

# Cultivating and Assessing Critical Thinkers through Second Language Writing

Tong Chu

*School of Foreign Language Education Jilin University, Changchun 130012, China*

**ABSTRACT.** *Cultivating and measuring critical thinking (CT) skills are an essential part in higher education. General assessment of CT skills existed for long but specific assessment of each dimension of CT skills is rare. The present study investigates one essential dimension of CT skills—inference and modified existent inference evaluation patterns to form an inference performance framework. This modified framework is applied in the experiment aiming to investigate the connection between inference abilities of the Chinese college students and their quality of argument in their English essays. 55 participants were assigned to write argumentative essays following the classical argumentation model. By modifying the CT Inference Performance framework, the study applies it in evaluating students' CT ability in argumentative writing, revealing several patterns of inadequacies in the inference dimension of the students' essays, exposing the need to bring more attention to the teaching of probability and form in the argumentative writing. Moreover, the study also indicates a significant connection between CT inference ability and quality of argument. On the basis of the above findings, inference-concentration teaching approach to improve quality of argument is recommended as a general guide for future classroom teaching, taking into account not only argumentative essay structure, content, and language, but also inference dimension of CT.*

**KEYWORDS:** *Critical thinking, Second language writing, Inference assessment, Quality of argument*

## 1. Introduction

Cultivating critical thinking skills has been a heated topic ever since the 1990s. At present, a variety of theoretical models and research methods have been successively introduced, and scholars have reached a basic consensus on such core concepts as critical thinking skills, critical thinking disposition and meta-critical thinking. The complexity and richness of studies on “critical thinking” has indeed enhanced the attractiveness of critical thinking research, but challenges still remain in the studies of assessment and cultivation of critical thinking ability.

Practice of second language writing has been proved to be an important way to enhance critical thinking (Stapleton, 2001). As a strategy for active learning, writing facilitates greater self-reflection and deeper thinking than speaking, listening and reading. Second language teachers can design appropriate writing tasks such as argumentative essay writing to help students think more critically by explaining, analyzing, inferencing, and evaluating multiple sides of controversial issues. Previous studies have provided a role model as to how to develop students' critical thinking skills through writing practice (Wade, 1995), but their application did not include Chinese context. This study intends to develop a practical framework for assessing students' critical thinking skills in the Chinese context and meanwhile, explore the applicable critical thinking assessment framework.

## 2. Cultivation of Critical Thinking

A number of approaches can be adopted to foster critical thinking in the context of English as a foreign language (EFL) courses. Activities such as writing arguments and conducting debates are especially meaningful tools to help students enhance critical thinking abilities. In the following sections, this study will showcase one phase of a semester's critical-thinking-embedded college writing course with detailed assignment design and critical thinking assessment. The participants were chosen from 61 second-year honored students of experimental class from a research university in northeast China. They were student scientists whose academic achievements outperformed other undergraduates in the same university, majoring in math, physics, chemistry, and biology. All students were native Chinese speakers and English was their first foreign language. After being assigned an argumentative essay topic, 60 students submitted the first draft and among them, 58 students submitted the second draft. 55 second-draft essays were properly peer-reviewed and teacher-reviewed, thus these

55 students became the final participants.

In this study, all participants had some experience of practicing critical thinking skills in the previous weeks of this course. They knew some essential dimensions of critical thinking skills such as analysis, evaluation and inference. They also displayed a high level of disposition in critical thinking when they first entered college and received a critical thinking disposition test. Even so, one invited English writing professor still gave a 40-minute lecture and instructed participants the essential elements of critical thinking skills: analysis, interpretation, inference and evaluation (Facione, 1990; Paul and Elder, 2006; Wen, 2009).

Participants were taught 8 class hours by the same professor of argumentative elements focusing on claims, counterarguments, rebuttals, and how to use reasons and evidence to develop a supporting claim, etc.. Toulmine model (1958, 2003) was also introduced in one of the eight class hours to acquaint students with a formalized persuasion process and enrich their experience in argumentative skills. Apart from instructions, four sample essays, two by native writers, and two by their seniors, were passed to students during these eight hours so that participants knew exactly what an argumentation essay required.

The following argumentation topic was determined through the students' brainstorming and voting, as students were more willing to write topics that they were familiar with and concerned about:

*In recent years, an increasing number of young people in China choose a lifestyle of spending much time inside instead of outside. In China, we call these group of people "Zhai"; In Japan, this group of people are named "Otaku". To what extent do you agree or disagree that "Zhai" or "Otaku" is not a recommended way of life for youngsters? (400 words or above)*

### 3. Assessment of Critical Thinking

To assess the exact critical thinking level of the students, the study selected the inference skill of all critical thinking dimensions. A coding scheme was modified and developed from Zarefsky's inferential patterns (2005, p.57) after the essays were collected. Students' essays were coded based on 6 elements: cause and effect, example, commonplace, form, probability, and analogy. Some of them have notable semantic or prepositional features. For example, "because" or "therefore" often indicated the causal relations; "for example" indicated example inferences; Maxim and proverbs as commonplace inferences are convenient to find with either quotation marks or a renowned guru's name adjacent to what they were intended to prove. For the rest of elements, the invited two experienced college professors specializing English writing, one from China (the one mentioned in the lecture and instruction session) and the other from the United States, determined the corresponding dimensions after discussion about 20 samples for each. One more thing to note was that for each coding, more than one sentence may suffice. For example, an example inference may contain as many as 6 sentences.

The elements of the above performance framework determined the basic coding principle; however, they were not adequate to measure the quality of each coded sentences. Therefore, we adapted inference hierarchy by Stapleton and Wu (2015) to develop another set of rubrics to assess quality of inference in this study, with irrelevant inference scoring 0, inaccurate inference scoring 1, illogical inference scoring 2, and Justifiable inference scoring 3. For measuring inference quality, I deemed that the first two criteria "relevance" and "accuracy" as the most basic requirement of an inference. Between the two, relevance was a necessary condition for further consideration of accuracy. Logic and justifiability then came next as higher-order criteria.

The above mentioned two experienced college professors were invited to discuss the rubrics and later coded and rated the participants' essays. To make the coding and rating process more reliable, the researcher and two professors selected at least 10 samples from the scripts, discussed and graded together to ascertain that their ratings of inference were within the acceptable range. I also treated the rating scale of the four categorical variables as Likert scales to make statistical calculations possible.

Lastly, the professors graded 55 essays based on the rubrics informed the participants during the 8 instruction hours. The evaluation of the essays focused on whether the participants effectively addressed the writing topic, whether the essays were well organized and well developed, whether participants used specific details and examples to support their views, and whether the essays displayed language facility by demonstrating syntactic variety, word choice and idiom. The scoring range was between 0-100. Meanwhile, participants were their peer reviewers. Generally speaking, each essay was graded 3 times by different participants other than the writer himself based on the same rubrics instructed by the professor. Both professors and participants' grading processes were assigned and operated on the website of "PEERCEPTIV Asia" which is an essay peer review

platform. When all grades and comments were submitted to the platform, the average score of each essay was automatically obtained and could be downloaded from the website.

Three set of results are shown in the following 2 tables, indicating the average inferencing ability of participants in relation to the quality of their argumentative essays, the strengths and weaknesses of participants in each inference pattern, and the relationships between each subskill of inference and the overall quality of the argumentative essays.

*Table 1 General Inference Skills in Relation to Grades of Arguments*

	highest	lowest	M	SD	r	Sig
inference	3	0.35	2.23	1.03	0.61	0.00**
Essay score	95	64	84.68	6.75		

\*\*p ≤ 0.01

Table 1 shows participants have displayed acceptable general inference skills with an average score of 2.23, showing that they have already achieved relevance and accuracy in their inferring process. However, weak in logic and justifiability may hinder their going further in both inference skills and argumentative writing. Therefore, in their later essay writing practices, they may be encouraged to pay special attention to these two higher inferring abilities.

The second line of Table 1 reveals the general argumentative writing ability of the participants. With a highest score 95, and a lowest score 64, participants had mastered the fundamental skills of persuasive essay writing, knowing that the organization, content, and language abilities are the essential elements. They also grasped the gist of Toulmin Model and knew most second language argumentative writing test rubrics such as essay rubrics of TOEFL test and other important English language tests. With a high correlation coefficient  $r=0.61$ , table 1 also indicates that the general quality of reference was closely correlated with the scores of argumentative essays, showing that, between these two variables, the progress in either inference or essay writing may facilitate the improvement of the other.

*Table 2 Inference Subskills In Relation to Grades of Arguments*

	Number	M	SD	r	Sig
Essay		84.68	6.75		
inference /a	188	2.29	0.93	0.52	0.00**
inference /b	88	2.28	1.11	0.56	0.00
inference /c	37	2.52	1.11	0.36	0.00
inference /d	19	2.15	0.96	0.74	0.00
inference /e	160	2.12	1.07	0.38	0.00
inference /f	13	2.31	1.18	0.61	0.00

\*\*p ≤ 0.01

The collection of the number of each inference subskill in Table 2 reveals that inference a, i.e. cause and effect (n=188) and inference e, i.e. probability (n=160), were participants' most favorite subskills in their writing practice whereas inference d (form, n=19) and inference f (analogy, n= 13) were least preferred among the participant. Each participant had used an average of more than 3 cause-and-effect inference, as "because" and "therefore" structure seemed to be practiced most in their previous writing and more accessible than other inference structures. Surprisingly, inference b (example, n=88) was not a most selected inference skill though it seemed to be easier to deal with. With a story happening to a person adjacent to a participant or a piece of relevant news would suffice, participants preferred more abstract reasoning than examples. Only 10 students used analogy to infer in their essays. The researcher interviewed 5 of those who did not apply this skill in their argument and their answers were similar: "I do not understand analogy", or "We don't use analogy often in Chinese." This indicates that they did not have adequate background knowledge about analogy and were not comfortable using this skill.

Table 2 also shows that inference c (commonplace M=2.52) was manipulated best among the students while inference e (probability M=2.12) achieved lowest average score. More than 30 participants used at least one adage or maxim in their writing and the relevance, accuracy, logic and even justifiability were well displayed in these commonplace inferences. In contrast, participants did not fare well in predicting the unknown from the known either because they were not good at using hedging so they used more boosters than they should, or because their predictions were based on wrong assumptions.

The connections between each subskill of inference and the quality of argumentative essays were core research item to investigate and elaborate in Table 2. Although inference d (form,  $r=0.74$ ,  $p \leq 0.01$ ) and inference f (analogy,  $r=0.61$ ,  $p \leq 0.01$ ) show very strong relations, they had too few samples ( $n=19$  and  $n=13$ ). Therefore, this study only discusses the other 4 elements. Causal (a) and example (b) inferences were significantly correlated with grades of argumentative essays, with respective  $r=0.52$ ,  $p \leq 0.01$ , and  $r=0.56$ ,  $p \leq 0.01$ , indicating that improvement in these two inference subskills may assist with the improvement of argument, and vice versa. The other two subskills, commonplace (c) and probability (e) were moderately correlated with quality of argument ( $r=0.36$ ,  $p \leq 0.01$  and  $r=0.38$ ,  $p \leq 0.01$ ), showing that maxims, adages, and prediction did not affect so much on the quality of an argument.

## 5. Conclusion

This study highlighted the need to assess the specific dimensions of critical thinking skills and selected one dimension- the inference skills – to experiment. Zarefsky's inferential patterns were adopted and adapted to form a performance framework of inference skills. After adaptation, we applied this framework to a group of college students who wrote argumentative essays on the same topic and found that inference skills are significantly correlated with quality of argument. Some other meaningful results were also obtained as following:

First of all, it is shown in the results that analogy was adopted least in students' writing, though it was a fairly common inference and rhetorical device in both Chinese and English language. Empowering students with more knowledge and practice of analogy is a task for second language writing teachers. Second, students in this study were weak in their prediction of probabilities and one important reason was that they were not familiar with the usage of hedging. Therefore, presenting more native samples to the Chinese students is one strategy to solve this issue. Third, results indicated that causal and example elements were remarkably correlated with quality of argument. Hence, if students, in their future second language writing classes, are trained better on how to use causal and example inference to improve their arguments, they would achieve higher grades on their essays.

However, we also acknowledge that there were limitations in this study. This modified inference performance framework had not been experimented in other studies; thus it still needs further research and modification by future studies. Moreover, strictly speaking, the rubric variables for assessing the level of inference ability – relevance, accuracy, logic, and justifiability-are categorical instead of continuous variables. The intervals between each two of them were difficult to stipulate. Therefore, it is essential to develop statistically more acceptable rubrics to measure categorical data.

## Acknowledgment

This study is supported by “Jilin University Undergraduate Teaching Reform Fund – Elite Program 2.0” (Fund Number: 2019XYB023). It is also supported by “Training program for top students in basic subjects” sponsored by Department of higher education, Ministry of Education. The project is entitled “The optimization of English education curriculum for top talents in the context of globalization” (Fund number: 20180401).

## References

- [1] Facione, P. A (1990). *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction*[M]. Millbrae,CA: The California Academic Press.
- [2] Paul, R.; Elder, L (2006): *Critical Thinking: Learn the Tools the Best Thinkers Use*[M]. Upper Saddle River, NJ: Pearson Prentice Hall.
- [3] Stapleton, P (2001). Assessing critical thinking in the writing of Japanese university students: Insights about assumptions and content familiarity[J]. In 4, no.18, pp. 506-548.
- [4] Stapleton, Paul; Wu, Yanming (2015): Assessing the quality of arguments in students' persuasive writing: A case study analyzing the relationship between surface structure and substance[J]. In *Journal of English for Academic Purposes*, no.17, pp. 12–23.
- [5] Toulmin, S (1958). *The uses of argument*[M]. Cambridge: Cambridge University Press.
- [6] Toulmin, S (2003). *The uses of argument*[M]. 2nd ed. Cambridge: Cambridge University Press.
- [7] Wade, C. (1995). *Using Writing to Develop and Assess Critical Thinking*[J], no.22, pp. 24-28.
- [8] Wen (2009). Developing a conceptual framework for assessing Chinese university students' critical thinking skills[J].Zarefsky, David (2005). *Argumentation: The Study of Effective Reasoning*[J]. 2nd. Virginia: The Great Courses, pp.57

