> <li>▪ <i>Convention</i> (the use of English grammar correctly)</li> <li>▪ <i>Integration (a holistic picture)</i> (how well the features from the five previous domains of the assessment are evident in the text)</li> </ul>	<p>1-6 levels of performance</p> <p>Score Level 1=<i>absent</i>,</p> <p>Score Levels 2-3=<i>developing</i>,</p> <p>Score Levels 4-6=<i>well-developed</i></p>	<p>Many domains of assessment contain unclear descriptors.</p> <p>Domain of <b>Focus</b>:</p> <p><u>Score Level 1</u> “Unclear; absent; insufficient length to ascertain maintenance”</p> <p><u>Score Level 2</u> “Confusing; attempted; main point unclear or shifts”</p> <p><u>Score Level 4</u> “Bare bones; position clear; main point previewed”</p> <p><u>Score Level 5</u> “Position clear; generally previewed”</p> <p>Domain of <b>Organization</b>:</p> <p><u>Score Level 2</u> “Attempted; plan is noticeable”</p> <p><u>Score Level 3</u> “Not knowledgeable in paragraphing”</p> <p>(Fiken &amp; Ennis, 1993, p.17)</p>

Analytic CT rubrics of the previous researchers	CT domains of assessment	Scoring	Evaluative descriptors that cannot make a clear distinction between different levels of critical thinking performance of the students
Neely (2009)	<ul style="list-style-type: none"> <li>▪ Problem statement</li> <li>▪ Audience awareness</li> <li>▪ Avoiding generalization</li> <li>▪ Logic</li> <li>▪ Counter-argument</li> <li>▪ Rebuttal</li> <li>▪ Source integration</li> <li>▪ Organization</li> <li>▪ Language use</li> <li>▪ Quality of source</li> <li>▪ Number of source</li> </ul>	2 levels of performance (high vs. low)	Neely's rubric has only two rating scales (high vs. low), so the rubric is very rough and cannot differentiate various levels of critical writing performance of the students.
MamourChoul (2010)	<ul style="list-style-type: none"> <li>▪ Rational appeal</li> <li>▪ Credibility appeal</li> <li>▪ Affective appeal</li> </ul>	Four rating scales for each domain of the assessment	<p>Domain of the credibility appeal</p> <p><u>Score Level 1</u>          "No writer credibility but <b>some awareness of audience's values</b> or some writer credibility (other than general knowledge) but <b>no awareness of audience's values</b>"</p> <p><u>Score Level 2</u>          "Some writer credibility (other than general knowledge) and <b>some awareness of audience's values</b>"          (MamourChoul, 2010, emphasis added, p.118)</p> <p>This rubric does not define the descriptors, namely "awareness of audience's values." Also, the rubric does not describe how to evaluate this specific descriptor in students' argumentative writing. The lack of the clear, detailed descriptors, therefore, affects the validity of the rubric.</p>

## APPENDIX 4

**A table summarizing the scoring criteria of the previous analytic critical thinking rubrics**

Analytic CT rubric of the previous studies	Crucial, relevant aspects of critical thinking in argumentative writing in the previous rubrics							
	Opening sentence(s) in the introduction	Claim	Reasons and Evidence	Counter-argument	Refutation	Summary and a final, constructive comment	Organization of ideas	Word choices and sentence structures
Hodges and his colleagues (2019)	✓	✓	✓	×	×	✓	✓	✓
Nejmaoui (2019)	×	✓	✓	✓	✓	×	✓	✓
Liu and Stapleton (2018)	×	×	×	×	×	×	✓	✓
Synthetic Skill Rubric (ETS, 2015)	×	✓	✓	×	×	✓	✓	×
Morozov (2011)	×	✓	✓	×	×	✓	×	✓
Pei, Zheng, Zhang and Liu (2017)	×	✓	✓	✓	✓	×	×	✓
Stapleton and Wu (2015)	×	✓	✓	✓	✓	×	×	×
Shehab and Nussbaum (2015)	×	✓	✓	✓	✓	×	×	×
Carwie (2009)	×	✓	✓	✓	✓	✓	×	×
Wright (2016)	×	✓	✓	×	×	×	✓	✓
Hughes (2000)	×	✓	✓	×	×	✓	×	×
Mulnix and Mulnix (2010)	✓	✓	✓	✓	✓	×	✓	✓
Rancourt (2012)	×	×	×	✓	✓	×	×	×
Mckitrick and Barnes (2012)	×	✓	✓	✓	✓	✓	×	×
Brookhart (2010)	×	✓	✓	×	×	×	×	×
Marzano and his colleagues (1993)	×	✓	✓	✓	✓	×	×	×
Haria (2010)	×	✓	✓	✓	✓	×	×	×
Franklin and his colleagues (2014)	×	✓	✓	✓	✓	✓	×	×
Metcalf (2007)	×	✓	✓	✓	✓	✓	×	×
Fulan (2014)	×	✓	✓	✓	✓	✓	×	✓
MamourChoul (2010)	✓	✓	✓	✓	✓	×	✓	✓
Chan, Ho and Ku (2011)	×	✓	✓	✓	✓	✓	×	×
Puthikanon (2009)	×	✓	✓	✓	×	×	×	×
McLendon (2008)	×	✓	✓	×	×	×	×	✓
Andrade (2000)	×	✓	✓	✓	×	×	✓	✓
VALUE Rubric (Association of American Colleges and Universities, 2019).	×	✓	✓	✓	✓	✓	×	×
Writing Center Partnerships, University of Houston (2009)	×	✓	✓	✓	×	✓	×	×
MAP Critical Thinking Project (1984, see Moss & Koziol, 1991)	×	✓	✓	×	×	✓	✓	×

**Note:** × = absent    ✓ = present

## APPENDIX 5

### The detailed descriptions of the six domains of the CTER

#### 1) Ability to write an introduction to gain readers' attention

“The mind is not a vessel to be filled, but a fire to be lighted”  
(Plutarch, n.d., cited in Halsted, 1998, p.1)

This domain can assess critical thinking skills, namely planning or designing how to write an introduction and developing a clear position. The introduction in argumentative essay should contain opening sentence(s), background information and a position (Folse et al., 2002). Writing an argument is an art of persuading readers (Crusius & Channell, 2000), so critical and creative writers should write an introduction to capture readers' attention (Ruggiero, 1984). To make opening sentences in the introduction interesting, writers can a rhetorical strategy: 1) using relevant, interesting quotation, 2) using a metaphor or a simile, 3) using a rhetorical question, 4) beginning with research results, 5) making a confession that is related to a topic, 6) stating a striking problem, 7) stating relevant, crucial fact or 8) stating an opposing view or a debate about a topic (Kalchayanant, 2009; Thongrin, 2009). Background information and thesis statement are also important. Background information links opening sentences to a thesis statement. A thesis statement is like a camera lens (Rosenwasser & Stephen, 2012). If the lens is unclear and is not focused enough, writers cannot give a clear direction to readers (Rosenwasser & Stephen, 2012). Thus, a thesis statement should be written clearly (Bierman & Assali, 1996). If a research article is a product, an opening paragraph of the article is a packaging that is designed for attracting readers' interest (Harwood, 2005). Similarly, if a student's argumentative essay is a book, an introduction of the essay is the books' cover that captures readers' attention.

The following is an example of an introduction in argumentative essay adapted from Oshima and Hogue (2006). The topic is “Separating the Sexes, Just for the Tough Years” (Oshima & Hogue, 2006, p.144). This introduction contains three crucial elements: 1) an opening sentence which is written by using a crucial, relevant fact (one of the rhetorical strategies for gaining readers’ attention), 2) detailed background information and 3) a clear claim.

The middle school years (grades 7 and 8) are known to be the tough years--the years when the uneven pace of girls’ and boys’ physical, emotional and cognitive development is most noticeable. (Opening sentences). Girls are ahead of boys on all counts, and both suffer; therefore, educators debate whether separating boys and girls during these years might improve students’ academic performance. Separate classes are now prohibited in public schools that receive federal funds, but a change in the federal law that prohibits them is under consideration. (Background information). Although some parents and educators oppose same-sex classes, there is some evidence that separating boys and girls in middle school yields positive results. (Claim).

## **2) Ability to give reasons and evidence to support a claim**

This domain can assess critical thinking skills, namely seeking, analyzing and evaluating information, constructing arguments and supporting claims with relevant and compelling evidence. Reasons are logical statements or explanations that support a claim (Butterworth & Thwaites, 2013; Stapleton, 2001a, 2001b). The purpose of giving evidence is to confirm that a reason is true (Crusius & Channell, 2000). Evidence can be personal experience, research studies, personal observation, opinions of the experts, interviews, comparison and contrast, anecdotes and so on (Stapleton, 2001a, 2001b; Kalchayanant, 2009; Lannon, 2009). Concrete evidence will be clear, and readers will have no doubt about it and know exactly whether it is true. As William and Colomb (2007) stated, to support the claim, writers should not merely use their feelings, but they should also use “reasons and evidence” (p.15). Also, writers must avoid fallacy, such as false authority, personal attack, hasty generalizations, non sequitur (making a conclusion that does not follow from a given fact)

and climbing on the bandwagon (arguing that everyone does it or believes it, so readers should do it) (Kalchayanant, 2009; Kirby & Goodpaster, 2007). Giving reasons and concrete evidence is the heart of English argumentative writing (Kalchayanant, 2009) and the core of critical thinking (Kirby & Goodpaster, 2007). In addition, giving reasons and evidence effectively reflects the ability to use two rhetorical appeals—i.e., 1) logos and 2) ethos or a writer’s credibility in making an argument (Connor, & Lauer, 1985; Verlinden, 2005).

### **3) Ability to address a counter-argument and a refutation**

This domain can assess critical thinking skills, namely analyzing and synthesizing the arguments by summarizing thoughts and reflecting final, constructive comments for readers. A counter-argument is an opposing view, whereas a refutation is a statement that refutes the opposing view. Walton (2013) defines the refutation as an “argument attack,” a “challenge” and a “defeater” (p.27). Stating the counter-argument and the refutation in argumentative essays is a problem solving process (Cheng, 2003), can make an argument more convincing (Hollihan & Baaske, 2005) and reflects writers’ critical thinking skills, such as analyzing, synthesizing and evaluating (Stapleton, 2001a, 2001b). Writers can refute a counter-argument by using three following ways: 1) logically explain why one side is stronger than the other, 2) identify flaw in reasoning of the weaker side and/or 3) give an alternative solution (Nussbaum & Schraw, 2007). The third domain of the assessment enables writing teachers to evaluate both depth and breadth of students’ thoughts and increases the students’ awareness of refuting an opposing view in writing. This, in turn, stimulates critical thinking of the students.

### **4) Ability to write a conclusion**

Conclusion reflects a writer’s critical thinking in argumentative writing (Cottrel, 2011). Conclusion should contain two elements: 1) a summary or a restatement of a claim

and 2) a final thought or a final, constructive comment. Constructive comment is useful and intends to help or improve something. It can be a suggestion, a recommendation (for someone to take action) and/or a solution (Kalchayanant, 2009). As Rottenberg (1994) stated, writers should not merely summarize ideas in a conclusion, but they should write a final thought; that is, should find “new” and relevant “idea that emerges naturally from the development of the whole argument” (p.266). Therefore, writing a final, constructive comment in argumentative writing manifests critical and creative thinking of writers (Wang, 2012). The following is an example of a student’ conclusion adapted from Oshima and Hogue (1991):

In conclusion, although the twentieth century gave us many advantages by making us richer, healthier and freer to enjoy our lives, it makes our world dirtier, our people less human and our spiritual lives poorer. (Concluding sentence). Yet, we still continue to make use of technological advancement because we can pursue our other interests and goals freely. Therefore, we must make a joint effort to preserve our natural environment for future generation and take the time now to make our lives more meaningful in our increasing mechanized world. (Final, constructive comment).

Below is the other example of a conclusion in one argumentative writing sample of Thai EFL university student. The topic is “Should a physical punishment be banned?” This conclusion contains a final, constructive comment as follow:

I think when children do the wrong thing, that because children are children and they do not know what is right. Parents should teach them and do not punish them. It is time for us to change our attitude toward physical punishment. When we mention about changing in society, we have to change thing in our family first. So, if we want to stop violence in our society, we have to stop violence in our family first. (No revision)

### **5) Ability to organize ideas in a logical way**

This domain can evaluate critical thinking skills, namely planning how to write and arrange ideas logically and analyzing and revising the structure of arguments to make ideas clear and easy for readers to understand. Organizing ideas in writing refers to developing random statements into general-to-specific ones, removing irrelevant details and arranging

the rest of ideas in a logical and easy-to-follow order (Raimes, 1983). Writers can organize ideas by using transitional or cohesive devices (Thongrin, 2002). The function of such devices is to connect, illustrate, compare, contrast, summarize and/or establish a logical sequence of ideas (Barnet & Bedau, 1993). To evaluate an organization of the essay, teachers should consider whether the essay starts with an introduction, followed by reasons, evidence, counter-argument, refutation and conclusion (Lau, 2011). Also, they should consider whether ideas in each paragraph and ideas between paragraphs are well-organized and connect to each other logically and whether logical connectors are used clearly and appropriately (Lau, 2011).

#### **6) Ability to word choices and sentence structures effectively**

Ability to use word choices and sentence structures skillfully manifests critical thinking in argumentative writing (Foresman, 2017; McLaughlin & Moore, 2012; Yusri, 2018). Word choices and sentence structures are language resources and deep features of writing (Brown, Glasswell & Harland, 2004). Word choices reflect the writer's cognitive process (Payutto, 2010). Word order creates meaning (Kirby & Goodpaster, 2007), and sentence structures reflect the writer's shape of thought (Rosenwasser & Stephen, 2012). A skilled writer will use language as a tool to convince readers to accept his or her argument (Epstein, 1999). The writer will analyze "grammatical structures" and the use of language carefully and accurately in order to convey ideas clearly (Pookcharoen, 2009, p.123). In addition, he or she can use literary features (i.e., metaphor and simile) in argumentative discourse (Maybin & Swann, 2006). Metaphor is a part of language and argumentation (Verlinden, 2005) and can be used to elaborate ideas or make the argument clearer, more effective and more interesting (Connor, 1990; Khopitak, 2015; Mayweg-Paus, Thiebach, & Jucks, 2016; Yeh, 1998a, 1998b). Metaphor reflects critical and creative thinking of the writer (Kirby & Goodpaster, 2007), so it is included as one element of this domain. This domain is very important because it raises the writing teachers' awareness of teaching

language use to EFL students so that they can express critical thinking in writing effectively (Manalo & Sheppard, 2016). This domain can evaluate critical thinking skills, namely analyzing, evaluating and carefully revising word choices and sentence structures for a clear communication.



## APPENDIX 6

### The first version of the CTER

Six domains of assessment	Performance Levels			
	Excellent 4	Good 3	Fair 2	Poor 1
<b>1) Ability to gain readers' attention in an introduction</b>	-The writer writes an effective opening sentence(s) and provides important, relevant background information and a clear claim.	-The writer writes a good opening sentence(s), and provides appropriate relevant background information and the clear claim.	- The writer provides some background information and the claim.	- The writer does not write the introduction.
<b>2) Ability to give reasons and evidence to support the claim</b>	-The writer gives very compelling reasons, concrete evidence and/or very persuasive examples to support the claim. His or her reasoning is insightful, and all views are clearly and thoroughly explained.	- The writer gives convincing reasons, evidence and/or examples to support the claim. His or her reasoning is mostly cogent and sound, and most views are clearly explained.	- The writer gives some reasons, evidence and/or example to support claim. His or her reasoning is adequate. - Some reasons may need more detailed explanations, but they do not cause misunderstanding	- The writer gives limited or no reason, evidence and/or example to support claim. His or her reasoning may be mostly weak, irrelevant or fallacious - Many reasons may be unclear, need many further detailed explanation and/or cause misunderstanding
<b>3) Ability to state a counter-argument and refutation</b>	-The writer states an important counter-argument and insightfully refutes it by using convincing reasons.	-The writer states the counter-argument and refutes it by using convincing reason.	-The writer states the counter-argument and refutes it by using some relevant reasons. Some details may need more explanations, but they do not cause misunderstanding	-The writer addresses counter-argument and refutes it with limited, irrelevant or no reason -May state a strong counter-argument but cannot refute it logically.

Domains of assessment	Performance Levels			
	Excellent 4	Good 3	Fair 2	Poor 1
<b>4) Ability to write a conclusion</b>	-The writer writes an effective conclusion by providing a summary AND a final, constructive comment.	-The writer draws a good conclusion by providing a summary AND/OR a final comment.	-The writer provides a summary WITHOUT a final comment.	- The writer does not write a conclusion. There is neither summary nor a final comment.
<b>5) Ability to organize the ideas logically</b>	-The essay exhibits a skillful organization. The essay is unified by a controlling idea, and a logical progression of ideas makes the argument effective. Transitions between and within paragraphs show the relationships among ideas clearly.	-The essay exhibits a clear organization. The response is mostly unified by a controlling idea, and a logical consequence of ideas mostly makes the argument effective. Transitions between and within paragraphs mostly link the relationships among ideas.	-The essay exhibits an adequate organizational structure.	-The essay exhibits a limited or no organization. Grouping of ideas may be inconsistent and unclear. Transitions between and within paragraphs may fail to connect ideas clearly and logically.
<b>6) Ability to use word choices and sentence structures</b>	Language use enriches the argument. Word choices are varied, appropriate and clear. May use metaphor and simile to convey meanings skillfully. Sentence structures are consistently varied and clear.	Language use helps enrich the argument. Most word choices and sentence structures are varied, appropriate and clear.	Word choice is adequate and somewhat precise. Sentence structures are understandable and demonstrate some variety.	The use of language is often unclear. Word choice is limited and frequently imprecise. Sentence structures are often unclear.

## APPENDIX 7

### The second version of the CTER

CT Domains	Performance levels			
	Excellent 4	Good 3	Fair 2	Poor 1
<b>1) Ability to write an introduction to gain readers' attention</b>	The introduction contains 3 elements: 1) interesting opening sentence(s), 2) relevant background information and 3) clear claim. Background information links opening sentence to the claim clearly. Introduction is outstanding and successfully captures readers' interest	The introduction contains 3 elements: 1) good opening sentence(s), 2) relevant background information and 3) a claim.	The introduction contains 2 elements: 1) opening sentence(s) and 2) relevant background information. No claim.	No introduction.
<b>2) Ability to give reasons and evidence to support claim</b>	Gives at least 2 strong, relevant reasons with concrete evidence to support a claim. Reasons and evidence are clearly and convincingly elaborated.	Gives at least 2 relevant reasons with concrete evidence to support the claim. However, the evidence is not elaborated.	Gives at least 2 relevant reasons without evidence.	No relevant reasons and no evidence.
<b>3) Ability to state counter-argument and refutation</b>	States a strong counter-argument and refutes it with much stronger reasons and concrete evidence. Reasons and evidence are clearly and convincingly elaborated.	States a strong counter-argument, and refutes it with much stronger reasons and evidence. However, the evidence is not elaborated.	States a general counter-argument and refutes it by using general reasons without evidence	-No counter-argument and no refutation. -OR, states strong counter-argument but cannot attack it. This makes a student's own argument less effective and weak.

CT Domains	Performance level			
	Excellent 4	Good 3	Fair 2	Poor 1
<b>4) Ability to write a conclusion</b>	The essay has two elements: 1) a summary AND 2) a final, constructive comment.	The essay has two elements: 1) a summary AND 2) a final, general comment. OR, may have only a final, constructive comment.	-The essay has a summary WITHOUT a final comment.	No conclusion.
<b>5) Ability to organize ideas in a logical way</b>	The essay is well-organized. That is, the essay has clear main parts: an introduction, a body paragraph and conclusion. Ideas within each paragraph AND between paragraphs connect with each other logically. Ideas flow smoothly and can be followed very easily	The essay is mostly coherent and organized. That is, the essay has an introduction, a body paragraph and a conclusion, and ideas in each part mostly connect with each other. Ideas can be followed easily.	Some parts of the essay are more coherent and organized than the other parts. The essay can be followed with some difficulties.	-Ideas lack organization. The connectors do not show a logical connection between ideas and cause misunderstanding.
<b>6) Ability to use word choices and sentence structures effectively</b>	-Uses word choices and sentence structures that are varied, appropriate and clear throughout the essay. <u>May</u> use a metaphor or a simile to convey ideas clearly and appropriately. Word choices are outstanding and enrich an argument	-Word choices and sentence structures are mostly appropriate and clear. Word choices enrich an argument	Word choices and sentence structures are quite clear and demonstrate some variety	Word choices and sentence structures are not varied and not clear.

## APPENDIX 8

### The final version of the CTER

Critical thinking domains	Levels of performance			
	Excellent (4)	Good (3)	Fair (2)	Poor (1)
<b>1) Ability to write an intro to gain readers' attention</b>	-Writes an opening sentence by using a rhetorical strategy for gaining readers' attention clearly, writes clear, detailed background information and writes a claim clearly. The introduction is clear and well-elaborated.	-Writes a relevant opening sentence, detailed background information and a clear claim.  -May only state a claim and briefly provide reasons.	-May write an opening sentence and background information without a clear claim. -May write an opening sentence and a claim without detailed background information. -OR writes a claim only.	-Write an unclear and irrelevant introduction.
<b>2) Ability to give reasons and evidence to support a claim</b>	-Gives at least 3 strong, relevant reasons with relevant, accurate concrete evidence. May use many kinds of solid and credible evidence, so the arguments are very strong. All reasons and evidence are non-fallacious and clearly, thoroughly and convincingly explained.	-Gives at least 2 strong, relevant reasons with relevant, concrete evidence. Reasons and evidence are convincing and non-fallacious, although few minor details may need more elaborations. -May give many reasons, and good reasons far outnumber poor reasons.	-May give 1 strong, relevant reason with relevant evidence. Yet, the evidence may not be concrete or clearly explained. -May give 2-3 relevant reasons without evidence. -May give many relevant reasons, but they are still unclear or need more evidence. -May give many reasons which have different levels of quality. -OR the number of good reasons equals the number of poor reasons.	-Reasons and evidence may be inaccurate, irrelevant, fallacious, too weak or too limited. -May give many reasons, but poor reasons far outnumber good reasons. -OR reasons and evidence are absent.
<b>3) Ability to state a counter-argument and a refutation</b>	-States at least 2 counter-arguments and refutes them by using convincing reason(s) with relevant, accurate, concrete evidence, AND explains all reasons and evidence clearly and thoroughly.	-States at least 1 counter-argument and refutes it by using more convincing reason(s) with relevant, concrete evidence, AND explains all reasons and evidence clearly.	-States 1 counter-argument and refutes it by using relevant reason(s) with relevant evidence. Yet, the evidence may not be concrete or clearly explained. -May state 2 counter-arguments, but can refute only one.	-States counter-argument(s) and refutation(s) inaccurately, unclearly, incompletely, irrelevantly or fallaciously. -May state a strong counter-argument but cannot refute it -OR a counter-argument and a refutation are absent.

Critical thinking domains	Excellent (4)	Good (3)	Fair (2)	Poor (1)
<b>4) Ability to write a conclusion</b>	-Gives a summary or a restatement of a claim AND writes a final, relevant, <u>constructive</u> comment.	-Gives a summary or a restatement of a claim AND writes a final, relevant comment  -OR only writes a final, relevant, <u>constructive</u> comment.	-Gives a summary or a restatement of a claim without a final, relevant comment (or vice versa)	-May write a conclusion unclearly, irrelevantly or illogically.  -OR a conclusion is absent.
<b>5) Ability to organize ideas in logical way</b>	-Ideas are carefully and very well organized, so ideas within each paragraph and between paragraphs flow smoothly and are very easy to follow. All connectors show a clear and logical connection between ideas.	-Overall, the essay has a clear organization. Although unclear connectors between ideas occasionally appear, they do not cause misunderstanding.	-Some parts of the essay are organized, but some parts are poorly organized or have connectors that obscure the connection of ideas and cause misunderstanding.	-Ideas lack organization. The connectors do not show a logical connection between ideas and cause misunderstanding.
<b>6) Ability to use word choices and sentence structures</b>	-Uses various, appropriate word choices and sentence structures throughout the essay. All word choices and all sentence structures convey meanings clearly. May use metaphor or simile to convey ideas clearly and appropriately.	-Most word choices and most sentence structures are appropriate and clear. Although unclear or inappropriate word choices occasionally appear, they do not cause misunderstanding.	-Word choices and sentence structures convey meanings to some extent. Some inappropriate and unclear word choices and/or sentence structures do appear and cause misunderstanding.	-All or almost all of word choices and sentence structures are unclear, inappropriate or distort meaning dramatically, causing misunderstanding.

## APPENDIX 9

### Rating manual for the CTER

#### Rating manual for the final version of the CTER

##### Domain 1

The effective rhetorical strategies for gaining readers' attention include: 1) using a relevant, interesting quotation, 2) using a metaphor or a simile, 3) using a rhetorical question, 4) using research results, 5) confessing something that is related to a writer's life and a topic, 6) stating a striking problem, 7) stating an opposing view or a debate pertaining to a topic and 8) stating a crucial fact that is related to the topic.

##### Domain 2

-Reasons are logical statements or explanations that are given to support a claim.

-Evidence proves that a reason is true. Evidence can be a logical elaboration, first or second-hand experience, research findings, personal observation, opinions of the experts in the field, interviews, comparisons and contrasts, anecdotes and/or a personal narrative.

-Concrete evidence will be clear and exist in a form that readers can see or feel. Readers will have no doubt about it and know exactly whether it is true.

##### Domains 3

A student can state counter-argument (s) and refutation (s) in three following ways. First, after giving the first reasons, the student can state and refute a counter-argument and then do the same with the second and the third reasons (See Model 1). Second, the student can wait until they finish giving all reasons/evidence and then state and refute the counter-argument in one separate paragraph (See Model 2). Third, the student can address the counter-argument and the refutation after writing an introduction (See Model 3).

###### **Model 1**

- a thesis statement
- reasons and evidence 1  
*& a counter-argument  
and a refutation*
- reasons and evidence 2  
*& a counter-argument  
and a refutation*

###### **Model 2**

- a thesis statement
- reasons and evidence 1
- reasons and evidence 2
- reasons and evidence 3  
*- a counter-argument  
and a refutation*

###### **Model 3**

- a thesis statement
- *a counter-argument and  
a refutation*
- reasons and evidence 1
- reasons and evidence 2
- reasons and evidence 3

##### Domain 4

A constructive comment is useful and intends to help or improve something. It can be a useful suggestion, a useful recommendation (for someone to take an action) or a useful

solution. The following is an example of a conclusion which contains a final constructive comment. This conclusion of an argumentative essay is adapted from Oshima and Hogue (1991).

In conclusion, although the twentieth century gave us many advantages by making us richer, healthier and freer to enjoy our lives, it makes our world dirtier, our people less human and our spiritual lives poorer. (Summary). However, we wish to continue to enjoy the benefits of technological advancement because they free us to pursue our other interests and goals. Thus, we must make a joint effort to preserve our natural environment for future generation and take the time now to make our lives more meaningful in our increasing impersonal, mechanized world. (Final, constructive comment).



## APPENDIX 10

### The checklist

#### 1. A checklist for investigating the use of a rhetorical strategy in the introduction

A user of the CTER can examine a rhetorical strategy for gaining readers' attention used by a student by filling the table below.

Tick (/) in front of a type of a rhetorical strategy for gaining readers' attention
<input type="radio"/> Use a relevant, interesting quotation
<input type="radio"/> Use a metaphor or a simile
<input type="radio"/> Use a rhetorical question
<input type="radio"/> Use research results
<input type="radio"/> Confess something that is related to a writer's life and a topic
<input type="radio"/> State a striking problem
<input type="radio"/> State an opposing view or a debate pertaining to a topic
<input type="radio"/> State crucial fact that is related to the topic
<input type="radio"/> Other (if you choose 'other' please specify .....)

#### 2. A checklist for investigating the use of evidence

Future researchers can use the checklist below to examine a type of evidence used by a student.

Tick (/) in front of a type of evidence
<input type="radio"/> logical elaboration
<input type="radio"/> personal experience
<input type="radio"/> research findings
<input type="radio"/> personal observation
<input type="radio"/> opinions of the expert in the field
<input type="radio"/> interviews
<input type="radio"/> comparisons and contrasts
<input type="radio"/> anecdotes
<input type="radio"/> Others (if you choose 'other' please specify .....)

### 3. A checklist for investigating the type of fallacy used by a student

Future researchers can examine a type of fallacy used by a student writer by filling the checklist below.

<b>Tick (/) in front of a type of fallacious reasons</b>
<input type="radio"/> Sweeping generalizations
<input type="radio"/> Post hoc, ergo propter hoc (“after this, therefore because of this”)
<input type="radio"/> Non sequitur (“it doesn’t follow”)
<input type="radio"/> False authorities
<input type="radio"/> Hasty generalization
<input type="radio"/> Climbing on the bandwagon
<input type="radio"/> Personal attack
<input type="radio"/> Circular reasoning
<input type="radio"/> Others (if you choose ‘other’ please specify .....)

#### 3.1 Detailed descriptions about a fallacy in English argumentative writing

The following are examples of the fallacies that students should avoid in their argumentative writing (Folse et al., 2002; Kalchayanant, 2009).

- **Sweeping generalizations**

A writer uses words that are too board, such as *all*, *always* and *never*. Also, he or she cannot support a claim with logical, concrete evidence.

Example 1: All Americans eat fast food.

Example 2: All police officers are corrupt.

- **Post hoc, ergo propter hoc (“after this, therefore because of this”)**

When one event happens, it does not necessarily make a second event happen.

Example 1: Since the democratic party came to power, we have been in an economic slump.

- **Non sequitur (“it doesn’t follow”)**

A writer makes a conclusion that does not logically follow from the facts given.

Example 1: Ben is a famous writer, so he will be a good teacher of English literature.

Example 2: Sam went to the football game, then he got drunk. Therefore, football games cause the drunkenness.

- **False authorities**

A writer uses names or opinions of famous people to support his or her claim, but those people are not experts on the topic of the argument.

Example 1: Katy Perry is a good singer, so she would make a good orchestra conductor.

Example 2: David Beckham drinks Pepsi, so Pepsi is better than Coke.

- **Hasty generalization**

Hasty generalizations are making quick judgments based on inadequate information.

Example 1: A new neighbor moved in downstairs on Saturday. My television disappeared on Sunday. Therefore, the new neighbor stole my TV.

- **Climbing on the bandwagon**

A writer does something because it is popular and everyone else is doing it.

Example 1: Almost everyone agrees that Professor Andrew is unfair in flunking his students, so why shouldn't you?

Example 2: Everyone's doing it. Why shouldn't I?

- **Personal attack**

A writer attacks person rather than an argument. Example below shows that the writer attacks Professor Andrew's sickness rather than his administrative incompetence.

Example 1: Prof. A cannot be a good department chair as he is lame in the left leg.

- **Circular reasoning**

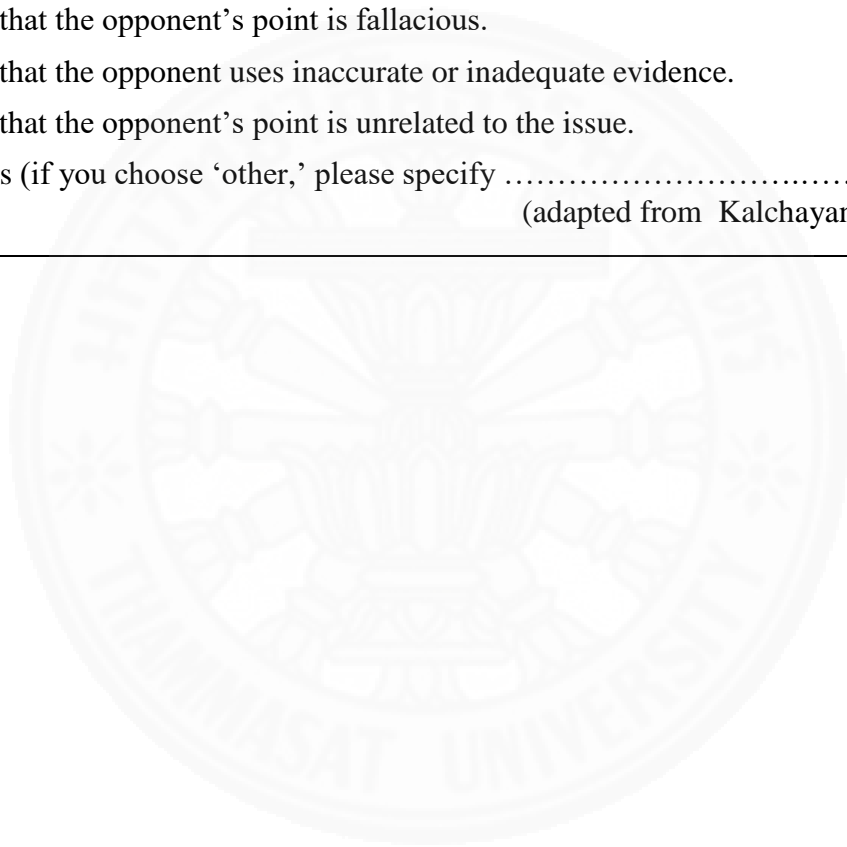
A writer restates a claim without providing any logical reasons to prove it.

Example 1: The traffic is heavy because there are many cars on the roads.

#### 4. A checklist for investigating a student's strategy of stating a counter-argument and a refutation

Future researchers can examine a student's strategy for addressing a counter-argument and a refutation by filling the checklist below.

<b>Tick (/) in front of the strategy of addressing a counter-argument</b>
<input type="radio"/> State that the opponent has a point, but your point is better or more logical.
<input type="radio"/> State that the opponent's point is untrue or partly true.
<input type="radio"/> State that the opponent's point is fallacious.
<input type="radio"/> State that the opponent uses inaccurate or inadequate evidence.
<input type="radio"/> State that the opponent's point is unrelated to the issue.
<input type="radio"/> Others (if you choose 'other,' please specify .....) (adapted from Kalchayanant, 2009).



## APPENDIX 11

### Anchor papers 1 – 8 for the final version of the CTER

#### Anchor Paper 1

**Topic: Separating the Sexes, Just for the Tough Years**  
(adapted from Oshima and Hogue’s (2006) argumentative essay sample)

The middle school years (grades 7 and 8) are known to be the tough years--the years when the uneven pace of girls’ and boys’ physical, emotional and cognitive development is most noticeable. Girls are ahead of boys on all counts, and both suffer. Educators debate whether separating boys and girls during these years might improve students’ academic performance. Separate classes are now prohibited in public schools that receive federal funds, but a change in the federal law that prohibits them is under consideration. Although some parents and educators oppose same-sex classes, there is some evidence that separating boys and girls in middle school yields positive results.

Opponents of single-sex education claim that test scores of students in all-girl or all-boy classes are no higher than those of students in mixed classes (“Study: All-Girls Schools Don’t Improve Test Scores, 1998”). However, the research is inconclusive. Despite the fact that some research shows no improvement in test scores, other research shows exactly opposite results (Blum, 2002). More importantly, many psychologists believe that test scores are the wrong measuring sticks. They believe that self-confidence and self-esteem issues are more important than test scores. In same-sex classes, girls report increased confidence and improved attitudes towards math and science, for example (“Study: All-Girls Schools Don’t Improve Test Scores, 1998”). These are results that cannot be calculated by a test but that will help adolescents become successful adults long after the difficult years of middle school are past. Carol Gilligan, New York University professor, is certain that girls are more likely to be “creative thinkers and risk-takers as adults if educated apart from boys in middle school” (Gross, 2004). Boys, too, gain confidence when they do not have to compete with girls. Boys at this age become angry and fight back in middle school because they feel inferior when compared to girls, who literally “out-think” them. With no girls in the classroom, they

#### **Introduction:**

- A student writes an opening sentence by stating a crucially important fact--one of the rhetorical strategies for gaining readers’ attention. He/she provides relevant background information which indicates people’s different opinions about the topic and its effects on education. He/she states claim clearly.

#### **Counter-argument 1**

**Refutation to counter-argument 1 with reason and evidence**

are more at ease with themselves and more receptive to learning (Gross, 2004).

Opponents also maintain that separate classes (or separate schools) send the message that males and females cannot work together. They say that when students go into the work force, they will have to work side-by-side with the opposite sex, and attending all-girl or all-boy schools denies them the opportunity to learn how to do so (“North Carolina school stops same-sex classes, 2004”). However, such an argument completely ignores the fact that children constantly interact with members of the opposite sex outside school. From playing and squabbling with siblings to negotiating allowances, chores and privileges with their opposite-sex parent, children learn and practice on a daily basis the skills they will need in their future workplaces.

Counter-argument 2

Refutation to counter-argument 2 with reason and evidence

The final argument advanced by opponents of same-sex education is that it is discriminatory and, therefore, unconstitutional. However, research supports exactly the opposite conclusion: that discrimination is widespread in mixed classes. Several studies have shown that boys dominate discussions and receive more attention than girls and that teachers call on boys more often than they call on girls, even when girls raise their hands (“North Carolina school stops same-sex classes, 2004”). Clearly, this is discriminatory. As discrimination occurs in mixed classes, discrimination is not a valid argument.

Counter-argument 3

Refutation to counter-argument 3 with reason and evidence

While it is evident that the arguments against same-sex classes are not valid, many people involved in middle-school education say that same-sex classes provide a better learning environment. Boys and girls pay less attention to each other and more attention to their schoolwork (Marquez, 2004). As one teacher noted, “Girls are more relaxed and ask more questions; boys are less disruptive and more focused” (“North Carolina school stops same-sex classes, 2004”). Girls are less fearful of making mistakes and asking questions in math and science, while boys are less inhibited about sharing their ideas in language and literature. Also, schoolchildren are not disadvantaged by lack of contact with the opposite sex because they have many opportunities outside the school setting to interact with one another.

Reason and evidence to support the claim

As discussed earlier, there are many advantages of separating boys and girls in middle school. Therefore, in my opinion, the law prohibiting same-sex classes in public schools should be changed in order to help middle school students fully develop their intellectual growth, academic

**Conclusion:**  
a concluding sentence and a final, relevant, constructive comment

performance, and higher-order thinking so that they can be the quality students in the future.

## References

- 1 “Study: All-Girls Schools Don’t Improve Test Scores.” CNNinteractive 12 Mar.1998. 2 June 2004 <<http://www.cnn.com/US/9803/same.sex.classes/>>.
- 2 Blum, Justin. “Scores Soar at D.C. School with Same-Sex Classes.” [washingtonpost.com](http://www.washingtonpost.com) 27 June 2002.
- 3 June, 2004 <<http://washingtonpost.com/ac2/wp-dyn/A52023-2002Jun26?language=printer>>.
- 4 “North Carolina school stops same-sex classes, 2004.” American Civil Liberties Union News 5 Apr. 2000
- 5 Marquez, Laura. “No Distraction? Proposed Title IX Changes Would Allow Separate Classrooms for Girls and Boys.” ABC News 13 May 2004. 2 June 2004<[http://abcnews.go.com/sections/WNT/US/same\\_sex\\_classrooms\\_040513.html](http://abcnews.go.com/sections/WNT/US/same_sex_classrooms_040513.html)>.

### Score results of anchor paper 1

<b>Domains of critical thinking assessment in argumentative writing</b>	<b>Score</b>
Domain 1: Ability to write an introduction to gain readers’ attention	4/excellent
Domain 2: Ability to give reasons and evidence to support claim	4/excellent
Domain 3: Ability to state a counter-argument and a refutation	4/excellent
Domain 4: Ability to write a conclusion	4/excellent
Domain 5: Ability to organize ideas in logical way	4/excellent
Domain 6: Ability to use word choices and sentence structures	4/excellent

## Anchor Paper 2

### **Topic: Should all students be required to wear a uniform?**

(adapted from Folse, Muchmore-Vokoun and Solomon's (2002) argumentative model essay)

Individualism is a fundamental value in the United States and many Americans believe in the right to freely express their own opinion without fear of punishment. However, this value is coming under fire in an unlikely place--the public school classroom. The issue is school uniforms. Should public school students be allowed to make individual decisions about clothing, or should all students be required to wear a uniform? In my opinion, school uniforms are the better choice for three main reasons.

#### **Introduction**

A student writes an opening sentence by stating a very important fact--one rhetorical strategy for gaining readers' attention. He/she also states claim clearly.

First, school uniforms influence students to act responsibly in groups and as individuals. Uniforms can create a feeling of unity among students. When students do something as a group, such as attend meetings in the auditorium or eat lunch in the cafeteria, the fact that they all wear the same uniform would create a sense of community. This togetherness felt by students can increase their pride of belonging to their educational institution which, in turn, can lead to a unity of students in school. More importantly, statistics show the positive effects that school uniforms have on violence and truancy. According to a recent survey in Hillsborough County, Florida, incidents of school violence dropped by fifty percent, attendance and test scores improved, and student suspensions declined approximately thirty percent after school uniforms were introduced. This evidence clearly proves that uniforms are better.

#### **Reason and evidence 1**

Second, school uniforms would help make all students feel equal. People's standards of living differ. Some people are well-off while others are not. Based on my experience as a student, some of my friends choose to wear only fashionable, costly clothes in order to show that they are rich and they likely choose to make friends with only those having the same standards of living. I think sometimes students forget that school is a place for learning, not for promoting a fashion show, wealth, appearance or popularity. Implementing mandatory school uniforms, therefore, would cut back on the socio-economic differences among students and make all students look the same regardless of their financial status. School uniforms would also promote pride and help raise the self-esteem of students who cannot afford to wear stylish

#### **Reason and evidence 2**

clothing. Students do not have to worry about what other students say about what they are wearing as well.

Opponents of mandatory uniforms may say that students who wear school uniforms cannot express their individuality. This point has some merit on the surface. However, as stated previously, school is a place to learn, not to flaunt wealth and fashion. Society must decide if individual expression through clothing is more valuable than improved educational performance. It is important to remember that school uniforms would be worn only during school hours. Students can express their individuality in the way they dress outside of the classroom.

counter-argument and refutation

In sum, there are many benefits to implementing mandatory school uniforms for students. Studies show that the students learn better and act more responsibly when they wear uniforms. Thus, public schools should require uniforms in order to benefit not only the students themselves but also society as a whole.

Conclusion (a summary and a final, relevant, constructive comment)

### Score results of anchor paper 2

<b>Domains of critical thinking assessment in argumentative writing</b>	<b>Score</b>
Domain 1: Ability to write an introduction to gain readers' attention	4/excellent
Domain 2: Ability to give reasons and evidence to support claim	3/good
Domain 3: Ability to state a counter-argument and a refutation	3/good
Domain 4: Ability to write a conclusion	4/excellent
Domain 5: Ability to organize ideas in logical way	4/excellent
Domain 6: Ability to use word choices and sentence structures	4/excellent

## Anchor Paper 3

### **Topic: Organic Food**

(adapted from Chase's (2011) argumentative essay model)

My opinion about organic food is it amazing and healthy for you. Many reasons why I believe organic food is good for you. First, it is free of pesticides. Second, studies have shown that a natural diet can lower your chances of acquiring unhealthy diszses. Third, organic food has some vitamin very useful for human body. Final, organic food often comes from independent farmers rather than large, money making corporations. It is best to support these farmers rather than big businesses.

#### **Introduction**

- A student writes states a claim and briefly gives reasons.

First, organic food is free of pesticides. This means that only natural products are used on the produce and for the animals. Conventional foods that are produced is usually has the use of methods to fertilize. See convential foods use fertilizers to promote plant growth. In Organic foods they will apply natural fertilizers, such as manure or compost, to feed soil and plants. When conventional foods are spray down to reduce pests and diease, you would not know what kind of things that might get you sick. Organic foods uses beneficial insects and birds mating disruption on traps to reduce pest and disease. By using no pesticides on growing or producing the food, we will not eat unnecessary pesticides and put it in our body.

#### **Reason and Evidence 1**

Since we are not eating pesticides, it lowers are chances of getting diseases from the pesticides. This supports the overall reason why organic food is so good and healthy for you. There is less fat and unnatural ingredients in the food, which can clog your arteries, increase your blood pressure, and be bad for your heart. Thus, organic good is better for your body and makes you less likely to get diseases from your food.

#### **Reason and Evidence 2**

In addition, the vitamins in organic foods, such as vitamin C, iron, Calicum, chromium, selenium, boron, litium, magnesium, betakeratin, and vitamin B, are very important to humans' bodies. These vitamins help you live a healthier life. You don't have to be over weight. These vitamins are beneficial to your health needs. These vitamins in this organic food prevents diseases which is a use of medicine.

#### **Reason and Evidence 3**

Lastly, supporting organic food helps to support the farmers and "mom and pop" shops. Lots of organic food comes from local farmers. In my neighborhood, I buy organic food

#### **Reason and Evidence 4**

from the local farmers; I feel good about myself when I do this because I'm supporting the local people who need it the most.

I know some people can argue that organic food is bad because it's so expensive. Yes, it's expensive, but it's totally worth it because, in the end it'll save you money because you won't have to pay so many doctors bills from getting sick so much. The organic food is useful to people because it can heal; for example, the boron on the organic foods prevents osteoporosis, a disease of the bone leading to an increase risk of fracture. Osteoporosis is usually found in women. The organic food so prevents diabet that is found in most people nowadays.

**Counter-argument  
and Refutation**

As an old saying goes "Let food be thy medicine, and medicine be thy food...", I think we should use organic, worth medicine, to save and improve our health.

**A final, relevant  
constructive comment**

### Score results of anchor paper 3

Domains of critical thinking assessment in argumentative writing	Score
Domain 1: Ability to write an introduction to gain readers' attention	3/ Good
Domain 2: Ability to give reasons and evidence to support claim	3/ Good
Domain 3: Ability to state a counter-argument and a refutation	3/ Good
Domain 4: Ability to write a conclusion	3/ Good
Domain 5: Ability to organize ideas in logical way	3/ Good
Domain 6: Ability to use word choices and sentence structures	3/ Good

### Comment on anchor paper 3

This student gives 4 logical reasons to support a claim and does a good job in giving supporting evidence. However, to gain 4 scores (Level Excellent), in the fourth paragraph, he/she should make a clear distinction between vitamins in organic food and vitamins in non-organic one because both organic and non-organic food have those kinds of vitamin. Providing clear, detailed explanations would help make the provided reason stronger and more convincing. Also, he/she reasoned that "You don't have to be over weight because these vitamins are beneficial to your health needs" (in paragraph 4). Although this reason are logical, it needs further evidence about how organic food helps prevent people from being overweight. Although few minor details may need more explanations or clarification, the given reasons and evidence are logical. Therefore, this essay should receive 3 scores in all domains of the assessment.

## Anchor Paper 4

Adapted from anchor paper in level 5 of TOEFL iBT independent writing task (Educational Testing Service, 2005)

**Direction: Do you agree or disagree with below statement?**

***“Always telling the truth is the most important consideration in any relationship.”***

**Use specific reasons and examples to support your answer**

Have you ever thought about the effects of always telling the truth in all situations? In today’s world, the traditional virtue of telling the truth in all situations is increasingly doubted by many people. While many think telling the truth is not always the best policy when dealing with people, some believe it is the best way to solve all kinds of problem.

-The student uses a rhetorical question (one of the rhetorical strategies for gaining readers’ attention) as an opening sentence. He/she also provides background information. However, he/she still lacks clear claim.

As we know, the truth is often offending and may not be a very nice thing to hear and say. Lies or white lies often have their advantages. Nowadays, the manipulation of white lies is the most obvious in the business world. How many times have we heard that some product is “the finest” or “the cheapest”? How many times have we heard that products have such “magical functions”? I think most advertisings are about persuasion, and many would agree that if a company is to tell the absolute truth about its products, no one would be interested in even having a look at the products. The same logic applies to human relationships. If your friend had worn a newly purchased dress on her birthday and energetically asked you if it was a worthy buy, would you freely express your opinion that you had never seen a dress as the one she’s currently wearing and spoil her birthday? Unarguably, hiding entirely or partially the truth in some situations can be quite handy indeed. Confrontations and disputes in family or among friends can seemingly be avoided.

**Counter-argument**

However, there is always the risk factor of the truth emerging sooner or later when telling an untruth. The basic trust in any relationships (businessman, customer, friends, parents or children), will be blotched, and would have an impact on the future relationship between both parties. The story of the “the boy who cried wolf” fully illustrates the consequences of telling untruths. No one will believe you when you’re telling the truth. Your word will have no weighting.

**Refuting the counter-argument**

**Reason and evidence**

Another “bad factor” of telling untruths is that you have absolutely no control over when the truth (of previous untruths) will emerge. Untruths breed pain in both parties: tears when the truth is uncovered after a period of time; fear and the burden of sharing a “secret” in the long run. So hiding the truth is not beneficial to either party.

Refuting the counter-argument 1

I believe everyone hates a lie. Although it is the trend to occasionally hide the truth in relationships, I recommended you not follow the trend since the risk and the consequences of the truth can unfold and overwhelm the minimal advantages one can derive from telling a lie. Since relationships are founded on “trust” which goes hand in hand with “truth”, it is better to tell the truth to keep a sustainable relationship with people, especially with your loved ones.

Conclusion (a summary and a final, constructive comment)

#### Score results of anchor paper 4

Domains of critical thinking assessment in argumentative writing	Score
Domain 1: Ability to write an introduction to gain readers’ attention	2/Good
Domain 2: Ability to give reasons and evidence to support claim	3/Good
Domain 3: Ability to state a counter-argument and a refutation	3/Good
Domain 4: Ability to write a conclusion	4/Excellent
Domain 5: Ability to organize ideas in logical way	3/Good
Domain 6: Ability to use word choices and sentence structures	4/Excellent

## Anchor Paper 5

Adapted from anchor paper in level 4 of TOEFL iBT independent writing task (Educational Testing Service, 2005)

**Direction: Do you agree or disagree with below statement?**

***“Always telling the truth is the most important consideration in any relationship.”***

**Use specific reasons and examples to support your answer**

I admit honesty is one most important thing in human relationship. Telling the truth is one ways to show one’s honesty. however, in some situations, white-lies are necessary not to damage the relationship. In this paper, I explain why i disagree with the given statement.

First, It is not good to tell someone his or her faults or defects directly. By doing so, you might feel good about telling the truth, but it could hurt the people who will listen. It is naturl that people make mistakes and not every one perfect. Although these shortcomings, we live together by understanding each other and tolerate the errors as we could be the one who makes the same mistakes in next time. By being pointed out our mistakes or things other people think abominable, such as the way we look, the clothes we wear, the way we talk and so on, we might hate the person who criticize us rather than being thankful or correcting those.

Second, It is better not to tell someone serious facts that can discourage him or her. I speak this from my experience. About one year ago, my nephew got a terrible car accident. Doctor said he won’t be able to walk again. We were all shocked and felt desprate. but, my other nephews decided not to tell the fact to their mother as she could pass off or be shocked to generate other problems of her own. I felt guilty about not telling the truth to her, who was relieved and become hopeful by the white-lies that her son would be okay. One year since, my nephew is getting better and his mother has been a good supporter. If we had told her the truth, she would have been so desparate that she couldn’t have taken good care of her son and her own health would have been in danger.

It is clear that honesty is great virtue, but “always telling the truth” is not always the key to maintain good human relationship. By refraining from telling the truth in certain occasions, we prevent other people from being hurt by the remarks.

### Score results of anchor paper 5

<b>Domains of critical thinking assessment in argumentative writing</b>	<b>Score</b>
Domain 1: Ability to write an introduction to gain readers’ attention	3/ Fair
Domain 2: Ability to give reasons and evidence to support claim	2/ Fair
Domain 3: Ability to state a counter-argument and a refutation	1/ Poor
Domain 4: Ability to write a conclusion	3/ Good
Domain 5: Ability to organize ideas in logical way	3/ Good
Domain 6: Ability to use word choices and sentence structures	3/ Good

## Anchor Paper 6

Adapted from anchor paper in level 3 of TOEFL iBT independent writing task (Educational Testing Service, 2005)

**Direction: Do you agree or disagree with below statement?**

***“Always telling the truth is the most important consideration in any relationship.”***

**Use specific reasons and examples to support your answer**

I think Telling true is important think in relationship between people cuz you must trust in him or her, nobody wants to be a fool, and truth makes a relation strong. So I agree that you should tell truth.

First I wanna tell you that the trust will keep a relation going on with confiden and confiden is something so important and you going to trust in person only one. If you lose trust in someone I think you will never be able to trust in this person again and again. Will you trust a very importan value to someone that has already stolen you? I bet you won't. Nobody also like to be seen as fool. It's not nice to be seen as stupid one. It's not a good sensation to be kept in lie while there are other who know the true. Can you imagine the sensation of been betrayed? How can you look at someone's face that betrayed you? A relationship in this way won't be nice and trustful. Once I saw that for a relationship be strong, you must trust in person at the point of leting yourself fall back in the person arms. I mean, you trust that the person is going to be on your back to hold you on. It won't be possible with a person that lies to you. The truth mean you can't lie. if you can't lie you won't be able to betray you lover. so you won't be able to hide anything. If you can't betrey you're going to have a perfect relationship. The truth is really basis of relationship. some may say we sometime need telling a lie to avoid conflicts. Like when your 49 year old aunt ask you whether she is still beautiful. If you say no (truth for sure) we can imagine result. In this case, we can use while lie to find good reason to use minor lies. but, I think telling truth is better. insteed of lie, why we don't tell her directly she is not beautiful and suggest her how to make her beautiful, so her looking will better.

I think truth is so important in relationship. it gives you some safety in relationship, the sensation that you won't be betrayed.

### Score results of anchor paper 6

<b>Domains of critical thinking assesment in argumentative writing</b>	<b>Score</b>
Domain 1: Ability to write an introduction to gain readers' attention	2/ Fair
Domain 2: Ability to give reasons and evidence to support claim	2/ Fair
Domain 3: Ability to state a counter-argument and a refutation	2/ Fair
Domain 4: Ability to write a conclusion	2/ Fair
Domain 5: Ability to organize ideas in logical way	2/ Fair
Domain 6: Ability to use word choices and sentence structures	2/ Fair

### Anchor Paper 7

Adapted from anchor paper in level 1 of TOEFL iBT independent writing task (Educational Testing Service, 2005)

**Direction: Do you agree or disagree with below statement?**

***“Always telling the truth is the most important consideration in any relationship.”***

**Use specific reasons and examples to support your answer**

yes, the true important in relation. person hope people love talk truyh The relationshid need building or true. The truth bad or hard better that lied.

#### Score results of anchor paper 7

Domains of critical thinking assessment in argumentative writing	Score
Domain 1: Ability to write an introduction to gain readers' attention	1/Poor
Domain 2: Ability to give reasons and evidence to support claim	1/Poor
Domain 3: Ability to state a counter-argument and a refutation	1/Poor
Domain 4: Ability to write a conclusion	1/Poor
Domain 5: Ability to organize ideas in logical way	1/Poor
Domain 6: Ability to use word choices and sentence structures	1/Poor

### Anchor Paper 8

Adapted from anchor paper in level 1 of TOEFL iBT independent writing task (Educational Testing Service, 2005)

**Direction: Do you agree or disagree with below statement?**

***“Always telling the truth is the most important consideration in any relationship.”***

**Use specific reasons and examples to support your answer**

relation need building on truth. Yes. The true good but thousand of lied i feel no good. Anyway telling littlr truth you tryng to help peple not hert them. when friend ask if she fat and she actially weighth 60 kilo, you answer her fat like cow bewcause you make her feel good and you just need to be sutil and friendle and responder her like that , it isnt something it cannot be fixed bewcause you can not probablie help her.

#### Score results of anchor paper 8

Domains of critical thinking assessment in argumentative writing	Score
Domain 1: Ability to write an introduction to gain readers' attention	1/Poor
Domain 2: Ability to give reasons and evidence to support a claim	1/Poor
Domain 3: Ability to state a counter-argument and a refutation	1/Poor
Domain 4: Ability to write a conclusion	1/Poor
Domain 5: Ability to organize ideas in logical way	1/Poor
Domain 6: Ability to use word choices and sentence structures	1/Poor

## APPENDIX 12

### Experts' validation survey for the CTER

#### Experts' validation survey for the second version of the CTER

Name-surname: \_\_\_\_\_

Department/faculty/Institute:

\_\_\_\_\_

Area of your research expertise:

\_\_\_\_\_

Years of your teaching experience: \_\_\_\_\_

You will be asked to validate the CTER—a critical thinking analytic rubric developed by the researcher (Mr.Nattawut Nakkaew).

#### **Part 1: Validation of the six domains of the critical thinking assessment in English**

##### **argumentative essays of EFL university students**

Please indicate the level of your agreement *or* disagreement with each domain of the critical thinking assessment proposed by putting a check mark (/) in the box. (SA = Strongly Agree, A= Agree, U = Undecided, D = Disagree, SD = Strongly Disagree)

Six domains of critical thinking assessment	Level of your agreement or disagreement towards the six domains					Additional comments or suggestions about each domain proposed
	SA 5	A 4	U 3	D 2	SD 1	
Domain 1: Ability to write an introduction to gain readers' attention						

Six domains of critical thinking assessment	Level of your agreement or disagreement towards the six domains					Additional comments or suggestions about each domain proposed
	SA 5	A 4	U 3	D 2	SD 1	
Domain 2: Ability to give reasons and evidence to support a claim						
Domain 3: Ability to state a counter-argument and a refutation						
Domain 4:  Ability to write a conclusion						
Domain 5:  Ability to organize ideas in a logical way						
Domain 6:  Ability to use word choices and sentence structures						

### **Part 2: Validation of the descriptions of the six domains of the CTER**

Please rate how strongly you agree *or* disagree upon whether the descriptions of each domain of the assessment are appropriate and clear by putting a check mark (/) in the box. (SA = Strongly Agree; A= Agree; U = Undecided; D = Disagree; SD = Strongly Disagree)

Six domains of critical thinking assessment	The descriptions of each domain of critical thinking assessment	Level of your agreement <i>or</i> disagreement towards the appropriateness of the descriptions of each domain					Additional comments or suggestions about the descriptions of each domain
		SA 5	A 4	U 3	D 2	SD 1	
<b>1) Ability to write an introduction to gain readers' attention</b>	Writes an opening sentence(s) by using a rhetorical strategy for gaining readers' attention clearly, provides clear, detailed background information and states a claim clearly.						
<b>2) Ability to give reasons and evidence to support claim</b>	Gives strong, accurate, relevant reasons with concrete evidence to support claim. Elaborates the reasons and evidence clearly and thoroughly.						
<b>3) Ability to state a counter-argument and a refutation</b>	States a counter-argument and refutes it with reasons and concrete evidence. Elaborates the reasons and evidence clearly and thoroughly.						

Six domains of critical thinking assessment	The descriptions of each critical thinking domain of the assessment	Level of your agreement <i>or</i> disagreement towards the appropriateness of the descriptions of each domain					Additional comments or suggestions about the descriptions of each domain
		SA 5	A 4	U 3	D 2	SD 1	
<b>4) Ability to write a conclusion</b>	Provides a summary or a restatement of a claim and a final, relevant, constructive comment.						
<b>5) Ability to organize ideas in a logical way</b>	Organizes ideas coherently. The essay has clear main parts: an introduction, a body paragraph and a conclusion. Ideas between paragraphs and ideas within each paragraph connect with each other logically. Uses logical connectors to link ideas clearly.						
<b>6) Ability to use word choices and sentence structures effectively</b>	Uses word choices and sentence structures that are varied, appropriate and convey meaning clearly and appropriately. <u>May</u> use a metaphor or a simile to convey ideas clearly.						

### **Part 3: Validation of the specific evaluative descriptors in each performance level of the six domains**

Please rate how strongly you agree *or* disagree upon whether the specific descriptors in each level of performance are appropriate and clear. Please indicate the level of your agreement or disagreement by putting a check mark (/) in the box.

(SA = Strongly Agree; A= Agree; U = Undecided; D = Disagree; SD = Strongly Disagree)

#### **3.1 The first domain of critical thinking assessment (i.e., the ability to write an introduction to gain readers' attention)**

Four rating scales	Specific evaluative descriptors of the first domain of assessment	Level of your agreement or disagreement towards the descriptions in each level of performance					Additional comments or suggestions about the descriptors in each score level
		SA 5	A 4	U 3	D 2	SD 1	
Excellent (4)	-The introduction contains 3 elements: 1) an interesting opening sentence(s), 2) relevant background information and 3) a clear claim. Background information links opening sentence to the claim clearly. Introduction is outstanding and successfully captures readers' interest.						
Good (3)	-The introduction contains 3 elements: 1) good opening sentence(s), 2) relevant background information and 3) a claim.						
Fair (2)	-The introduction contains 2 elements: 1) opening sentence(s) and 2) relevant background information. No claim.						
Poor (1)	-No introduction.						

**3.2 The second domain of critical thinking assessment (i.e., the ability to give reasons and evidence to support claim)**

Four rating scales	Specific evaluative descriptors of the second domain of assessment	Level of your agreement or disagreement towards the descriptions in each level of performance					Additional comments or suggestions about the descriptors in each score level
		SA 5	A 4	U 3	D 2	SD 1	
Excellent (4)	-Gives at least 2 strong, relevant reasons with concrete evidence to support claim. Reasons and evidence are clearly and convincingly elaborated.						
Good (3)	-Gives at least 2 relevant reasons with concrete evidence to support claim. However, the evidence is not elaborated.						
Fair (2)	-Gives at least 2 relevant reasons without evidence.						
Poor (1)	-No relevant reasons and no evidence.						

### 3.3 The third domain of critical thinking assessment (i.e., the ability to state counter-argument and refutation)

Four rating scales	Specific evaluative descriptors of the third domain of assessment	Level of your agreement or disagreement towards the descriptions in each level of performance					Additional comments or suggestions about the descriptors in each score level
		SA 5	A 4	U 3	D 2	SD 1	
Excellent (4)	-States a strong counter-argument and refutes it with much stronger reasons and concrete evidence. Reasons and evidence are clearly and convincingly elaborated.						
Good (3)	-States a strong counter-argument and refutes it with much stronger reasons and evidence. However, the evidence is not elaborated.						
Fair (2)	-States a general counter-argument and refutes it by using general reasons without evidence.						
Poor (1)	-No counter-argument and no refutation -OR, states a strong counter-argument but cannot refute it. This makes a student's own argument less effective and weak.						

**3.4 The fourth domain of critical thinking assessment (i.e., the ability to write a conclusion)**

Four rating scales	Specific evaluative descriptors of the fourth domain of assessment	Level of your agreement or disagreement towards the descriptions in each level of performance					Additional comments or suggestions about the descriptors in each score level
		SA 5	A 4	U 3	D 2	SD 1	
Excellent (4)	-The essay has two elements: 1) a summary AND 2) a final, constructive comment.						
Good (3)	-The essay has two elements: 1) a summary AND 2) a final, general comment. -OR, may have only a final, constructive comment.						
Fair (2)	-The essay has a summary WITHOUT a final comment.						
Poor (1)	-No conclusion.						

**3.5 The fifth domain of critical thinking assessment (i.e., the ability to organize ideas in a logical way)**

Four rating scales	Specific evaluative descriptors of the fifth domain of assessment	Level of your agreement or disagreement towards the descriptions in each level of performance					Additional comments or suggestions about the descriptors in each score level
		SA 5	A 4	U 3	D 2	SD 1	
Excellent (4)	-The essay is well-organized. That is, the essay has clear main parts: an introduction, a body paragraph and a conclusion. Ideas within each paragraph AND between paragraphs connect with each other logically. Ideas flow smoothly and can be followed very easily.						
Good (3)	-The essay is mostly coherent and organized. That is, the essay has an introduction, a body paragraph and a conclusion, and ideas in each part mostly connect with each other. Ideas can be followed easily.						
Fair (2)	-Some parts of the essay are more coherent and organized than the other parts. The essay can be followed with some difficulties.						
Poor (1)	-Ideas lack organization. The connectors do not show a logical connection between ideas and cause misunderstanding.						

**3.6 The sixth domain of critical thinking assessment (i.e., the ability to use word choices and sentence structures)**

Four rating scales	Specific evaluative descriptors of the sixth domain of assessment	Level of your agreement or disagreement towards the descriptions in each level of performance					Additional comments or suggestions about the descriptors in each score level
		SA 5	A 4	U 3	D 2	SD 1	
Excellent (4)	-Uses word choices and sentence structures that are varied, appropriate and clear throughout the essay. May use a metaphor or a simile to convey ideas clearly and appropriately. Word choices are outstanding and enrich an argument.						
Good (3)	-Word choices and sentence structures are mostly appropriate and clear. Word choices enrich an argument.						
Fair (2)	-Word choices and sentence structures are quite clear and demonstrate some variety						
Poor (1)	-Word choices and sentence structures are not varied and not clear.						

**Part 4: Evaluating the descriptors' ability to make a clear distinction among the four score levels**

Please rate how strongly you agree *or* disagree upon whether the descriptors of each domain of the assessment can make a clear distinction among the four score levels. Please indicate the level of your agreement or disagreement with each statement by marking (/) in the box. (SA = Strongly Agree; A = Agree; U = Undecided; D = Disagree; SD = Strongly Disagree)

The six statements about the descriptors' ability to make a clear distinction among the four score levels.	Level of your agreement <i>or</i> disagreement towards the six statements					Additional comments or suggestions about the descriptors of the six domains
	SA 5	A 4	U 3	D 2	SD 1	
1) The descriptions of the first domain of assessment can make a clear distinction among the four score levels						
2) The descriptions of the second domain of assessment can make a clear distinction among the four score levels						
3) The descriptions of the third domain of assessment can make a clear distinction among the four score levels						
4) The descriptions of the fourth domain of assessment can make a clear distinction among the four score levels						
5) The descriptions of the fifth domain of assessment can make a clear distinction among the four score levels						
6) The descriptions of the sixth domain of assessment can make a clear distinction among the four score levels						

Name: \_\_\_\_\_

Signature: \_\_\_\_\_

Date: \_\_\_\_\_

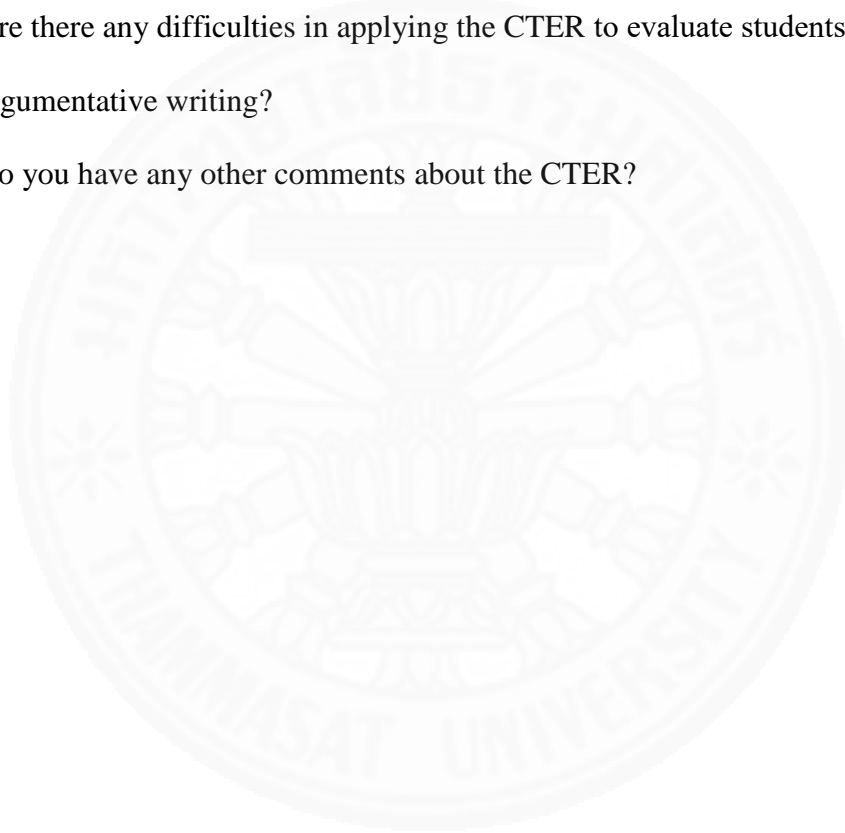
\*\*\*\*\*End of the validation survey\*\*\*\*\*

**Thank you very much for your corporation ☺ ☺ ☺**

## APPENDIX 13

### Interview questions for three raters and three writing teachers

- 1) Overall, what do you think of the CTER? What do you like or dislike about this rubric? Why or why not?
- 2) Overall, do you think the CTER is a usable/practical scoring tool? Why or why not?
- 3) Are there any difficulties in applying the CTER to evaluate students' argumentative writing?
- 4) Do you have any other comments about the CTER?



**APPENDIX 14****Open-ended questionnaire for the writing teachers****Writing teachers' questionnaire asking about the attitudes towards the use of the CTER**

Name \_\_\_\_\_ Years of teaching \_\_\_\_\_

Research Expertise \_\_\_\_\_

1) Overall, do you think the CTER is a suitable tool for assessing critical thinking in EFL students' argumentative writing? Why or why not?

---

---

---

---

---

2) Do you think this rubric can be used for developing students' critical thinking in argumentative writing? How?

---

---

---

---

---

---

3) Do you think the score results from this rubric can be used for planning an instruction?

How?

---

---

---

---

---

---

---

---

4) If possible, would you use the rubric in the future? Why?

---

---

---

---

5) Do you have any other comments about the rubric?

---

---

---

---

---

---

---

---

( \_\_\_\_\_ )

\_\_\_\_\_

Date \_\_\_\_\_

**Thank you very much for your corporation 😊😊😊**

## BIOGRAPHY

**Name** Mr.Nattawut Nakkaew

**Date of Birth** October 4, 1991

**Educational Attainment** 2018: Master of Arts (English Language Studies),  
Thammasat University

2014: Bachelor of Arts (English) (First Class Honors)  
Srinakharinwirot University

**Work Position** Full-time lecturer in the Department of Business  
English, College of International Business,  
Rattana Bundit University, Thailand

### Conference Presentations

Nakkaew, N. (2017). *Developing a powerful model of the teaching writing process along with an 'assessment-for learning' instrument for evaluating students' writing abilities*. Research paper presented at the 16<sup>th</sup> Symposium on Second Language Writing, Chulalongkorn University, Thailand.

Nakkaew, N., & Adunyarittigun, D. (2017). *Initial development of rubric to assess critical thinking in writing*. Research paper presented at the 37<sup>th</sup> Thailand TESOL International Conference. Bangkok, Thailand.

Nakkaew, N., & Adunyarittigun, D. (2018). *Developing a rubric for assessing critical thinking in argumentative writing: Challenges and opportunities for using assessment to promote learning*. Research paper presented at the 5<sup>th</sup> International Conference on FLLT, Chiangmai, Thailand.