

An Initial Development of an Analytic Rubric for Assessing Critical Thinking in English Argumentative Essays of EFL College Students

NATTAWUT NAKKAEW*

Department of English, Faculty of Liberal Arts, Thammasat University, Bangkok, Thailand

DUMRONG ADUNYARITTIGUN

Department of English, Faculty of Liberal Arts, Thammasat University, Bangkok, Thailand

*Corresponding author e-mail: nattawut_bast@hotmail.com

Article Information

Article History:

Received: July 12, 2019

Accepted: November, 17, 2019

Available online:

December 30, 2019

Keywords:

The initial development of rubric

Critical thinking

Argumentative writing

Writing assessment

Abstract

This study aims to initially develop a Critical-Thinking-in-Argumentative-Essay Rubric (CTER) for EFL college students. Participants of this study were five experts and two groups of raters. Data sources included the experts' validation survey for the CTER, interviews and writing samples. Three phases for developing the CTER were conducted, and the evaluative descriptors of the rubric were revised based on the experts' comments. To complete the initial development of the rubric, the raters and the first researcher used the CTER to evaluate the writing samples. The scores obtained from the evaluation were analyzed to examine the inter-rater reliability of the rubric. The findings showed that the CTER contained six clear and valid domains for assessing critical thinking in argumentative essays of EFL students. The total scoring results from the six domains achieved a moderate inter-rater reliability with ICC of 0.70 and Kendall's W of 0.5 ($p < 0.05$). The raters perceived that the CTER could be used to promote learning and critical thinking of EFL learners. Pedagogical implications were presented based on the findings.

INTRODUCTION

Critical thinking is an ability to think clearly and logically (Kirby & Goodpaster, 2007; Moskal & Leydens, 2000) and has been considered as a higher-order thinking skill (Liu & Stapleton, 2018) and a lifelong learning skill (Celuch & Slama, 2010). It includes many crucial subskills, such as analyzing and evaluating information from various perspectives (Chatfield, 2018) and making a logical argument and providing a constructive comment (Lau, 2011).

To write an argumentative essay, students are required to engage in critical thinking. To illustrate, they have to think about propositions and succinct claims, plan and seek information, construct their advocacy carefully, support claims with relevant and compelling evidence, organize arguments in a logical way, and make the most of language to help readers make sense of the meaning conveyed in arguments. Thus, argumentative writing is an effective means for learning how students think and for assessing their critical thinking (Hughes, 2000).



Evaluating critical thinking in writing needs a valid and reliable rubric. A rubric is a scoring tool containing domains of assessment, rating scales and evaluative descriptors (Stevens & Levi, 2005). It helps scorers understand and make distinctions between constructs of critical thinking clearly and accurately (Williams, 1999) and assess all important aspects of critical thinking. An effective rubric needs to have clear domains of assessment, clear rating scales and clear, sufficient and appropriate descriptors in each score level (Moskal & Leydens, 2000). It also needs to provide consistent scoring results.

Existing rubrics for assessing critical thinking in argumentative writing suffer from crucial limitations. For instance, existing holistic rubrics cannot diagnose strengths and weaknesses of each critical thinking element in argumentative writing of EFL students. Also, existing analytic rubrics contained evaluative descriptors that could not make a clear distinction among the score levels. Thus, this study aims to initially develop and validate the CTER—an analytic rubric—for EFL college students. The CTER contains six domains of critical thinking assessment and evaluative descriptors in four levels of performance (See Appendix 1).

LITERATURE REVIEW

A valid and reliable rubric is crucial for critical thinking assessment since it can be used to evaluate critical thinking accurately and provide students with diagnostic data for improving their critical thinking. However, the existing rubrics for assessing critical thinking in English argumentative writing suffer from four striking flaws. Firstly, existing checklist rubrics (Liu, Wu & Shieh, 2015; Liu & Stapleton, 2014; Stapleton, 2001) do not provide teachers or scorers with diagnostic information about EFL students' varying levels of performance, especially in constructing a strong argument. For example, the rubrics developed by Stapleton (2001) and by Liu and Stapleton (2014) give a number or frequency of each critical thinking element in argumentative writing (e.g., argument, evidence, counter-argument). Yet, it is apparent that the degree of quality of these elements is not seriously taken into consideration. This affects the validity of these rubrics (Green & Hawkey, 2002).

Secondly, existing holistic rubrics (e.g. Angeli & Valanides, 2009; Welch, Hieb, & Graham, 2015) cannot provide an accurate picture of a student's critical thinking ability in argumentative essays. It is the nature of a holistic rubric to grant a single score to illustrate a comprehensive picture of a student's critical thinking ability which includes such important subskills as the abilities to state a claim, provide reasons and evidence, refute an opposing view, organize ideas logically and use effective language to convey meanings. It could be inferred from a single score that certain critical thinking subskills of the student are developed at the same rate (Saxton, Belanger & Backer, 2012). In fact, the student may be able to state a claim and give very strong evidence to support his or her claim but might not be able to state a counter-argument or refutation in his or her essay (Saxton, Belanger & Backer, 2012). Therefore, generalizing all performance levels of all critical thinking skills to a single score cannot provide useful, diagnostic feedback to be used for helping EFL students improve specific weaknesses and support their strengths in critical thinking (Saxton, Belanger & Backer, 2012).

Thirdly, existing analytic rubrics (e.g., Daud, 2012; Dong, 2015) contain descriptors that cannot make clear distinctions among score levels. For instance, in a rubric developed by Saxton, Belanger and Becker (2012), descriptors in the domain of 'Inference' are subjective and hard to make distinctions from one score level to another. The descriptors include subjective adjectives, such as 'surprising', 'insightful' and 'strong' as the followings.

Score level 6: “demonstrates surprising/insightful ability to take concepts further”

Score level 5: “demonstrates strong ability to take concepts further”

(Saxton, Belanger & Becker, 2012, p.268).

To make this rubric clearer, the subjective adjectives should be removed or revised, and the number of reasons or evidence needs to be quantified to make a clear distinction among the score levels. Lacking clear and effective descriptors contributes to the threat to the validity of the rubrics (Weigle, 2002).

Fourthly, existing analytic rubrics (e.g., Liu & Stapleton, 2014, 2018; McLendon, 2008; McKittrick & Barnes, 2012; Mulnix & Mulnix, 2010; Stapleton & Wu, 2015) fail to include some crucial, relevant aspects of critical thinking to be measured in argumentative writing, such as writing an introduction to gain readers' attention and providing constructive comments. To achieve validity, a rubric needs to include all of the important aspects of critical thinking (Moskal & Leydens, 2000). Otherwise, this will lead to one major threat to the validity, Construct Underrepresentation (CU) (Messick, 1996).

Therefore, there is a need to develop a more valid and more reliable rubric used to assess EFL college-aged students' critical thinking in argumentative writing. The objective of this study is to develop a valid and reliable rubric for evaluating EFL college-aged students' critical thinking in argumentative essays. There are two research questions: 1) Is the CTER a valid and reliable rubric? and 2) What are the attitudes of the raters towards the CTER?

METHODS

Research design

This study used quantitative and qualitative methods in the process of developing and validating the CTER (Johnson & Onwuegbuzie, 2004). The researchers collected and analyzed data and integrated the findings through using both methods (Johnson & Onwuegbuzie, 2004). This strengthened the research study in terms of data triangulation (Denzin, 1978) and enhanced the credibility of the findings (Greene, Caracelli & Graham, 1989).



Participants

1. Experts

- **Three experts in the first level of validation**

These experts (Experts A, B and C) were university professors from two universities in Thailand. Expert A had expertise in critical thinking, Expert B in L2 writing, and Expert C in language assessment. The three experts were asked to validate the CTER and provide constructive comments for the initial version of the rubric.

- **Five experts in the second level of validation**

The three experts from the first level of validation and two professors from a university in Thailand having expertise in L2 writing were asked to validate the revised version of the CTER.

2. Raters

- **Three raters in the first tryout of the CTER**

Three doctoral students (Raters A, B and C) were asked to be raters in the first tryout. Raters A and B were doctoral students in a university in Thailand. They did research about language assessment and had expertise in using rubrics to assess English writing of EFL students. Rater C was also a doctoral student at a Chinese university with expertise in analyzing discourse and argumentation. The three raters were asked to try using the CTER to grade some of the writing samples and share their opinions about validity and practicality of the rubric. Their comments about the CTER would be considered for revising the rubric before it was implemented in the second tryout.

- **Three raters in the second tryout of the CTER**

Three instructors from three universities in Thailand were asked to be raters in this tryout. These raters (Raters D, E and F) had at least 10 years of teaching experience and were familiar with teaching and evaluating EFL students' writing. They were asked to implement the final version of the CTER to evaluate the writing samples and share their opinions about the validity and practicality of the rubric. Their scoring results were analyzed to examine the inter-rater reliability of the CTER.

It is important to note that the researchers' roles were to develop the CTER, investigate the validity and inter-rater reliability of the CTER, and examine the raters' opinions about using the CTER to evaluate the writing samples. It is with this in mind that the researchers would subsequently revise and edit the CTER to improve the quality of the rubric.

Instruments

1. The experts' survey for validating the CTER

This survey was developed from existing surveys used in previous studies (e.g., Castilleja, 2012; Moskal & Leydens, 2002). It contained four parts and used a five point Likert Scale (5=Strongly agree to 1=Strongly disagree). In part 1, the five experts were asked to rate how strongly they agreed that the six domains of the CTER were valid and clear indicators of critical thinking in argumentative essays of EFL students. In part 2, they were asked to rate how strongly they agreed that descriptions of each type of abilities were valid and clear. In part 3, they were asked to rate how strongly they agreed that specific descriptors in each rating scale were valid and clear. In part 4, they were asked to rate how strongly they agreed that the descriptors could make a clear distinction among the four score levels. The five experts were asked to complete the survey in the second level of validation.

2. Free interview for the experts

Individual, face-to-face, free interviews were conducted with the five experts to obtain their opinions about the CTER and their suggestions for revising the rubric. It took around 30 minutes to interview each expert. The interview was conducted in Thai in order to facilitate their discussion and was transcribed and translated into English later.

3. The writing samples

Forty-one argumentative essays written by Thai EFL college students were used as the writing samples for the two tryouts of the CTER. The topics of their writing varied and could be categorized into six issues: technology, ethics, health, education, media and others. The length of each essay was approximately 600-700 words. The writing samples were used for the two tryouts by the raters.

In the first tryout, fifteen essays were randomly selected from the pool. Five selections were used for training Raters A, B and C how to use the CTER to evaluate the essays. After the raters were competent and confident of using the CTER, they were asked to evaluate 10 selected essays. The first tryout was conducted with the purpose of validating the CTER. The raters tried using the rubric to mark the writing samples. Their comments would be taken into consideration for revising the rubric before it was implemented in the second tryout.

In the second tryout, twenty-six essays were randomly selected from the pool. Raters D, E and F were asked to use the CTER to grade the essays. Six essays were used for the training, whereas twenty essays were used for implementing the rubric. The second tryout was conducted with the purpose of investigating the raters' attitudes towards the validity and practicality of the CTER. The scores obtained from their evaluation were also used to examine the inter-rater reliability of the CTER.

4. Semi-structured interview questions for the raters

The interview questions adapted from Zhao (2010) were used to investigate the raters' attitudes towards the use of the CTER for assessing critical thinking in argumentative essays. In the present study, the attitudes were defined as "an evaluative reaction" to the rubric based on the participants' "beliefs or opinions" (Gardner, 1985, p.9). Four questions (See Appendix 2) were used as a guideline for the interview. It took approximately 30 minutes to interview each participant. The interview was conducted in Thai in order to facilitate their discussion and was transcribed and translated into English later.

Research procedures

1. The process of developing and validating the CTER

There were three phases for developing and validating the CTER (See Figure 1). These phases were adapted from Daud's study (2012).

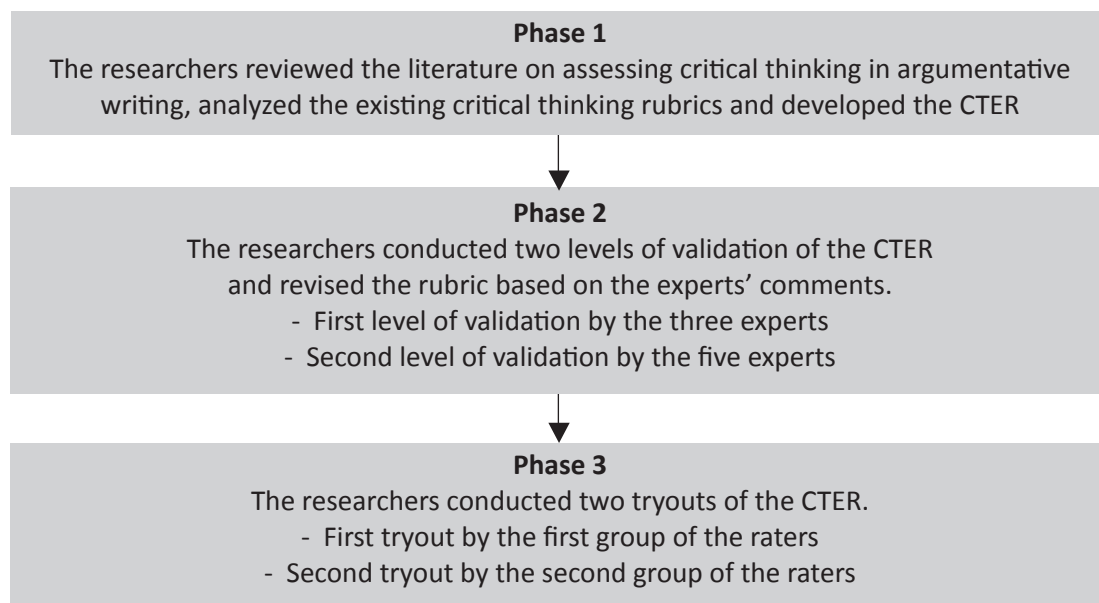


Figure 1 Three phases for developing and validating the CTER

Literature in the fields of critical thinking, argumentative writing, language assessment, and assessing critical thinking in argumentative writing was extensively reviewed. Also, more than 40 rubrics for evaluating critical thinking were reviewed and analyzed, and their strengths and weaknesses were identified. The researchers then developed the initial version of the CTER based on the review and analysis.

The CTER contains six crucial aspects of critical thinking in argumentative essays:

1. Writing an introduction to gain readers' attention
2. Giving reasons and evidence
3. Stating a counter-argument and a refutation
4. Writing a conclusion
5. Organizing ideas logically
6. Using effective language to convey thoughts.

Descriptions of the six domains are presented in Appendix 3. The CTER was designed to be an analytic rubric, so it can specifically evaluate each domain of the assessment and offer detailed, diagnostic feedback for each domain.

- **Validating the CTER**

The experts from the two levels of validation validated the CTER by reviewing whether the six types of abilities in the CTER and the evaluative descriptors in each score level were clear, accurate and appropriate. The descriptors were revised based on the experts' comments.

- **Training the raters in the first tryout and their implementation of the CTER**

The first group of the raters received group training on how to use the CTER (See Appendix 1). The training aimed to ensure clear understanding of how to use the rubric. In the training, Raters A, B and C discussed the rubric with the first researcher and were asked to review anchor papers developed by the two researchers. The anchor papers served as a benchmark illustrating each score level of the rubric. Thus, after carefully reading and discussing the anchor papers in each performance level with one another, the three raters could develop clear understanding about the scoring rubric. Subsequently, both the raters and the first researcher tried using the rubric to grade 5 randomly selected essays and discussed the scoring results altogether. When the ratings had more than a 1-point difference, they would discuss until a consensus was reached (Zhao, 2010). This enabled the raters to mark the essays more consistently. Afterwards, the raters were asked to implement the rubric by grading 10 randomly selected essays independently and discussing the scoring results with the first researcher in the training session.

After the group training, the first group of the raters were provided with “the flexibility and time” to carefully review the rubric (Tulgar, Yagiz, & Han, 2017, p. 65) and to reflect on how they felt about the rubric. Then, the first researcher conducted individual, face-to-face interviews with the raters the day after to investigate their opinions about the validity and practicality of the CTER.

- **Training the raters in the second tryout and their implementation of the CTER**

The second group of the raters received individual training on how to use the CTER. It is important to note that the training was done individually to facilitate these raters because they had urgent work and could not come and receive the group training together. In this training, Raters D, E and F discussed the CTER with the first researcher and used it to grade



6 randomly selected essays, compared, contrasted and discussed the scoring results with the researcher. This enabled the raters to have a clear understanding about using the CTER to mark the essays. After the training, the raters were asked to implement the CTER by marking 20 randomly selected essays independently and submitting the scores within three weeks. To examine an inter-rater reliability of a rubric, it was acceptable to use this number of essays (Qin, 2009; Rowicki, 2001). The raters were also provided with the descriptions of the CTER (See Appendix 3) and the anchor papers so that they could understand the scoring criteria clearly. The scores obtained from their evaluation were analyzed to examine the inter-rater reliability of the CTER. The second group of the raters were also provided with “the flexibility and time” to carefully review the rubric in the same procedure as that of the first group (Tulgar, Yagiz, & Han, 2017, p. 65). After submitting their scores, Raters D, E and F were interviewed in order to examine their attitudes towards the use of the CTER.

Data analysis

To investigate the validity and inter-rater reliability of the CTER, the data analyses were conducted. The quantitative data from the experts’ validation survey were analyzed using SPSS to compute mean (\bar{x}) and standard deviation (S.D.). To maintain the validity of the domains and descriptors of the CTER, the confidence level of means had to meet or exceed 3.50; domains or descriptors not meeting the criterion had to be revised based on the experts’ comments (Adunyarittigun, 2015). To investigate the inter-rater reliability of the CTER, scoring results from the first researcher and the three raters in the second tryout were examined by using two statistical approaches: Kendall’s coefficient of concordance (Kendall’s W) and an Intra-class correlation coefficient (ICC). The Kendall’s W criterion for judging the inter-rater reliability of the CTER was: 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). The ICC criterion for judging the inter-rater reliability of the CTER was: less than 0.5=poor, 0.5-0.75=moderate, 0.75-0.90=good, and greater than 0.90=excellent (Koo & Li, 2016). The data from ICC were used to triangulate with those from Kendall’s W to confirm the consistency of the findings.

The qualitative data from the interviews were analyzed using a thematic analysis (Merriam, 1998). The analysis was used to support the validity and reliability of the rubric and to identify the raters’ attitudes towards the CTER. The interviews were transcribed and translated from Thai into English. Some of the transcripts were randomly selected. The interpretation of the selection was verified and checked for accuracy by the raters.

RESULTS AND DISCUSSION

Research question 1: Is the CTER a valid and reliable rubric?

- **Findings from the first level of validation**

The results from the qualitative data analysis demonstrated that the three experts strongly agreed that the CTER included all crucial, relevant aspects of critical thinking in English argumentative essays of EFL college students. These experts also agreed that the descriptors

in each rating scale were comprehensive, accurate and appropriate. To make the CTER more valid and clearer, some descriptors were revised based on the experts' comments. Also, the number of reasons was clearly quantified. Therefore, the descriptors could make a clear distinction among the score levels.

- **Findings from the second level of validation**

The results from the quantitative data analysis indicated that the CTER was valid. It contained six clear and valid indicators of critical thinking in EFL students' argumentative essays: 1) writing an introduction to gain readers' attention ($\bar{x} = 5.00$, S.D. = 0.00), 2) giving reasons and evidence to support a claim ($\bar{x} = 5.00$, S.D. = 0.00), 3) stating a counter-argument and a refutation ($\bar{x} = 4.80$, S.D. = 0.45), 4) writing a conclusion ($\bar{x} = 5.00$, S.D. = 0.00), 5) organizing ideas logically ($\bar{x} = 5.00$, S.D. = 0.00), and 6) using word choices and sentence structures effectively ($\bar{x} = 5.00$, S.D. = 0.00). Descriptions of these domains were valid and clear ($\bar{x} = 4.63$, S.D. = 0.40). In addition, the descriptors in each score level of the six domains were valid and clear with means ranging from 3.60 (S.D. = 1.14) to 4.75 (S.D. = 0.50) and could make clear distinctions among the score levels ($\bar{x} = 4.00$, S.D. = 0.72).

Careful revisions of the descriptors based on the experts' comments enhanced the validity of the CTER. Such revisions included the followings. First, important terminology in the rubric was clearly defined and elaborated. For example, the word 'reason' was defined as logical explanation that supported a claim, while the meaning of the word 'evidence' was made clearer by including the following: first or second-hand experience, research findings, personal observations, opinions of experts in the field, interviews, personal narratives and so on. Second, more descriptors were added to make the descriptors capture what EFL students write in their argumentative essays. For instance, the descriptors in Level Fair of the second domain of the CTER, after being revised, could capture all possibilities of the responses as follows (Green & Hawkey, 2002):

- 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.

Third, the number of counter-arguments was quantified in each performance level of the CTER (See Appendix 1). Thus, the descriptors can make a clear distinction among the four score levels.

Moreover, analysis of the qualitative data obtained from the two groups of the raters support that the CTER is valid because it covers all crucial aspects of critical thinking in argumentative writing and contains clear, detailed evaluative descriptors.

The six domains of the assessment are clear. The rubric can assess critical thinking in argumentative writing. (Rater A, Interview)



I think the scoring criteria are clear and can help me assign a score to a text accurately. (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. (Rater D, Interview)

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. (Rater E, 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. (Rater F, Interview)

Contrary to the results of several studies which proposed checklist or holistic critical thinking rubrics (e.g., Angeli & Valanides, 2009; Stapleton, 2001), the present study initially developed and validated the CTER—an analytic rubric. Both quantitative and qualitative data analyses demonstrated that the CTER is valid since it provides scorers with detailed evaluative information. It includes valid and clear indicators of critical thinking in argumentative essays of EFL students and has descriptors that can make clear distinctions among the score levels.

Two important factors account for the validity of the CTER. First, the rubric has clear operational definitions of critical thinking—the six critical thinking abilities in argumentative writing. These abilities are developed from extensively reviewing and examining the literature related to critical thinking, argumentative writing, language assessment and existing rubrics for assessing critical thinking in argumentative writing. Second, the CTER has clear rating scales and descriptors that are specific, comprehensive, concrete, relevant and appropriate for each domain of the assessment (Weigle, 2002; Zhao, 2010). The CTER was validated thoroughly by the experts and was revised carefully and continuously by the two researchers under the experts' supervision. This made the descriptors of the rubric illustrate a clear distinction among the score levels (Hodges et al., 2019). When the raters used the CTER to evaluate students' essays, they could clearly identify different levels of critical thinking in the writing of each student.

Table 1
The inter-rater reliability of the CTER

Critical thinking domains	Kendall's W values	ICC values
Domain 1	0.5*	
Domain 2	0.3	
Domain 3	0.4*	
Domain 4	0.5*	
Domain 5	0.4*	
Domain 6	0.5*	
Total scores from the six domains	0.5*	0.70**

*p < 0.05

**p < 0.01

According to the above table, Kendall's W values indicated that domains 1, 3, 4, 5 and 6 and the total scoring results from the six domains achieve the moderate inter-rater reliability (Kendall's W = 0.4 - 0.5) which is statistically significant at the 0.05 level, whereas domain 2 exhibits a weak inter-rater reliability (Kendall's W = 0.3) that is not statistically significant at the 0.05 level. ICC values also showed that the total scoring results from the six domains achieve the moderate inter-rater reliability (ICC = 0.7) which is statistically significant at the 0.01 level. Thus, it can be concluded that the CTER achieves the moderate inter-rater reliability. The CTER is reliable (Kendall's W = 0.3 - 0.5; ICC = 0.70) because of two important reasons. First, the clarity of the CTER reduces the raters' subjectivity and promotes consistency in the scoring. It is undeniable that an unclear rubric can cause raters to have different perceptions about good or bad writing and different understanding about criteria when grading an essay (Plakans & Gebril, 2017), leading to inconsistency in grading essays (Quintieri, 2005). In addition, Janssen and her colleagues' finding (2015) shows that it is hard to make a clear distinction among the four score levels—especially "in the middle levels" (p.64). Yet, the CTER has clear operational definitions of critical thinking and has specific, concrete and detailed descriptors. This helped the raters develop clear understanding about the assessment and enabled them to mark the essays more reliably (Gebril & Plakans, 2014).

Second, providing the raters with careful training on how to utilize the CTER potentially helped them grade the essays consistently. In the training of this study, the raters discussed the rubric with the first researcher, practiced using it to grade the writing samples, discussed the scoring results with the first researcher and received constructive comments about their grading. Thus, training is a key factor in making the raters and the researcher understand the rubric clearly and grade the writing samples consistently (Gebril & Plakan, 2014), leading to the effectiveness of the rubric.

It is important to note that the inter-rater reliability of the second domain of the assessment is weak (Kendall's W = 0.3) and is not statistically significant at the 0.05 level. It is possible that the raters in this study could have interpreted the descriptors of this domain differently due



to the fact that they received the individual training. At that time, Raters D, E and F were unavailable to come and obtain the training altogether; therefore, they were trained in different sessions at a different period of time, based on their convenience. Thus, although the raters discussed the rubric with the first researcher, they did not have an opportunity to discuss the scoring criteria with one another. The second domain of the CTER contains many descriptors that require careful consideration and interpretation from the scorers. Examples of such descriptors include “reasons and evidence may be fallacious, too weak or too limited.” Thus, without the group training, the three raters might have interpreted these descriptors differently. 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 raters need to be trained together. For example, they need to discuss the scoring criteria together, grade essays, compare, contrast and discuss the scoring results with each other and justify their grading by referring to the descriptors in order to reach an agreement (Weigle, 2002).

Research question 2: What are the attitudes of the raters towards the CTER?

Analysis of the qualitative data revealed three emerging themes that show the participants’ positive attitudes towards the CTER.

- a. The CTER can be further improved, especially in the aspect of being time-consuming

The raters thought that the CTER contains many detailed descriptors, so using it to grade many students’ argumentative essays demands lots of time and effort from scorers. This aspect of the study resonates with previous research (e.g., Han & Huang, 2017) which found that an analytic rubric may be impractical if it requires lots of time and energy from raters. However, raters 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 they can see the evaluative descriptors clearly and grade the essays faster. Also, providing more frequent and effective training to scorers will familiarize scorers with the rubric and help them grade essays faster (Chinda, 2014). Chinda (2013) argued that although it is time-consuming to use the analytic rubric, it is worthwhile and helpful for teachers to utilize it since it can promote Thai EFL writers’ learning to write effectively.

- b. Teachers can use the CTER as assessment for learning (Shepard, 2000) to promote learning and critical thinking of EFL students.

The participants believed that writing teachers can use the CTER as an instructional tool for helping EFL students understand the six domains of critical thinking and skillfully write argumentative essays including these elements. To illustrate, the six domains of the CTER are rhetorical and cognitive strategies for English argumentation. Also, there is clear, detailed evaluative information in each level of performance. Thus, this rubric serves as a “writing development model” (Bailey et al., 2015, p.525) that helps students learn how to develop an argument. If teachers use the CTER to introduce students to the crucial elements in argumentative writing (e.g., claims, reasons and evidence) 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 and write an argument more efficiently (Reznitskaya et al., 2009, p.219).

Moreover, the participants believed that teachers can encourage EFL students to use the CTER to do self-and peer- assessments which will help them improve their critical thinking and become aware of their own learning, strengths and weaknesses and areas for improvement. To illustrate, when learners use the CTER to evaluate their own or their peers’ essays, they need to review the rubric carefully and evaluate the quality levels of each domain of the assessment judiciously. Also, they can use the CTER as a guide that assists them in giving feedback to peers and challenge peers to think more critically about their own writing. Tangkiengsirisin (2014) found that in the revision of writing, many EFL students were not aware of crucial skills, such as writing an introduction strategically, writing a claim clearly, writing a constructive comment and using connectors to link ideas logically. Importantly, the CTER provides useful, diagnostic feedback about critical thinking features in writing, so students can use the rubric to analyze their levels of critical thinking abilities and learn the crucial skills from the rubric. Also, they can use feedback from the rubric to revise their own essays and provide peers with useful comments about writing. In sum, clear, detailed, diagnostic information is a crucial part of an “effective formative assessment” (Katz, 2012, p.72); therefore, it can be said that the CTER is an assessment-for-learning instrument that can foster critical thinking of the students.

- c. The CTER can help teachers plan and improve their writing instruction and provide appropriate instruction that best serves students’ needs.

The participants believed that the CTER provides teachers with scoring results that can identify students’ levels of critical thinking abilities, including their strengths and weaknesses, so teachers can use these to plan and improve teaching or learning activities that are appropriate for the needs and performance levels of the students. For example, they can use the rubric to assess students’ critical thinking before, during and after they write argumentative essays to diagnose their abilities and monitor their development (Saxton et al., 2002). If students in a class have different levels of critical thinking abilities in argumentative writing, teachers can have them do group activities, such as group 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 teach reasoning skills that are appropriate for the students’ performance levels at hand. For example, they can teach students to use personal opinions and logical explanations as evidence to support a claim. After the students master these techniques, teachers can teach other kinds of evidence, such as research results and opinions of experts in the field, and encourage learners to use various types of evidence. If most students are proficient learners, teachers can teach them all critical thinking abilities which are identified in the CTER and challenge them to improve their writing in order to reach the score level of Excellent in all domains of the assessment.



IMPLICATIONS

The findings from this study suggest a number of implications. First, teachers should encourage EFL students to use the CTER as an initial tool to assess and improve their own and their peers' argumentative writing in the process of writing. As mentioned earlier, the CTER contains clear domains of the assessment, clear rating scales and clear descriptors. Hence, 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. A writing process is "a recursive, dynamic and exploratory process in which writers search for their ideas, evaluate them and develop their arguments" (Yu, 2008, p.iii). Thus, incorporating the CTER in the writing activities will help support students to think about, share and write their ideas more clearly. Also, they will be able to evaluate their arguments, obtain useful, diagnostic feedback from the rubric, learn their own and their peers' strengths and weaknesses in writing, make self-improvements and share constructive feedback about writing to their peers.

Second, EFL teachers should train students to use the CTER to assess critical thinking in writing, so they can do the self-and peer-assessment effectively. The trainings can include explaining crucial elements of critical thinking in writing to the students and demonstrating how to use the CTER to grade the essays. Teachers need to go through the scoring with the students to gain their feedback about their scoring and provide the learners with constructive comments or guidance in order to help them do the self-and peer-assessment effectively (Ross, 2006). For example, in the training, teachers can have students review the anchor papers—a benchmark representing each score level of the CTER, so they can clearly understand how to mark the essays and evaluate their levels of critical thinking abilities accurately. This enables learners to adhere to the evaluative criteria and avoid over-estimating or under-estimating their own and their peers' abilities, leading to the effective use of the assessment instrument (Anderson, 2012).

Third, EFL teachers can incorporate the CTER into their writing instruction in order to facilitate EFL students to learn how to write effective argumentative essays. The findings showed that the CTER is valid and reliable and has clear, comprehensive descriptions about critical thinking features in argumentative writing. Thus, teachers can use the rubric as a supplementary tool for teaching argumentative writing to promote students' thinking and writing abilities. For example, teachers can elucidate the crucial elements of critical thinking, such as claims, reasons, evidence, counter-arguments, refutations, constructive comments and effective language use in English argumentation. This will help EFL students see a clearer picture about how to write an argument and be able to include these crucial features in their argumentative essays 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 argument (Newell et al., 2011, p.295);
- 2) realize that "Thai and English compositions have different styles and organizations" (Dhanarattigannon, 1995, p.70);
- 3) practice thinking and writing in English clearly (Oshima, & Hogue, 2006);

- 4) improve communicative competence in English writing (Matsuda & Hammill, 2014);
- 5) develop creative ideas in writing, such as using metaphors or similes to make arguments more interesting (Thongrin, 2002);
- 6) express voice or opinion in writing clearly (Zhao, 2010);
- 7) be aware of audience by answering all potential questions in the minds of readers (Malcolm, 2011); and
- 8) demonstrate knowledge about a rhetorical convention of English argumentative writing (Tapinta, 2006).

CONCLUSIONS

This research aimed to initially develop the CTER for EFL college students. The analysis of both quantitative and qualitative data demonstrated that this rubric contains six clear and valid indicators of critical thinking in English argumentative essays and clear, evaluative descriptors in the four levels of performance. The total scoring results from the six domains of the CTER achieved the moderate inter-rater reliability which was statistically significant ($p < 0.05$). In addition, the raters expressed the unanimous agreement that the CTER could be used to promote learning and critical thinking of EFL learners.

Although all research questions have been addressed, two limitations exist in this study. First, the writing samples were limited in terms of the number and the representation of the population—EFL college students. To gain the optimal inter-rater reliability of the CTER, a large number of writing samples should be utilized. Unfortunately, only 20 writing samples were utilized in the second tryout. The researchers were aware that the sample size was rather small and had to admit that an uncontrollable factor took place during the validation process. The raters in the second tryout happened to have so much work that they could not handle with marking a lot of essays at the time that the validation took place. They kindly agreed to help use the CTER to mark 20 essays. Also, the writing samples were obtained from only Thai EFL college students. Thus, the findings may not be generalizable to other groups of the population in different countries. To investigate the inter-rater reliability of the rubric, future studies should include writing samples that have a larger sample size and represent other groups of EFL students.

Second, conducting training for the raters on how to implement the rubric individually is very likely to affect the inter-rater reliability of the rubric. During the study, the raters in the second tryout happened to have so much work and could not come and obtain group training together. If they had received the group training and discussed the CTER and their scoring results altogether, the inter-rater reliability of the rubric could have been higher. Thus, future researchers should provide the raters with effective group training and have them discuss their scoring results in the training session thoroughly, so they will be aware of their strengths or problems in the scoring, improve their grading and mark the essays more reliably.



ACKNOWLEDGEMENTS

The authors would like to thank all of the students who allowed the authors to use their writing samples and the participants in this study for their cooperation. Appreciation is also extended to the anonymous reviewers for their constructive comments on the earlier draft of this manuscript.

THE AUTHORS

Nattawut Nakkaew received his BA in English (first class honors) from Srinakharinwirot University and his MA in English Language Studies from Thammasat University. Currently, he is a lecturer in the Department of Business English, College of International Business, Rattana Bundit University. His research interests include English writing, critical thinking and language assessment.
nattawut_bast@hotmail.com.

Dumrong Adunyarittigun is an associate professor in the Department of English and Linguistics, Faculty of Liberal Arts, Thammasat University in Thailand. His research interests include comprehension, self-perception and motivation to read, language assessment and critical literacy to promote peace.
dumrong.a@arts.tu.ac.th.

REFERENCES

- Adunyarittigun, D. (2015). Developing a scale to measure reader self-perception for EFL students *PASAA*, 50, 1-30.
- 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 United Kingdom: Cambridge University Press.
- Angeli, C., & Valanides, N. (2009). Instructional effects on critical thinking: Performance on ill- defined issues. *Learning and Instruction*, 19, 322-334.
- 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.
- 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, Wisconsin, U.S.A.
- 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.
- 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, Florida, United States of America.
- 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.
- Chatfield, T. (2018). *Critical thinking: Your guide to effective argument, successful analysis and independent study*. London, United Kingdom: Sage.
- Cheng, F. (2003). *The effects of rhetorical specification in writing assignments on EFL (English as a foreign language) writing* (Unpublished doctoral dissertation). Boston University, Boston, U.S.A.

- 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.
- Daud, N. S. B. M. (2012). *Developing critical thinking skills in tertiary academic writing through the use of an instructional rubric for peer evaluation* (Unpublished doctoral dissertation). The University of Canterbury, Christchurch, New Zealand.
- Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods* (2nd ed.). New York, NY: McGraw-Hill.
- Dhanarattigannon, J. (1995). Journal writing: Language activity and teaching/learning evaluation. *Humanities Journal*, 3(1), 69-80.
- Dong, Y. (2015). *Critical thinking in second language writing: Concept, theory, and pedagogy* (Unpublished doctoral dissertation). The University of British Columbia, Vancouver, Canada.
- Epstein, R. L. (1999). *Critical thinking*. Belmont, CA: Wadsworth.
- Gardner, R. C. (1985). *Social psychology and language learning: The role of attitudes and motivation*. London, United Kingdom: Edward Arnold.
- Gebriel, A., & Plakans, L. (2014). Assembling validity evidence for assessing academic writing: Rater reactions to integrated tasks. *Assessing Writing*, 21, 56-73.
- 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. A. Coombe, P. Davidson, B. O'Sullivan & S. Stoyhoff (Eds.), *The Cambridge guide to second language assessment* (pp. 299-306). Cambridge, UK: 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.
- Han, T., & Huang, J. (2017). Examining the impact of scoring methods on the instructional EFL writing assessment: A Turkish perspective. *PASAA*, 53, 112-147.
- 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.
- 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, Virginia, U.S.A.
- Janssen, G., Meier, V., & Trace, J. (2015). Building a better rubric: Mixed methods rubric revision. *Assessing Writing*, 26, 51-66.
- Johnson, B., & Onwuegbuzie, A. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14-26.
- Kalchayanant, T. (2009). *Writing essays*. Bangkok, Thailand: Thammasat University Press.
- Katz, A. (2012). Linking assessment with instructional aims and learning. In C. A. Coombe, P. Davidson, B. O'Sullivan & S. Stoyhoff (Eds.), *The Cambridge guide to second language assessment* (pp. 66-73). Cambridge, UK: Cambridge University Press.
- Kirby, G. R., & Goodpaster, J. R. (2007). *Thinking: An interdisciplinary approach to critical and creative thought* (4th ed.). Cranbury, NJ: Pearson Education, Inc.
- 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.
- Lau, J. Y. F. (2011). *An introduction to critical thinking and creativity: Think more, think better*. Hoboken, NJ: John Wiley & Sons, Inc.
- 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, 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.
- Manalo, E., & Sheppard, C. (2016). How might language affect critical thinking performance? *Thinking Skills and Creativity, 21*, 41–49.
- 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, United Kingdom: Oxford University Press.
- Maybin, J., & Swann, J. (Eds.) (2006). *The art of English: Everyday creativity*. New York, NY: Palgrave Macmillan.
- 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.
- McKittrick, 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, Alabama, U.S.A.
- Merriam, S. B. (1998). *Qualitative research and case study applications in education*. San Francisco, CA: Jossey-Bass Publishers.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13*(3), 241-256.
- Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research & Evaluation, 7*(10), 1-10.
- Mulnix, J. W., & Mulnix, M. J. (2010). Using a writing portfolio project to teach critical thinking skills. *Teaching Philosophy, 33*(1), 27-54.
- 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.
- Nussbaum, E. M., & Schraw, G. (2007). Promoting argument–counterargument integration in students' writing. *The Journal of Experimental Education, 76*(1), 59-92.
- Oshima, A., & Hogue, A. (2006). *Writing academic English* (4th ed.). New York, NY: Pearson Education.
- Payutto, P. A. (2010). *The language and the truth*. Bangkok, Thailand: Chandpen Publishing.
- Plakans, L., & Gebriel, A. (2017). An assessment perspective on argumentation in writing. *Journal of Second Language Writing, 36*, 85-86.
- Qin, J. (2009). *The analysis of Toulmin elements and use of sources in Chinese university EFL argumentative writing* (Unpublished doctoral dissertation). Northern Arizona University, Arizona, U.S.A.
- Quintieri, M. (2005). *Measuring rater consistency: An investigation into the effects of two testing instruments on raters' scores* (Unpublished master's thesis). Concordia University, Montreal, Canada.
- 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.
- 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.
- 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, Alabama, U.S.A.



- 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.
- Schmidt, R. C. (1997). Managing Delphi surveys using nonparametric statistical techniques. *Decision Sciences, 28*(3), 763-774.
- Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4-14.
- Stapleton, P. (2001). Assessing critical thinking in the writing of Japanese university students: Insights about assumptions and content familiarity. *Written Communication, 18*(4), 506-548.
- 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.
- Stevens, D. D., & Levi, A. (2005). *Introduction to rubric: An assessment tool to save grading time, convey effective feedback, and promote student learning*. Herndon, VA: Stylus Publishing.
- 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, Texas, U.S.A.
- Thongrin, S. (2002). *E-mail peer response in collectivist Thai culture: Task, social, and cultural dimensions* (Unpublished doctoral dissertation). Indiana University of Pennsylvania, Pennsylvania, U.S.A.
- Tittle, P. (2011). *Critical thinking: An appeal to reason*. New York, NY: Routledge.
- Tulgar, A. T., Yagiz, O., & Han, T. (2017). Bridging the gap between linguistic and pragmatic assessment in Turkish contexts. *PASAA, 54*, 58-81.
- Verlinden, J. (2005). *Critical thinking and everyday argument*. Boston, MA: Wadsworth, Cengage Learning.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge, United Kingdom: Cambridge University Press.
- 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.
- Williams, R. L. (1999). Operational definitions and assessment of higher-order cognitive constructs. *Educational Psychology Review, 11*(4), 411-427.
- Yu, H. J. (2008). *Interactional structure in the writing process: A comparison of three ESL writing classes* (Unpublished doctoral dissertation). Arizona State University, Arizona, U.S.A.
- Zhao, C. G. (2010). *The role of voice in high-stakes second language writing assessment* (Unpublished doctoral dissertation). New York University, New York, U.S.A.



Appendix 1

The CTER

Critical thinking domains	Levels of performance			
	Excellent (4)	Good (3)	Fair (2)	Poor (1)
1) Ability to write an intro to gain readers' attention	-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.	-Writes an unclear and irrelevant introduction.
2) Ability to give reasons and evidence to support a claim	-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 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.

			<p>reasons which have different levels of quality.</p> <p>-OR the number of good reasons equals the number of poor reasons.</p>	
Critical thinking domains	Excellent (4)	Good (3)	Fair (2)	Poor (1)
3) Ability to state a counter-argument and a refutation	-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 counter-argument.	-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.
4) Ability to write a conclusion	-Gives a summary or a restatement of a claim AND writes a final, relevant, 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.	-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.
5) Ability to organize ideas in a	-Ideas are carefully and very well organized, so ideas within each	-Overall, the essay has a clear organization. Although unclear	-Some parts of the essay are organized, but some parts are poorly organized or	-Ideas lack organization. The connectors do not show a logical



logical way	paragraph and between paragraphs flow smoothly and are very easy to follow. All connectors show a clear and logical connection between ideas.	connectors between ideas occasionally appear, they do not cause misunderstanding.	have connectors that obscure the connection of ideas and cause misunderstanding.	connection between ideas and cause misunderstanding.
Critical thinking domains	Excellent (4)	Good (3)	Fair (2)	Poor (1)
6) Ability to use word choices and sentence structures effectively	-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 2

Interview questions for the raters

1. What do you think about the rubric?
2. What do you like or dislike about the rubric? Why or why not?
3. Do you face any difficulties when you use the rubric to mark students' essays?
4. Do you have other comments about the rubric?

Appendix 3

Descriptions of the six domains of the CTER

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

The introduction of an argumentative essay should contain opening sentence(s), background information and a claim. To make the opening sentences interesting, EFL students can use a rhetorical strategy, e.g., beginning with a relevant, interesting quotation, using a metaphor or a simile, using a rhetorical question, beginning with research findings, making a confession that is related to a topic, or stating a striking problem (Kalchayanant, 2009). Background information links the opening sentence to a thesis statement. The thesis statement should be written clearly to help readers clearly understand the writer's point (Rosenwasser & Stephen, 2012). This domain can assess critical thinking subskills, namely planning or designing how to write an introduction and developing a clear position.

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

Reasons are logical statements or explanations that support the claim (Stapleton, 2001). Evidence can be personal experience, research studies, personal observation, opinions of the experts, interviews, comparison and contrast, anecdotes and so on (Stapleton, 2001; Kalchayanant, 2009). In providing reasons and evidence, students need to avoid fallacy, such as a false authority, a personal attack, a hasty generalization, non sequitur (making a conclusion that does not follow from a given fact) and climbing on the bandwagon (arguing that everyone does it, so readers should do it) (Kalchayanant, 2009; Kirby & Goodpaster, 2007). Giving reasons and evidence is the core of critical thinking and argumentation (Kirby & Goodpaster, 2007). This domain can assess critical thinking subskills, namely seeking, analyzing and evaluating information, constructing arguments and supporting claims with relevant and compelling evidence.

3) Ability to state a counter-argument and a refutation

A counter-argument is an opposing view, whereas a refutation is a statement that refutes such opposing view. Stating this in an argumentative essay requires a problem-solving skill (Cheng, 2003). Students can refute the counter-argument by explaining why one side is stronger than the other, identifying a fallacy in reasoning of the weaker side or providing a solution (Nussbaum & Schraw, 2007). This domain can assess critical thinking subskills, namely analyzing and



evaluating information from various perspectives, and critiquing or defending a position with logical reasons and evidence.

4) Ability to write a conclusion

Writing a conclusion in argumentative writing manifests students' critical thinking. Conclusion should contain 1) a summary or a restatement of a claim and 2) a final thought or a final, constructive comment (Oshima, & Hogue, 2006). Constructive comment intends to help or improve something. It can be a suggestion, a recommendation for someone to take an action or a solution (Kalchayanant, 2009). This domain can assess critical thinking subskills, namely analyzing and synthesizing the arguments by summarizing thoughts and reflecting final, constructive comments for readers.

5) Ability to organize ideas in a logical way

Effective organization of ideas in argumentative writing reflects critical thinking (Bauer, 2016). Students can use transitional or cohesive devices to connect ideas logically. To evaluate the organization of ideas, teachers need to consider whether an essay starts with an introduction, followed by body paragraphs and a conclusion, whether ideas in each paragraph and those between paragraphs are well-organized and connect altogether and whether logical connectors are used clearly and appropriately (Lau, 2011). This domain can evaluate critical thinking subskills, 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.

6) Ability to word choices and sentence structures effectively

Using effective language in argumentative writing shows critical thinking skills, such as analyzing, evaluating and carefully revising word choices and sentence structures for a clear communication. Word choices and sentence structures are language resources (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 language use carefully to convey ideas clearly. Moreover, he or she can use metaphors and similes in argumentative writing (Maybin & Swann, 2006). Metaphor is a part of language and argumentation (Verlinden, 2005) and can be used to make arguments clearer and more interesting (Mayweg-Paus, Thiebach, & Jucks, 2016). Metaphor reflects the writer's both critical and creative thinking (Kirby & Goodpaster, 2007). This domain helps raise teachers' awareness of teaching language use to EFL students so that they can express thoughts in writing effectively (Manalo & Sheppard, 2016).