

Quality of Argumentation Models

Chamnong Kaewpet
Rajamangala University of Technology Krungthep, Bangkok, Thailand

Abstract—This study investigated the effectiveness of updated argumentation quality criteria. It evaluated the scale and quality of selected argumentation models judged by the new criteria. Effectiveness concerned content validity, reliability, and practicality of the criteria. The argumentation models were regarded as possessing good quality when they featured important elements in the criteria and received high scores. Five argumentation models were purposefully selected from an argumentative writing course. The models were evaluated by three evaluators with expertise in academic writing. Analysis confirmed the effectiveness of the quality criteria and scale. It addressed all important concerns in evaluation of argumentation; evaluation scores were in accordance with each other, and the important argumentation elements carried equivalent weight. Only three of the models received a quality score of 4 on a scale of 0 (null) to 5 (highest), because they did not feature all quality elements required by the criteria. The updated framework and argumentation models can be further employed for teaching, learning and evaluating argumentation.

Index Terms—argumentation, English, criteria, model, quality, Thai

I. INTRODUCTION

Argumentation has been taught and learned in classrooms around the world. Ability in the skill of argumentation is measured in high-stake English standardized tests including Test of English as a Foreign Language (TOEFL) and The International English Language Testing System (IELTS). In education, argumentation has been identified as the language ability at the B2 level of the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2001). In the classroom, for example, the Common Core State Standards, adopted by more than 40 states in America, stresses argumentation skills, especially skills in logical arguments and using relevant evidence (Deane & Song, 2014). In Malaysia, trainee teachers study argumentative writing because it helps their future students become critical and reflective thinkers (Bipinchandra, Shah, Puteh, Din, Rahamat & Aziz, 2014). In Chile, argumentation is researched as a predictive tool of university study and academic achievement over time (Preiss, Castillo, Flotts & Martin, 2013).

Among written text types, argumentation has been reported to be the hardest, in comparison with others such as description and narration (Aertselaer & Dafouz-Milne, 2008; Yang & Sun, 2012). In fact, progress in mastering a new type of text covers three categories, namely, 1) narratives, 2) procedures, and 3) argumentative essays (Martin, 1989; Schleppegrell, 2004). The argumentative ability of school children progressively matures through the grade levels (McCann, 1989; Crammond, 1998). Narratives tend to be well mastered by around age ten, but argumentative writing develops at a later stage (Berman & Nir-Sagiv, 2007). By the time students enter university, they should already be able to write argumentation effectively. However, this conclusion may not always be applicable. According to one study, students at a college were not very effective at producing written arguments. They failed to include critical argumentative elements such as thesis, reasons and evidence. In other cases, they failed to clearly present the elements. As Golder and Courier (1996) point out, argumentative writing is difficult in itself due to its language-specific nature. It is even more challenging for students who write in different languages (Aertselaer & Dafouz-Milne, 2008).

When it comes to argumentation in teaching English as an additional language (TEAL), some form of instruction is important to yield satisfactory results. Findings of one L2 study were truly surprising. Knoch, Rouhshad, Oon and Storch (2015) report that the writing of undergraduate students in their study improved after three years of study, but only in terms of fluency. Changes in accuracy, grammatical and lexical complexity, as well as global scores of writing fell short of significance. More often than not, formal instruction effectively minimizes such problems. In another L2 context, Varghese and Abraham (1998) taught argumentative writing using Toulmin's classic scheme/structure of argumentation. They found that students made obvious progress after instruction. They produced more claims, specific and developed data, and reliable warrants. They also became more aware of views from opposing sides. In another study, Liu and Stapleton (2014) examined argumentative essays written by students at a Chinese university. The students received instruction in argumentation which included counter arguing and refuting. It was found that inclusion of the elements was positively correlated with the overall score on argumentative essays.

More studies such as these have yielded positive results as a consequence of instruction. Many researchers have investigated teaching techniques such as public Internet discussion forums (Ritchie & Black, 2012), online peer feedback (Noroozi, Biemans & Mulder, 2016), parent-child discussion (Domenech & Krah, 2014), and debates (Dickson, 1994). The present study is motivated to investigate the quality of the techniques themselves. It focuses on argumentation models employed in the instruction. This consideration is based on a genre-based approach to teaching writing. In this approach, learners are first exposed to examples of target text types. They have a chance to study the

examples before they produce their own. Studies have demonstrated the usefulness of the genre-based approach to develop the writing ability of English learners (e.g., Bhatia, 1993, Paltridge, 2001, Hyland, 2004). This study is based on the premise that using good models will have a positive influence on the quality of novice writers' argumentation skills. In line with Noroozi, Biemans and Mulder (2016), the quality of edited essays in their study was dependent on the quality of peer feedback itself. In this scenario, the schemes of argumentation, such as those pioneered by Aristotle in ancient Greece, and revised by Qin and Karabakak in 2010, have been recently updated to keep pace with advancements in the field. The latest scheme has been transformed into a quality criteria and scale for teaching, learning and evaluation of argumentation. Like CEFR (2001), it is designed to be an all-inclusive framework. The criteria and scale were created by the researcher as a result of examination of related literature in the field of argumentation. It comprised five main categories: 1) relevancy, 2) reasoning, 3) language use, 4) organization and 4) writer's voice. The criteria and scale must be examined for quality (see sections C, F) when used to evaluate the argumentation models. Investigation into such criteria, scale and argumentation model is relatively scarce in recent publications except for one published by Khodabandeh (2014). The study also looked into model essays. However, the research emphasized moves, functions and linguistic forms, which are different from this study.

Research questions

Specifically, this study has aimed to find answers to two questions.

- (1) How effective are the updated quality criteria and scale for argumentation?
- (2) How complementary are the argumentation models with the updated quality criteria and scale?

II. MODELS AND QUALITY

It is considered important in this research that EAL learners have the opportunity to learn from good-quality models when they study argumentation, and instruction is organized around effective quality criteria and scale. Concepts of quality may be defined differently in various situations. Therefore, both the models and concepts of quality are re-examined in this study.

A. Argumentation Models

Learners can access argumentation examples from a variety of sources, for instance, from English standardized tests, instructional websites, academic papers, web articles, and commercial textbooks. The examples might be selected as models for teaching, learning, and evaluation. Based on these sources, argumentation is present in five main channels of communication: an independent task on a standardized English test; a student essay; a letter to an editor/principle/teacher; part of an academic paper/article; and a magazine/web page/columnn.

Argumentation is measured specifically on independent writing tasks on two English standardized tests, namely, TOEFL and IELTS. In TOEFL, test takers are asked to give reasons to support their choice of position – to agree or disagree with a given statement. They are allowed 30 minutes to complete their answer with a length of approximately 300 words. In IELTS, test takers are asked to write approximately 250 words in response to a point of view, argument or problem. They are given 40 minutes to complete the task. Hawthorn English Language Centres (1997) gives an example of a response to an IELTS task. A communication situation is given for test takers to present arguments to an educated non-specialist audience: "In the last 20 years, the assessment of students has undergone major transformation. Many educational institutions no longer use formal examinations as a means of assessment as they believe formal examination results are an unfair indication of a student's ability" (p. 33). The sample response is 370 words long and comprises six paragraphs. In the first paragraph, the controversial situation is repeated. In the second paragraph, the writer states her position –she believes that examinations can give an unfair assessment of students' ability. In the third paragraph, the writer gives examples and reasons for her position in more detail. In the fourth paragraph, she reports on an opposing position and one reason for that position. The writer quickly returns to her stated position. When she comes back, the opposing position and reasons are rebutted. Reasons are given for the rebuttals with examples of possible and fairer evaluation methods. In the concluding paragraph, she reviews the situation and restates her position, as well as gives recommendations for future examinations that should include the evaluation methods exemplified in the fourth paragraph.

Another source of argumentation models are instructional websites, such as one that teaches the skill of writing argumentative essays by Daly (1997). One of the two sample essays available on the website concerns marine parks. The essay length is approximately 734 words, containing six paragraphs. A controversial situation and issue are reported in the introductory paragraph, where the writer also positions himself – marine parks should be prohibited. After that a series of reasons for the position are laid out in the next four body paragraphs. The body paragraphs discuss important topics one by one. The first three body paragraphs establish the opposing position and its reasons first and then switch to the writer's chosen perspective. The opposing contents are split almost into equal halves by a single connective, the word 'however.' The fourth body paragraph is dedicated to the writer's position with reasons to support it. The essay ends with a concluding paragraph, where the writer's position is restated, and the previously debated topics are mentioned. Also in the last paragraph, a strong final warning is given regarding negative consequences of the position in opposition to the writer. An outstanding feature of this essay is the inclusion of in-text citations to support both the writer's and the opposing views. Consequently, a list of references is given after the essay. This feature makes this essay unique because most examples of student essays do not contain citations and references.

Argumentation can also be found in the bodies of theses, academic reports and papers. In an introduction section to research (Chien, Lee & Kao, 2010), argumentation is analyzed in a sample paragraph. The paragraph is 176 words in length. The researchers support inclusion of English for specific purposes in university study programs. They begin by citing two experts' opinions regarding the possibility for English learners to transfer general English skills to tasks required in their specific disciplines. This suggests that English for specific purposes is not necessary. However, the writers take a turn in the middle of the paragraph to present an opposing recommendation, which encourages university students to improve their English ability specifically in their area of expertise. The proposition is followed by three supporting reasons and evidence. Like the student essay, relevant sources are cited to support both positions. This example of argumentation differs from the student essay, and differs from the response to the English standardized test, in that the writers' position is not so clearly stated.

Letters to editors/principles/teachers are frequently presented at school as examples of argumentation. An example of such a letter is given in a commercial textbook in writing (Azar, 1999). The argumentation concerns the high cost of college education. The letter has only one paragraph, and is approximately 147 words in length. The writer starts the letter with her family background. Then she moves on to a controversial issue under discussion, where she reports on her choice of a lower-cost college. After that, she describes the disadvantages of a possible circumstance that she might have faced if she had chosen a higher-cost institution. She also explains the advantages of the situation that she has chosen. The writer ends the letter with a conclusion: "After all, in four years, when you graduate, what you have achieved will reflect far more about you than where you achieved it" (p. 110). Personal stories count in this argumentation. The writer does not provide extensive explanations as shown in the earlier student essay. This is most likely due to the space limit of such a channel of communication.

Argumentation can also be observed in web articles. In the New York Times (Matter, 2013), a columnist wrote an article of 814 words. He focused on whether or not communication technology such as text messaging and Facebook can drain brain power and weaken work performance. He began with a discussion of the deleterious effect of multi-tasking. Then he reported his own investigation in collaboration with a university to measure the brain power lost when someone is interrupted. They asked research participants to complete a standard cognitive skill test, with some of them being purposely interrupted during the test. Although the participants who were interrupted performed lower than those who were not, they learned to adapt themselves and performed better later. In this context, participants who were warned of interruptions that never came performed better than participants who were left alone doing the test. At the end of the article, unlike the other argumentation models, the writer repeated both the opposing view and his own position. Another observable feature was the way the author presented content in very short paragraphs. Links to cited sources were also given intermittently throughout the article.

All of the examples cited thus far can be used as argumentation models in the classroom. Most of them have probably been edited before publication. Therefore, they should be of good quality. It is re-examined in this study, however, because an updated evaluation framework is in place (see section B).

B. Evaluation of Argumentation Quality

When argumentation is evaluated, the focus can be divided into three main categories, namely, general writing elements, argumentation elements, and other elements.

General writing elements mostly refer to task relevancy, organization, grammar, vocabulary, cohesion and coherence. The notion of task relevancy in argumentation differs from other kinds of writing. In argumentation, the content concerns a controversial issue and the writer must take a position regarding it. Relevancy is measured as part of idea development in TOEFL and task response/achievement in IELTS. It was measured in a study carried out by Stapleton and Wu (2015), analyzing the argumentative essays of students in Hong Kong. The essays were scored, firstly, for their relevancy to the topic, and, secondly, for three levels of acceptability: not acceptable, weak, and acceptable. Evaluation of the other five elements was generally conducted in the same way as other kinds of writing.

The evaluation is an overall approach, such as when an essay receiving a top score of five is described as "effectively addresses the topic and task," "is well organized and well developed, using clearly appropriate explanations, exemplifications and/or details," "displays unity, progression and coherence," and "displays consistent facility in the use of language, demonstrating syntactic variety, appropriate word choice and idiomaticity, though it may have minor lexical or grammatical errors" (Educational Testing Service–TOEFL 2016). Examination of the general elements may also be centered in only some of their features, such as when Maxwell-Reid (2015) analyzed clausal embedding in argumentative texts written by Spanish junior secondary students. In the study, the students' control of the language resources required for clausal embedding was found to vary depending on their ability.

Evaluation of argumentation more often focuses on general writing elements than argumentation elements. Common elements of argumentation are understood to include 1) the writer's statement/s of his/her position regarding a controversial issue; 2) reasons in support of the position; 3) a statement/s of the opposing position; 4) reasons in support of the opposing position; 5) the writer's statement rebutting the opposing position and reasons; and 6) his/her reasons for the rebuttals. The six elements or reasoning scheme are called claim, data, counter-argument claim, counter-argument data, rebuttal claim and rebuttal data, respectively (Qin & Karabakak, 2010). These elements have been applied to many research contexts. The elements originated in ancient Greece by Aristotle and were widely recognized when Toulmin introduced his scheme of argumentation in 1958 (Qin & Karabakak, 2010). For example, in

an investigation into argumentative papers written by second-year English-majors at a Chinese university, it was found that an average paper had at least one claim supported by four pieces of data. The papers contained fewer uses of counter-argument claims, counter-argument data, rebuttal claims, and rebuttal data. However, employment of these elements served as a predictor of the overall quality of the argumentative papers (Qin & Karabakak, 2010). As noticed, reasoning is the heart of argumentation. It should be more frequently emphasized in the evaluation of argumentation.

Other elements to be considered in the evaluation of argumentation include persuasiveness, argumentative strategies and writer's voice. Ferretti, Lewis and Andrews-Weckerly (2009) compared persuasiveness and quality of argumentative essays written by students who were assigned two instructional conditions: elaborated and general conditions. The students in both conditions wrote a letter to a teacher taking a position regarding an issue on homework assignments. The students' responses to the prompt were scored for overall persuasiveness, argumentative structures, and argumentative strategies. The analysis found that the students in the elaborated condition wrote more persuasively than those in the general condition. The argumentative strategies used were highly predictive of the essay quality, and nearly all students used the argument from consequences strategy to defend their standpoint. With regard to writer's voice or authorial voice, research has more frequently reported on its positive influences on quality of argumentation.

Yoon (2017) examined how the quantity and diversity of textual voice elements that a writer used would contribute to holistic voice strength and essay quality. Argumentative essays written by Greek-speaking EFL students were collected from a standardized English examination, which was developed at Michigan State University. The results suggested quite positive correlations between textual voice and holistic voice strength. The textual voice features explained 26 percent of the variance in voice strength scores. The results also indicated weak correlations between textual voice and essay quality.

To summarize, an investigation into the quality of argumentation could make use of the examples presented in section A, depending on the research purposes. More importantly, it should take both the general writing elements and the argumentation elements into account. As for the other elements, writer's voice should be built into the existing criteria because it has received more attention and casts a stronger influence on the writer's position. It can actually be regarded as advancement in the field.

III. RESEARCH METHODS

Three research methods were accomplished in this study: selection of research materials, invitation of evaluators, and data analysis.

A. Materials

Two kinds of materials were evaluated in this study; 1) an argumentation quality criteria and scale; and 2) argumentation models used in a real classroom.

The criteria and scale were created by the researcher (Kaewpet, in press) based on examination of related literature to argumentation. It has drawn on CEFR specifications of argumentation at the B2 level (Council of Europe, 2001), independent tasks of TOEFL (ETS–TOEFL, 2017) and IELTS (IELTS, 2016), elements of argumentation such as those reported in section B (Qin & Karabakak, 2010; Stapleton & Wu, 2015), plus other recently published articles in argumentation (e.g., Yang & Sun, 2012; Liu, 2013; Bastukmen & Randow, 2014; Crossley, Kyle & McNamara, 2016; Lancaster, 2016; Yoon, 2017; Zhao, 2017). The criteria and scale consist of five main categories and eleven items in total (cf. Introduction, see TABLE I). The criteria and scale were further investigated for content validity and reliability in this study.

TABLE I.
ARGUMENTATION QUALITY CRITERIA AND SCALE

Directions: Please choose the degree to which the argumentation has met the criteria (0 = not at all, 5 = most).						
Criteria	0	1	2	3	4	5
I. Relevancy						
1. The writer has addressed a/ the given controversial situation.						
II. Reasoning						
2. The writer has stated his/her position regarding the controversial issue.						
3. The writer has given adequate reasons in support of his/her position.						
4. The writer has shown his/her awareness of the opposing position and included opposing reasons in the essay/response.						
5. The writer has given reasons to rebut the opposing reasons and to maintain his/her position.						
III. Language Use						
6. Use of grammar is effective.						
7. Use of vocabulary is effective.						
IV. Organization						
8. The essay has good organization.						
9. Use of connectives is effective.						
10. Ideas are developed effectively.						
V. Writer's voice						
11. You can feel the presence of the writer's authority and confidence.						

Argumentation models in this study were meant to be those easily and internationally accessible such as those presented in section A, and they had to be employed in real instruction. This ensures wide and real applicability of the models. To fulfill this purpose, a research site had to be selected, and the argumentation models had to be utilized at the site. In accordance with the requirements, a course on argumentative writing was offered at a university in Thailand. The course was selected as the site of the study. According to the course instructor, ten argumentation models were used in the course, and approximately five of them could be covered at a time. The course length was 15 weeks with a lesson of three hours each week. The instructor was the only person teaching the course. She had been teaching this course for approximately ten years. Therefore, she was the only expert available at the site. However, that was not to guarantee the best quality of the models, because the models could have been employed for many purposes such as for analysis or improvement. The purposeful sampling method was similar to that employed in Stapleton and Wu's study (2015), in which research materials were first examined to ensure elements of argumentation before they were evaluated for their reasoning quality. Five of the course's argumentation models were randomly selected for this study. Argumentation model IV was the same as the formal examination reported in section A.

Argumentation model I was organized the same as the marine parks essay in section A. It came from the same source (Daly, 1997). The writer prepared it for teaching argumentative essay to students undertaking the Victorian Certificate in English for Vocational Education and Further Study course at Victoria University of Technology in Melbourne, Australia. The essay was 379 words long. The issue concerned which was the better condition for caring for children: in their own homes or at childcare centers. The writer argues for children's own homes. The differences between the marine parks and childcare essays were that the childcare essay had one less paragraph. The missing paragraph was the one dedicated for the writer's side of reasons. The body paragraphs are also organized around one topic at a time. They began with the opposing side and then switched to the writer's side by using the same connective 'however.' There were no in-text citations in this essay, however, and references were not provided. As a result, the body paragraphs and reasoning were briefer. The concluding paragraph was also shorter – only one long sentence to repeat the writer's position and main idea.

Argumentation model II came from a commercial textbook focusing on writing. The essay was on whether or not people should choose faith healers over modern medicine. It was 276 words long and contained five paragraphs. In the first paragraph, the controversial situation and issue were introduced and the writer's position was stated: "Going to a faith healer is a risky thing to do, even if science cannot solve your problem" (Pavlik & Segal, 2002, p. 150). The second paragraph had only one sentence and functioned to restate the writer's position and main idea. In the third paragraph, an anecdote was used as evidence of the position and main idea. The fourth paragraph was also short and used to emphasize the writer's position. In the last paragraph, the position was restated. In this model, the opposing position was presented along with the writer's position in the first paragraph. However, it was not explained in the other paragraphs.

Argumentation model III was drawn from a website which gave examples of argumentation for free (English Club 1997-2018). The essay was 346 words long and offered two options: blocking employees from social media or giving access to it. A prompt was given before the essay. The example essay in response to the prompt contained five paragraphs. The first paragraph repeated the controversial situation and announced the writer's position: "It is a wise decision to remove these Internet temptations." In the second paragraph, reasons were given to explain the position. Negative consequences on employers were demonstrated in the third paragraph, and the negative consequences on employees were explained in the fourth paragraph. In the concluding paragraph, the writer's position was further explained and restated.

Argumentation model V was an example of a TOEFL test. The issue in question concerned which was the more important quality of teachers – the ability to relate well with students or excellent knowledge of the subject taught (Educational Testing Service–TOEFL 2016). The length of the essay was 436 words. A prompt was given before the essay. The prompt consisted of the issue in question and directions for writing an answer. The essay consisted of five paragraphs. In the first paragraph, the writer provided a background to her position settling in a kindergarten where she remembered all of the teachers. In the second paragraph her position was stated –she believed that teachers should relate well with students. In the third paragraph, the position was explained using the writer's personal experiences. The fourth paragraph was again dedicated to the writer's position. The opposing position was referred to in this paragraph, but it was used to highlight the writer's own position. In the last paragraph, the personal experiences shared in the third paragraph were extended to a more general situation before the writer's position is restated.

The criteria and scale and the models were accessible to the research's evaluators (see section D). A Google site web page was created and links to them were provided. The models were kept on Google Drive and the evaluation forms were created using Google Forms.

B. Evaluators

Three TEAL instructors who had experience teaching writing and/or English standardized tests were invited to evaluate the models using the created criteria and scale. All of them were working in Thailand. The first evaluator was from the United Kingdom and was working at the research site. He had been teaching English for seven years. He taught courses including Essay Writing, English Standardized Tests and Independent Study. He had also published a number of research papers. The second and third evaluators were not working at the research site, but had expressed

their interest in working there. The second evaluator was an instructor from South Africa who had been working in Thailand for nine years. He had taught English courses at a high school and an undergraduate institution. The courses he taught included English for Writing and some preparation courses for English standardized tests. The third evaluator was from India. She had been teaching English in India, Cambodia and Thailand for 25 years. She had extensive experience in academic work and activities such as participating in conferences, chairing meetings and supervising doctoral theses. Links to the argumentation models as well as the evaluation forms/quality criteria and scale were sent to the evaluators. The evaluators were contacted by email and evaluated the models using the online evaluation form.

C. Data Analysis

The evaluation data was analyzed to determine the effectiveness of the updated quality criteria and scale, and the respective argumentation model with the updated quality criteria and scale. The criteria and scale were considered as effective when they addressed all important concerns in evaluation of argumentation, when scorers' evaluations agreed with each other, and when the important argumentation elements carried the same weight. The existing models were regarded as of good quality if they received a quality score of 4 on a scale of 0 (null) to 5 (highest), when they featured all quality elements in the criteria and scale, and when the scorers' evaluation agreed with each other. The analysis was carried out through free statistical calculators provided for students and researchers working in the social sciences. The output of the calculators has been audited for accuracy against the output produced by a number of established statistics packages, including SPSS and Minitap (Stangroom, 2017).

In order to analyze the argumentation quality criteria and scale in terms of effectiveness in addressing all important concerns, content validity was carried out by re-examining related literature to argumentation. When the related literature was examined the first time with the aim to update the quality criteria and scale (see section C), the literature was limited to publications in high-stake databases in the past ten years. Later, interesting issues raised in the publications, although outside the period, were also explored. In the first round, the focus was on existing schemes of argumentation, conceptions of argumentation in four obvious sources (CEFR, TOEFL, IELTS, TOEIC), plus any others that could contribute criteria for quality of argumentation. In this study, the literature was re-examined focusing on important general writing elements, argumentation elements, and other elements for evaluating argumentation. Knowledge obtained from this round was regarded as principles to look into the initial framework to determine its validity. The procedure was in fact qualitative content analysis. New knowledge emerged through processes of focusing, organizing, coding, classifying, reducing the data, as well as interpreting and making conclusions (Lynch, 1996; Neuman, 2000). The procedure arrived at the stage of measuring the framework for its credibility, one of the important measurements of trustworthiness in qualitative studies. Credibility is another word to refer to validity in qualitative research.

Establishing credibility began with evaluating the effectiveness of the argumentation criteria and scale in terms of scorers' agreement in the evaluation, its effectiveness to balance important argumentation elements to equal weight, and to measure the mean scores for quality of the argumentation models. Next, it was necessary to determine whether or not the models featured all quality elements in the criteria and scale. The raw data collected from the Google Forms was entered into the one-way analysis of co-variance (ANOVA) calculator available on the web site by scorer, criterion, and model, respectively. Statistically, the procedures were carried out to establish the reliability of the data.

IV. FINDINGS AND DISCUSSION

The research was conducted to find answers to questions regarding the effectiveness of the updated quality criteria and scale, and to enhance the argumentation models with the updated criteria, in other words, to improve the quality of the argumentation models.

A. Effectiveness of the Updated Quality Criteria and Scale

Effectiveness of the updated criteria and scale was observable from its content validity, the evaluators' agreement in the evaluation, and its effectiveness in giving equal weight to the important argumentation elements. The literature review reported in section II confirmed that the updated criteria and scale drawn from related literature in the first round (cf. section E) was effective in terms of content validity. It comprised: 1) relevance, which was one of the common general writing elements; 2) reasoning, which embraced common argumentation elements; 3) language use: grammar, vocabulary, which were two of the six common elements of general writing; 4) organization, including organizational patterns, cohesion and coherence, which were three additional general writing elements; and 5) writer's voice, which was an important element in the other concerns category (cf. Introduction, section C).

The analysis of the evaluation data by ANOVA revealed that the updated quality criteria and scale were also effective in terms of reliability. TABLE II and TABLE III show that they were effective because 1) the evaluators' evaluations agreed with each other, 2) their evaluation was consistent across the criteria and models, and 3) the quality criteria and scale gave equal weight to the important argumentation elements.

TABLE II.
EFFECTIVENESS OF THE UPDATED CRITERIA AND SCALE BY EVALUATOR

Summary			Result Details			
Evaluator	Mean	Std. Dev.	Source	SS	df	MS
I.	3.6111	1.28	Between evaluators	57.7901	2	28.8951
II.	4.7037	2.5	Within evaluators	3402.0926	159	21.3968
III.	5	7.5035	Total	3459.8827	161	

The *f*-ratio value is 1.35044. The *p*-value is .262082. The result is *not* significant at $p < .05$.

TABLE III.
EFFECTIVENESS OF THE UPDATED CRITERIA AND SCALE BY CRITERION

Summary			Result Details			
Criterion	Mean	Std. Dev.	Source	SS	df	MS
I.	4.1333	1.1872	Between criteria	3.5273	4	0.8818
II.	3.7333	1.3387	Within criteria	212.2667	160	1.3267
III.	4.0667	0.8277	Total	215.7939	164	
IV.	3.8000	1.1201				
V.	3.8667	0.9155				

The *f*-ratio value is 0.66469. The *p*-value is .61741. The result is *not* significant at $p < .05$.

The results provide insights into the number of evaluators that should be invited to mark a subjective test, the significance of availability of a practical evaluation method, and the importance of each of the evaluation criteria in the updated framework. Firstly, it is generally recommended that a subjective or essay test should be scored by more than one evaluator. In TOEFL for example, trained raters plus automated scoring technology are used to score the writing section (ETS–TOEFL, 2017). However, the results of this study suggest that only one evaluator is also dependable, when he/she is experienced and knowledgeable in the evaluation area. In this study, the evaluators were experts in the field of academic writing, and so no significant differences were found among their scores. Secondly, many subjective tests are scored using a score band descriptor – holistic or analytical, which requires specialized training, and makes marking a difficult job. In this study, a marking tool, or the quality criteria and scale, was designed to be as marker-friendly as possible. The technical terms, such as claim and data, were presented in the form of the vernacular language and a rating scale. Therefore, it was practical for the evaluators, and no additional training was necessary. The evaluators were not trained in using the updated quality criteria and scale before the research. A design like this one could be proved useful in other situations. Finally, it was found that all of the criteria were of the same significance – there were no significant differences among the scores. This implies that instruction and evaluation on argumentation should take all of the important argumentation elements into account.

B. Quality of the Argumentation Models

The quality of the argumentation models was observable from their means scores, the way they were enhanced by the quality criteria and scale, and whether or not the scorers' evaluations were in accordance with each other. The means score defined as good quality in this study was a minimum of 4 on a scale from 0 to 5. TABLE IV demonstrates that the quality scores of the argumentation models ranged from 2.76 (Model 2) to 4.58 (Model 1). The scores were significantly different as judged by the updated quality criteria and scale and the three evaluators. As reported in the previous section, the evaluation scores of each evaluator were found to concur with each other. Therefore, it was most likely the presence or absence of the important argumentation elements that caused differences among the scores.

TABLE IV.
EFFECTIVENESS OF THE UPDATED CRITERIA AND SCALE BY MODEL

Summary			Result Details			
Model	Mean	Std. Dev.	Source	SS	df	MS
I.	4.5758	0.5019	Between models	66.703	4	16.6758
II.	2.7576	1.1734	Within models	149.0909	160	0.9318
III.	4	1.199	Total	215.7939	164	
IV.	4.3333	0.736				
V.	3.6364	1.0252				

The *f*-ratio value is 17.89594. The *p*-value is $< .00001$. The result is significant at $p < .05$.

When looking into the individual argumentation models described in section C, one would understand that it is essential to include all of the important elements in his/her argumentation. It has been reported by many studies that some argumentation elements, such as counter-argument data and rebuttals, more strongly influence the quality of argumentation than the others (e.g., Qin & Karabakak, 2010). This study has confirmed that the two models receiving the lowest scores were missing the opposing views and reasons. Although one of the higher-scoring models did not contain many of the elements, it compensated for this lack through stronger usage of other elements. The optimum quality of argumentation should contain all of the important elements.

V. CONCLUSION

The study reported in this paper has confirmed the validity, reliability and practicality of the updated quality criteria and scale. It can be used as principles for teaching, learning and evaluation of argumentation particularly in TEAL, where students have to deal with the specific nature of argumentation and write in a different language (Golder & Courier, 1996; Aertselaer & Dafouz-Milne, 2008). In addition, the argumentation models investigated can be further employed as good or needing improvement, depending on the context of use.

REFERENCES

- [1] Aertselaer, J. A. N. & Dafouz-Milne, E. (2008). Argumentative patterns in different languages: An analysis of metadiscourse markers in English and Spanish texts. In Putz, M. & Aertselaer, J. A. N. (eds.), *Developing contrastive pragmatics interlanguage and cross-cultural perspectives*. Berlin: Mouton de Gruyter, 87–102.
- [2] Azar, M. (1999). Argumentative text as rhetorical structure: An application of rhetorical structure theory. *Argumentation*, 13, 97–114.
- [3] Basturkmen, H. & Randow, J. v. (2014). Guiding the reader (or not) to re-create coherence: Observations on postgraduate student writing in an academic argumentative writing task. *Journal of English for Academic Purposes*, 16, 14–22.
- [4] Berman, R.A. & Nir-Sagiv, B. (2007). Comparing narrative and expository text construction across adolescence: A developmental paradox. *Discourse Processes*, 43: 79–120.
- [5] Bhatia, V. K., (1993): Analysing genre: language use in professional settings. London: Longman.
- [6] Bipinchandra, J. S., Shah, P. M., Puteh, S. N., Din, R., Rahamat, R. & Aziz, J. A. (2014). User needs analysis in learning argumentative writing via mobile platform. *Procedia Social and Behavioral Sciences*, 118, 198–205.
- [7] Chien, C., Lee, W. & Kao, L. (2010). Collaborative teaching in an ESP program. *The Asian Journal Quarterly*, 10, 114–133.
- [8] Council of Europe (2001). Common Europe Framework of Reference for Languages: Learning, Teaching, Assessment. http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf (accessed 11/3/ 2016).
- [9] Crammond, J. (1998). The uses and complexity of argument structures in expert and student persuasive writing. *Written Communication*, 15: 230–268.
- [10] Crossley, S.A., Kyle, K. & McNamamura, D.S. (2016). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing*, 32, 1–16.
- [11] Daly, B. (1997). Writing Argumentative Essays. Available: <http://www.ltn.lv~markie/essaywriting/frntpage.htm> (accessed 2/11/ 2008).
- [12] Deane, P. & Song, Y. (2014). A case study in principled assessment design: Designing assessments to measure and support the development argumentative reading and writing skills. *Psicologia Educavita*, 20, 99–108.
- [13] Dickson, R. (1994). Developing "real-world intelligence": Teaching argumentative writing through debate. *The English Journal*, 94, 34–40.
- [14] Domenech, M. & Krah, A. (2014). What familial aspects matter? Investigating argumentative competences of learners at the beginning of secondary schooling in the light of family-based resources. *Learning, Culture and Social Interaction*, 3, 77–87.
- [15] Educational Testing Service–TOEFL. (2016). For test takers. Available: <https://www.ets.org/toefl> (accessed 25/4 2016).
- [16] English Club. (1997–2018). TOEIC practice – writing. Available: <https://www.englishclub.com/esl-exams/ets-toeic-practice-writing.htm> (27/4 2015).
- [17] ETS–TOEFL. (2017). Understanding your TOEFL iBT test scores. Available: <https://www.ets.org/toefl/ibt/scores/understand/> (22/4/ 2017).
- [18] Ferretti, R.P., Lewis, W.E. & Andrews-Weckerly, S. (2009). Do goals affect the structure of students' argumentative writing strategies? *Journal of Educational Psychology*, 101, 577–589.
- [19] Golder, C. & Coirier, P. (1994). Argumentative text writing: Developmental trends. *Discourse Processes*, 18, 187–210.
- [20] Hawthorn English Language Centres (1997). IELTS to success. Queensland: John Wiler & Sons.
- [21] Hyland, K. (2004). Genre and second language writing. Ann Arbor: University of Michigan Press.
- [22] IELTS (2016). About Us. <https://www.ielts.org/> (accessed 12/4/ 2016).
- [23] Khodabandeh, F. (2014). Argumentation across L1 and L2: Examination of three instructional treatments of genre-based approach to teaching writing. *Procedia Social and Behavioral Sciences*, 98, 968–975.
- [24] Kaewpet, C. (in press). Criteria and scale for argumentation. *Theory and Practice in Language Studies*, 8, –.
- [25] Knoch, U., Rouhshad, A., Oon, S.P. & Storch, No. (2015). What happens to ESL students' writing after three years of study at an English medium university. *Journal of Second Language Writing*, 28, 39–52.
- [26] Lancaster, Z. (2016). Expressing stance in undergraduate writing: Discipline specific and general qualities. *Journal of English for Academic Purposes*, 23, 16–30.
- [27] Liu, F. & Stapleton, P. (2014). Counterargumentation and cultivation of critical thinking in argumentative writing: investigating washback from a high-stakes test. *System*, 45, 117–128.
- [28] Liu, X. (2013). Evaluation in Chinese university EFL students' English argumentative writing: an appraisal study. *Electronic Journal of Foreign Language Teaching*, 10, 40–53.
- [29] Lynch, B. K. (1996). Language program evaluation: Theory and practice. Cambridge: Cambridge University Press.
- [30] Martin, J. R. (1989). Factual writing: Exploring and challenging social reality. Oxford: Oxford University Press.
- [31] Matter, G. (2013). Brain Interrupted. Available: <https://www.nytimes.com/2013/05/05/opinion/sunday/a-focus-on-distraction.html> (28/4/2017).
- [32] Maxwell-Reid, C. (2015). The role of clausal embedding in the argumentative writing of adolescent learners of English. *System*, 49, 28–38.
- [33] McCann, T. (1989). Student argumentative writing knowledge and ability at three grade levels. *Research in the Teaching of English*, 23, 62–76.
- [34] Neuman, W. L. (2000). Social research methods: Qualitative and quantitative approaches. Boston: Allyn and Bacon.

- [35] Noroozi, O., Biemans H. & Mulder, M. (2016). Relations between scripted online peer feedback processes and quality of written argumentative essay. *Internet and Higher Education*. 31, 20–31.
- [36] Paltridge, B. (2001). *Genre and the language learning classroom*. Ann Arbor: University of Michigan Press.
- [37] Pavlik, C. & Segal, M. K. (2002). *Interactions 2 writing*. New York: McGraw-Hill Contemporary.
- [38] Preiss, D. D., Castillo, J. C., Flotts, P. & Martin, E. S. (2013). Assessment of argumentative writing and critical thinking in higher education. *Educational Correlates and Gender Differences*. 28, 193–203.
- [39] Qin, J. & Karabacak, E. (2010). The analysis of Tomlin's elements in Chinese EFL university argumentative writing. *System*. 38, 444–456.
- [40] Ritchie, M. & Black, C. (2012). Public internet forums: Can they enhance argumentative writing skills of second language learners? *Foreign Language Annuals*. 45, 349–361.
- [41] Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics perspective*. New York: Routledge.
- [42] Stangroom, J. (2017). Social Science Statistics. Available: <http://www.socscistatistics.com/tests/Default.aspx> (Assessed 28/4/2017).
- [43] 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*. 17, 12–23.
- [44] Varghese, S.A., Abraham, S.A. (1998). Undergraduates arguing a case. *Journal of Second Language Writing*. 7, 287–306.
- [45] Yang, W. & Sun, Y. (2012). The use of cohesive devices in argumentative writing by Chinese EFL learners at different proficiency levels. *Linguistics and Education*. 23, 31–48.
- [46] Yoon, H. (2017). Textual voice elements and voice strength in EFL argumentative writing. *Assessing Writing*. 32, 72–84.
- [47] Zhao, C. G. (2017). Voice in timed L2 argumentative essay writing. *Assessing Writing*. 31, 73–83.

Chamnong Kaewpet is an assistant professor at Faculty of Liberal Arts, Rajamangala University of Technology Krungthep, where she teaches argumentation and independent study to Thai students. She obtained a doctoral degree from Melbourne University, Australia. Her research interests lie primarily in the same area as the courses taught. She can be contacted at chamnong.k@mail.rmutk.ac.th.