Critical Thinking

AND LANGUAGE LEARNING • CTLL

LUIS F. SANTOS • ADAM GYENES
• SANAE FUJISAKI • SEAN GAY
• GRETCHEN CLARK • HIROSHI NAKAGAWA

Critical Thinking in Language Learning • CTTL
The Journal of the JALT Critical Thinking SIG • CT SIG
The Japan Association for Language Teaching • JALT
Volume VIII, Issue 1, December 2021.

Find out more at:
http://www.jaltcriticalthinking.org.

Exploring Fairminded Critical Thinking in Argumentative Writing with Chinese EFL Users: Teaching Model and Rubric Design

LUIS F. SANTOS
NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF ECONOMICS

ADAM GYENES
RYUTSU KEIZAI UNIVERSITY

Abstract: "Fairminded critical thinking" (FCT) is an alternative to traditional critical thinking conceptions as it places greater emphasis on ethical reasoning. Argumentative writing can provide a means to observe and cultivate such thinking qualities. This study introduces a non-linear instructional model to teach FCT, and through a small-scale intervention, explores its potential for developing four FCT intellectual standards in argumentative writing (intellectual accuracy, depth, breadth, and logic) with a group of Chinese university English (EFL) users. Adopting a one-group, pretest-posttest exploratory design, writing samples were assessed with a locally created rubric tool which was also triangulated through an adapted content analysis. Overall, pretest writing samples showed that some thinking skills had already been developed, yet there was a tendency toward one-sided argumentation and the use of informal evidence to justify claims, frequently culminating in a persuasive defense of one position. On the posttest, logic was the only intellectual standard that showed a more noticeable improvement, potentially attributed to the teaching model, representing development of a more moderated, multi-sided articulation of logic. These findings provide impetus for a discussion of how FCT can be further taught and assessed. Theoretical and methodological implications of FCT as a non-traditional approach to critical thinking and argumentation are discussed.

Keywords: Argumentative writing; Critical thinking; accuracy; Logic; Depth; Open-mindedness

Introduction

Education has been considered through many iterations of civilization to be the foundation for a balanced society. For many critical thinking scholars, this demands a wider conception of the development of human intellect that transcends conventional educational models (Davies & Barnett, 2015, pp. 9-19; Dewey, 1933; Facione, 2015, p. 24; Kelly, 2004, pp. 74, 88). Fairminded critical thinking (FCT), as an object of education, provides one such alternative understanding, and suggests an approach to the cultivation of intellect that emphasizes rationality toward critical/rational societies (Elder & Cosgrove, 2019; Paul & Elder, 2012). Yet education systems in many parts of the world continue to rely on policies and practices that neglect 21st-century competencies, among them critical thinking (CT). While education levels are rising in many parts of the world, serious social issues persist, and new challenges to rationality and ethics are appearing; critical issues that cry out for a critical society: "a community of people who value critical thinking and value those who practice it" (Elder & Cosgrove, 2019). In the shadow of multifarious 21st-century problems, the distinction between an education that develops intelligent people and one that cultivates balanced, rational thinkers is all the more clearly illuminated (Halpern, 2014, pp. 31, 244; Kahneman, 2011, p. 49; Stanovich, 2011).

Thinking is a complex construct to observe, but it can

be manifested for instructional purposes through thinking products such as written and spoken discourse (Ku et al., 2014; Paul & Elder, 2007; 2005, p. 19). While speech acts are interactional, spontaneous and exist in a moment of time, argumentative writing represents for many the most suitable means to observe and evaluate CT virtues (Doherty et al., 2019; Jonsson & Svingby, 2007; Liu & Stapleton, 2014; Liu, Frankel, & Roohr, 2014; Paul & Elder, 2007). To this end, this study proposes an original perspective for the teaching and assessment of an alternative approach to CT based on Paul and Elder's (2012, 2014) framework. Focusing particularly on four thinking standards (intellectual accuracy, breadth, depth, and logic), we explore the use of these standards in teaching practice with a small group of Chinese university students, and assess FCT in their writing samples through an original rubric design and content analysis method.

Literature Review

Fairminded critical thinking

Paul and Elder's conception of FCT (2012, Elder & Paul, 2013) includes core components of any prominent CT conceptualization such as analysis, evaluation, and synthesis (Facione, 1990; Paul & Elder, 2012, pp. xix, xxiv, xxxi). Moreover, it applies intellectual standards such as intellectual accuracy, depth, breadth, logic, and clarity (pp. 91, 100) that prompt reflection on the quality of thought. Most striking is the emphasis put on ethics in reasoning,

which is the essential meaning of 'fairmindedness' in this conceptualization. This is expressed through intellectual traits or virtues; including an intellectual sense of justice, humility, empathy and autonomy, among others (Elder & Paul, 2013, p. 43). In this way, their approach to CT explicitly highlights and challenges human tendencies that oppose rationality such as egocentrism and sociocentrism (Elder & Paul, 2013, p. 30), necessitating a consciousness of bias, self-deception, and self-serving tendencies in human thought and action (Paul & Elder, 2012, pp. 1, 6).

In doing so, Paul and Elder (2012) separate critical thinkers in a cognitive sense from critical thinkers in a fairminded sense. The former are highly skilled intellectual thinkers but not necessarily fair or ethical, and are termed as "weak-sense critical thinkers" (p. 300). Meanwhile, the latter are not only highly effective thinkers, but measure their own thinking against intellectual and ethical standards, and are termed as "strong-sense critical thinkers" (p. 301). As Higgins (2014) observes, an emphasis on "logically valid arguments" while dismissing ethics or values leads to sophistry or an unethical use of premises.

While FCT is comprehensively articulated by Paul and Elder, similar CT conceptions have been approached by different authors, mainly in reference to "fair-minded objectivity" (Facione & Gittens, 2016, p. 89); the recognition of bias in one's own beliefs ("confirmation bias"), and the ability to reevaluate an argument in light of better evidence

or reasons (Halpern, 2014, p. 243-245).

Watson (1925) was perhaps the first to propose fairmindedness in critical thinking explicitly, developing a "test of fair-mindedness" in 1925, focused on exploring test-taker's prejudicial tendencies concerning economic and moral topics. This involved (i) weighing evidence, (ii) matching factual statements with conclusions, (iii) observing moral judgments, (iv) bias tendency, and (v) generalization tendency (Rainey, 1926). Years later, Paul adapted these components in his own fairmindedness test (Fisher, 1991), the core components of which are: impartiality, truth seeking, and open-mindedness or breadth of thought through recognition and tolerance of opposing beliefs (Paul & Elder, 2007).

Four intellectual standards of FCT

Under the FCT approach, intellectual accuracy means to "obtain conformity with fact or truth" (Paul and Elder, 2012, p. 430). In argumentation, this conception is commensurate with notions such as credibility and reliability of the premises (Fisher, 2011, pp.24, 60, 84-86); "acceptability" or "trustworthiness" (Halpern, 2014, pp. 246, 248); "validity, evidential warrant, and consistency" (Lipman, 2003, p. 213); healthy skepticism and withholding judgment (Ennis, 1985; Facione & Gittens, 2016, p. 134; Halpern, 2014, p. 258; Norris, 1992); using reliable or credible sources (Ennis, 1985; Norris, 1992), and supporting opposing positions with

relevant evidence (Kuhn, 1991, pp. 51, 268; Halpern, 2014, p. 243).

Breadth of thought refers to considering an issue from every relevant alternative viewpoint, giving fair consideration to every perspective (Paul & Elder, 2012, p. 98). This may also encompass "considering alternative interpretations" (Norris, 1992), or "conceiving an alternative theory" and "counterarguments" (Kuhn, 1991, pp. 51, 266), or recognizing "missing components" and being willing to abandon "cherished belief[s] when given good reasons to do so" (Halpern, 2014, pp. 244- 246).

Depth of thought refers to recognizing the complexities of an issue beyond simplistic answers (Paul & Elder, 2012, p. 97). This conceptualization relates to recognizing "the unseen part of the argument" (Halpern, 2014, p. 246), the provision of causes or antecedents at different levels in an argument until "genuine evidence" is reached (Kuhn, 1991, pp. 45, 51, 266) or until "basic reasons" or "basic premises" are achieved (Fisher, 2004, p.15). These are reasons beyond the first or second responses to a "why" question (Facione & Gittens, 2016, pp. 90-91), which can also be called root reasons.

Finally, Paul and Elder (2012) conceptualize logic as organized "mutually supporting thoughts" or ideas that make sense in combination (p. 99) and equate to a "relationship between the premises" (Halpern, 2014, p. 246) or the relationship between the conclusion and the premises in an

argument (Fisher, 2011, pp 60, 84-86).

Critical thinking in the Chinese context

A number of studies have reported concerns about Chinese students' CT competencies (e.g., Chan, Ho, & Ku, 2011; Ku, Lai, & Hau, 2014; Pei et al., 2017; Qian, 2015; Tian & Low, 2011; Xiaobei, 2017), while there are others that have taken a more optimistic view (Chen, 2017; Davies & Barnett, 2015, p. 1; Hernández, 2016). Contextual and pedagogical aspects of Chinese education such as teacher-centered practices and a rigorous examination system that promote rote-learning and rigid thinking are among the perceived barriers to CT development in China (Dong, 2015, pp. 356, 358; Liu & Stapleton, 2014; Shaheen, 2016; Tan, 2017; Tian & Low, 2011). Moreover, deferment to authority in Chinese and other Asian cultures influenced by Confucianism is perceived to hinder CT (Chan et al., 2011; Davies & Barnett, 2015, p. 1; Dong, 2015, p. 357; Fung & Howe, 2014; Ho, 1994; Li, 2012; Paton, 2005). Current socio-political circumstances in China add another dimension to this (Xiaobei, 2017; Zhang, 2016).

Research has pointed out particular concerns regarding Chinese students' CT competencies, such as limited autonomous/independent thinking, linked to freedom of speech (Fung & Howe, 2014; Pei et al., 2017; Tan, 2017; Tian & Low, 2011; Zhang, 2016), dogmatism toward truth and knowledge (Dong 2015, p. 361), as well as fixed conceptions of knowledge and a lack of consideration of

multiple perspectives (Chan et al., 2011; Ku et al., 2014). For instance, Ku et al. (2014) report that 92% of Hong Kong Chinese students did not generate rebuttals, whilst 75% did not provide counterarguments in an argumentation task when exposed to an authority figure. Liu and Stapleton (2014) report a similar tendency while, Pei et al (2017) report that only 35% of research participants challenged an authoritative viewpoint, while a lack of "flexibility", and "profundity" were other noted weaknesses.

Research Questions

Considering findings such as these, this study sought to explore the interplay in teaching and assessment of logic, intellectual accuracy, breadth, and depth of thought under the fairminded conception of CT, as observable components of argumentative writing with a group of Chinese students. We hypothesized that explicit instruction with active learning/ student-centered pedagogies, can help to improve at least some of these traits, as such pedagogies have produced gains in CT skills and dispositions in previous experiments (e.g., Abrami et al., 2015; Cargas et al., 2017; Dwyer et al., 2010; Ku et al., 2013; Liu & Stapleton, 2014; Nold, 2017). A mixed method, exploratory design was adopted to enlighten the central research questions:

Q1. How do the research participants perform in terms of logic, intellectual accuracy, breadth, and depth before the teaching intervention?

Q2. To what extent can a 15-hour explicit CT instructional program utilizing a non-linear teaching approach, produce gains in participants' argumentation performance on logic, accuracy, breadth, and depth of thought, measured via two non-standardized locally created assessment tools?

Method
Research Design

This study adopted a one-group, pretest-posttest exploratory design (Gall et al., 2015, p. 306). The effects of the teaching program were explored through participants' performance on pretest and posttest argumentative writing tasks, measured with both a locally created rubric and content analysis focused on the four thinking standards. With an exploratory scope, conclusive observations are not necessarily sought, and the aim is to explore the efficacy of one pedagogical approach in nurturing the four FCT markers.

Participants

The participants in this study were 12 Chinese EFL university students (2 male and 10 female), from the same university in Mainland China, aged 21 to 25 years old. They included eight graduate students and four bachelor students in social science fields. This sample was made based on locally availability: participants recruited via open announcements on campus noticeboards. The 12 participants who completed the program (eight others did not) matched the intended

sample size of 10 to 20 students. The participants received no reward for their participation, and their motivations included interacting with a foreign researcher, practicing English skills, and learning more about CT. Two students reported an English level equivalent to IELTS writing bands 5-5.5, seven were in the 6-6.5 band, and three at 7-7.5. These levels were reported by participants based on recent tests, and could be verified in some cases. These levels generally matched each student's spoken and written competence, during the class.

Due to the limited number of available research participants, there was no chance to include a control group. No volunteers were found to form a control group for this study. The 12 participants were willing to be part of the experimental group rather than a control group because they found more advantages in the treatment process in terms of experiencing the activities around the new CT conception and others. We also considered that performing the experiment with only six students in the treatment group, if we had split the participants into two groups, would have been less significant than doing it with 12.

Setting and material

The intervention took place in meeting rooms which were equipped with an audio-and-video system. Provided materials included: reading materials, videos, paper, and markers. Edited video clips of debates on controversial

social topics such as climate change, GM food, the effects of artificial intelligence in the job market, and causes of poverty in developing countries were used to introduce topics and argumentative writing tasks during the instructional process. Transcripts from the videos were also provided in both English and Chinese. In addition, Excel was extensively used for drawing argument maps, as these were frequently developed with the students as a teaching-learning technique. A rubric for assessing the quality of the learners' FCT performance was designed and presented to students.

Procedures

Using English as a means of instruction (without explicitly teaching language), the 15-hour syllabus was implemented in six 2-hour sessions, and one 3-hour session. The group met once a week over seven weeks in September and November, 2019. Argumentative writing tasks served as pretest and posttest assessments, using rubric-based assessment (RBA) and content analysis methodology. For both the pre-and-post tests, the participants were required to write 20-minute free-written argumentative essays, using "questions of judgment" as prompts in order to evoke responses to social issues (Paul & Elder, 2012, pp. 127-128).

The teaching model implemented was based mostly on student-centered pedagogies involving Socratic questioning, argument mapping, writing tasks, rubric-based learning, and feedback (Dwyer, 2015, 2010; Freire, 2005; Lee, 2018;

Paul & Elder, 2012). This could be classified as a domain-general CT course, embracing mixed teaching methods (Nold, 2017; Abrami et al., 2015; Ku et al., 2014; Alwehaibi, H., 2012; Smith et al., 2018). The teaching model aimed for a comprehensive approach, not limited to traditional cognitive and dispositional realms, but reaching ethical, civic, and cultural dimensions of CT (Davies & Barnett, 2015; Santos, 2020). These latter dimensions were addressed by emphasizing the conception of fairmindedness in critical thinking, by confronting students with their social reality and challenging it, as well as by recognizing the participants' cultural teaching and learning styles.

Figure 1 shows the non-linear teaching-learning moment's framework, which was the basis of the pedagogical design. In this model, the activities/moments do not necessarily follow a sequential pattern. The purpose was creating a bridge between the teaching-learning moments based on instructor's insight (Kelly, 2004, pp. 64-81), to overcome potential issues of a rigid sequential model, (Ku et al, 2014; Kuhn, 2007, p. 112). The implementation of this approach is elaborated in Table 1.

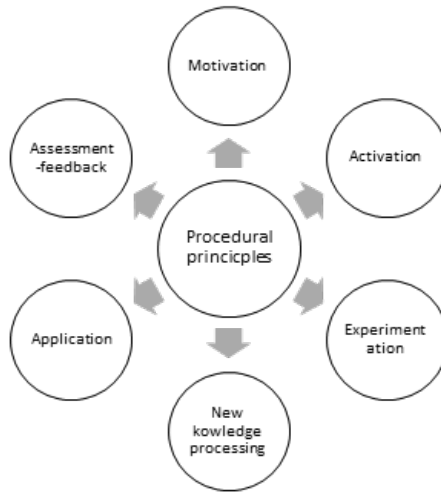Figure 1. The non-linear teaching-learning moments' framework

Table 1. Implementation of the non-linear teaching-learning moments' framework

Teaching-learning moments	Activities
Motivation	- Watching thinking-engaging videos. - Highlighting the importance of CT via PPT (presenting images and reflections).
Prior knowledge activation	- Socratic questioning (based on casual and pre-elaborated questions). - Question posing via cards for anonymous answers.
Experimentation	- Prompting- argumentation tasks and videos. - Reading argumentation samples/models. - De-/constructing an argument via argument mapping. - Feedback provisioning. - Teacher-student pair-group role-playing. - Joint analysis and writing.
New knowledge/information processing	- Reflecting on prior experience via questioning. - Direct instruction/presentation of concepts. - Procedural demonstration of target skills, virtues and standards via lecturing, PPT, argument mapping, and casual questioning. - Individual and assisted reading of new information and argumentation samples/models.
Application	- Writing argumentation exercises based on controversial topics/prompts. - Constructing an argumentation via argument mapping. - Providing feedback.
Assessment and feedback	- Rubric-based self-assessment and feedback. - Rubric-based peer-assessment and feedback. - Rubric-based instructor-assessment and feedback.

Instruments

Data was collected through two 20-minute argumentative writing tasks prompted before and after the intervention, which served as the pre-test and post-test: a constructed-response or open-response-based test (Livingston, 2009; Popham, 2003; Tankersley, 2007). Twenty minutes were allotted for the tasks, while providing time for other teaching activities within the sessions. The participants were prompted to write spontaneously regarding controversial social issues. The topic for the pretest was "Does Artificial Intelligence or new technology create or eliminate jobs?" For the posttest, a variety of topics were posed by the participants through a collective reflection: they were generated following Freire's "problem-posing methodology" (Freire, 2005, p. 79-82), from which participants compiled a list (Appendix D) and then randomly picked one topic each to spontaneously write about, with the aim of making the assessment process democratic.

Participants' thinking qualities -accuracy, depth, breadth, and logic -were assessed using two locally created, non-standardized measures. These were a five-level analytical rubric tool (RBA), and a list of CT qualitative indicators for content analysis-based assessment (Appendices A and B). These two assessment tools adopted four intellectual standards from Paul and Elder's fairminded CT approach, which in essence are commensurate with the common criteria implicit in argument development and evaluation with a

focus on CT (AACU, n.d.; P. Facione & Gittens, 2016, pp. 24-25,89; Fisher, 2004, pp. 24–25, 2011, pp. 60,84-86; Halpern, 2014; Kuhn, 1991; Paul & Elder, 2014, p. 127; Pearson, 2012; Watson, 1925); yet the particularity in this fairminded argumentation approach is the emphasis given to two-sided reasoning and intellectual honesty, as noted in the literature review and shown in the rubrics (Appendices A and B). These instruments were locally/contextually designed to address the specific characteristics of the students, particularly following Rezaei and Lovorn's (2010) recommendations about avoiding the use of decontextualized measures. Based on this, writing and linguistic skills including grammar and spelling were overlooked, placing emphasis on the participants' ideas, content, or thinking qualities.

Both assessments supported analysis of verbal data that could be transformed and evaluated numerically. Both the RBA and the content analysis were carried out with the purpose of triangulating findings in order to reduce the subjectivity inherent in these evaluation methods, following the suggestions of Williams and Lahman (2011), Ku et al. (2014), and Dwyer et al. (2015).

In addition, the rubric tool attempts to characterize the three types of thinkers described by Paul and Elder (2012), namely "uncritical thinker", "weak-sense critical thinker", and "strong-sense critical thinker" (pp. 299-301). In this sense, the rubric proposes 5 levels, where level 1 represents "uncritical thinker", level 3 "weak-sense critical thinker",

and level 5 "strong-sense critical thinker". As Paul and Elder note, weak-sense critical thinkers do not necessarily lack cognitive critical thinking skills, yet have little inclination toward fairmindedness (pp. 3-4).

Newman (1995) and Williams and Lahman (2011) served as references to design the content analysis-based assessment, which involved coding and marking "textual indicators of critical thinking" (Newman, 1995) and their opposites in the students' essays based on the four thinking standards (see Appendix B). Afterwards, those marks were counted under each category following the same procedure indicated by Newman (1995). This procedure was carried out on both pretests and posttests by the first researcher, while three raters were used for the RBA.

Validity and reliability

Cronbach's Alpha computed with the results from grader 1 via SPSS yielded values of .91 for pre-test and .94 for post-test, representing high coefficients and thus an excellent internal consistency in the pre-test and post-test. None of the item-total correlations came in below .55 in either test, and only one criterion in the posttest (Breadth) would slightly improve Cronbach's Alpha if deleted (by .0003). Meanwhile with the RBA, Cronbach's Alpha computed with the mean results from three raters yielded .91 for pre-test and .97 for post-test, also representing excellent consistency.

Intraclass Correlation Coefficient (type Two-Way Mixed)

computed with the mean gradings from the three raters that conducted the RBA was 0.46 for the pretest and 0.50 for the posttest, indicating low to moderate inter-rater reliability (Hamp-Lyons, 1991; Rezaei & Lovorn, 2010; van Helvoort, Brand-Gruwel, Huysmans, & Sjoer, 2017).

For the sake of attaining content validity, the rubric tool underwent revision by three independent experts before use in this research. Additionally, students in a pilot study provided feedback regarding the content of the rubric before its final use. The final rubric was therefore developed over a period of two years (2017-2019).

Triangulation of results of the rubric and the content analysis was intended to reduce the degree of subjectivity inherent in these non-standardized, open-response measures. The triangulation consisted of observing the results from the descriptive statistics and inferential statistics of the RBA, along with the results from the content analysis system. This process offered the possibility to contrast numerical and qualitative data and to confirm tendencies. The results from three raters using the rubric tool were averaged to generate fair judgment, echoing van Helvoort's et al. (2017) recommendation.

Data analysis

A comparison of results between test 1 (a baseline) and test 2 (an immediate-after-intervention posttest) was conducted through the Wilcoxon Signed-Rank test - the

non-parametrical test equivalent to pair-sample t-test -via SPSS, as data did not satisfy the conditions of normality (see Appendix C for numerical and graphical normality analysis/results). This test was applied using the results stemming from the RBA. As for the content analysis, a simple counting or tabulation of positive and negative indicators of CT was carried out in Excel. Results were not broken down by gender to analyze data discretely due to the small sample and the unrepresentative number of male participants (n = 2). In this way, a baseline of participants' four thinking standards could be established, while the efficacy of the experimental teaching design to produce gains in the four thinking standards could be measured, in response to the central research questions.
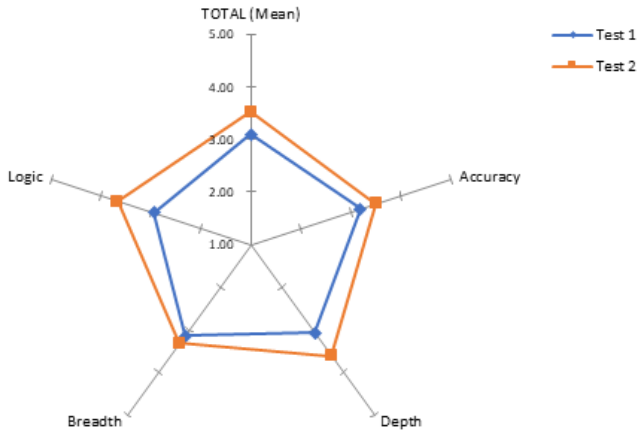
Results

RBA Descriptive statistics

In general, the results from the descriptive statistics showed an increase from test 1 (baseline/pre-test) to test 2 (immediate post-test) in the four intellectual standards (see graphic 2). The most striking variations were observed in the results for "logic" with the highest positive difference, and "breadth of thought" with the lowest positive difference. The general average in this 1-to-5-point scale of test 1 was, M = 3.09, SD = 0.68; and in test 2, M = 3.52, SD = 0.70, which denotes a 0.43 point difference. Table 1 shows the mean (M), standard deviation (SD), and mean difference for each thinking standard.

Table 2. Descriptive statistics

		Mean	N of cases	Std. Deviation	Mean difference
Pair 1	Logic-Test 1	2.94	12	0.91	0.72
	Logic-Test 2	3.67	12	0.79	
Pair 2	Accuracy-Test 1	3.19	12	0.78	0.31
	Accuracy-Test 2	3.50	12	0.77	
Pair 3	Depth-Test 1	3.08	12	0.64	0.53
	Depth-Test 2	3.61	12	0.62	
Pair 4	Breadth-Test 1	3.14	12	0.70	0.17
	Breadth-Test 2	3.31	12	0.74	

Figure 2. Comparison between ratings of test 1 and test 2 – Descriptive statistics

Content analysis

Results from the content analysis are less striking when compared with the descriptive statistics, except for logic, which shows a significant increase from test 1 to test 2. The only thinking standard that showed discrepancy between the RBA and the content analysis is depth of thought, whose value showed a slight decrease in test 2 compared with test

1, which had shown a notable increase in the descriptive statistics.

Table 3. Total number of positive and negative indicators of CT

	Test 1 (n)		Test 2 (n)		Test 1 (%)		Test 2 (%)		Ratios	
	+	-	+	-	+	-	+	-	Test 1	Test 2
Logic	13	27	36	12	33%	68%	75%	25%	-0.35	0.50
Accuracy	20	10	41	15	67%	33%	73%	27%	0.33	0.46
Depth	15	3	25	6	83%	17%	81%	19%	0.67	0.61
Breadth	9	3	10	3	75%	25%	77%	23%	0.50	0.54
Total	57	43	112	36	57%	43%	76%	24%	0.14	0.51

Note. *n* = Number of positive and negative instances observed/counted. *Ratio* = Value converted from the positive and negative counts to a ratio in the scale, -1 to +1.

Figure 3. Comparison between ratios of test 1 and test 2 – Content analysis method
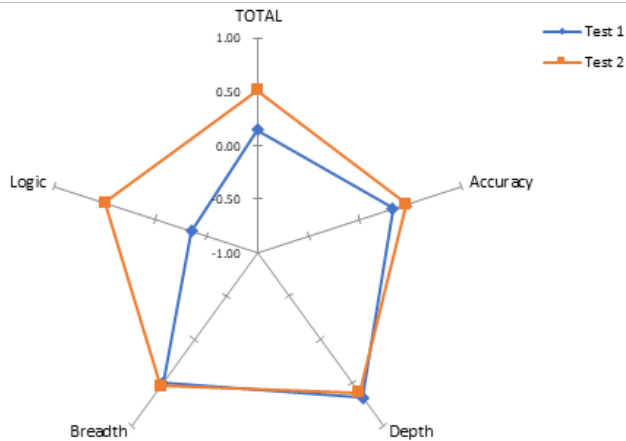
Wilcoxon Signed-Rank test

The individual comparison between tests 1 and 2 of each thinking standard conducted through Wilcoxon Signed-Rank test indicated that there was a statistically significant difference in the results for logic, Mdn = 3.67, Z = -2.01, p < .05, effect size r = 0.18. However, no statistically significant

difference was observed for accuracy, depth, and breadth of thought; or for general mean scores from both tests, as shown in Tables 4 and 5, respectively.

Table 4. Wilcoxon test output of the comparison between two testing times of individual thinking standards

	Logic_Test2 - Logic_Test1	Accuracy_Test2 - Accuracy_Test1	Depth_Test2 - Depth_Test1	Breadth_Test2 - Breadth_Test1
Z	-2.01	-0.79	-1.73	-0.54
Asymp. Sig. (2-tailed)	0.04*	0.43	0.08	0.59

* Significance of p at the .05 level

Table 5. Comparison between general results in two tests

	Mean	Median	N	Std. Deviation	Std. Error Mean	Z	Asymp. Sig. (2-tailed)
Total average_Test_2	3.52	3.5	12	0.70	0.20	-1.33	0.18
Total average_Test_1	3.09	3.38	12	0.68	0.20		

Discussion

An overall comparison of total mean scores between the pretests and posttests show no significant statistical difference. Moreover, individual analysis indicates that participants' performance in logic is the only intellectual standard with statistically significantly gains after intervention. Meanwhile, no statistically significant changes were found in intellectual accuracy, depth, and breadth, although the descriptive statistics and qualitative analysis both indicate slight improvements in all intellectual standards. Implications regarding each thinking standard, participants' traits and tendencies, and the effects of the instructional model are described over the next four sections

in response to the research questions.

Logic

Triangulating the results of the RBA and content analysis, a gain could be confirmed regarding participants' performance in logic. This suggests that participants improved the logic of their written argumentation following the intervention and likely as a product of it, which infers some degree of effectiveness in the teaching model to boost this particular standard, at least in the terms measured here. Effectiveness of the teaching approach in this regard can perhaps be attributed to the frequent use of argument mapping. Dwyer et al (2010) also found "logical reasoning" to be effectively developed by such an approach.

The RBA shows an increase from, M = 2.94 on test 1 to M = 3.67 on test 2 for logic in the 5-point scale, representing a difference of 0.72. This may suggest an important gain after rounding, given that the qualitative characteristics that a single point represents on the 1-to-5-point scale differ substantially. Hence, 3 represents a logical flow of ideas but a weak disposition toward the inclusion of opposing ideas, whereas 4 shows a conscious effort to represent opposing views to some extent. This is consistent with Nejmaoui's (2018) research, which reports a similar gain found in 36 Moroccan EFL university learners, who after explicit instruction in CT, showed a statistically significant increase in their performance from a below-average level

in the pretest to an average level in the posttest, where CT skills assessed included "addressing alternative arguments", and maintaining "the logical flow of ideas" tested through an argumentative essay and rated via the Illinois Critical Thinking Essay Scoring Rubric (Finken, 1992). In another study with 110 Chinese English majors, Pei et al (2017) determined a score equal to 3.47 in the category "logicality" observed via an argumentative writing test and a 1-to-5-point holistic rubric scale where "1 = poor, 3 = average, 5 = excellent." Pretest results in Pei et al.'s study (M = 3.47) and this study (M = 2.94) for logic indicate that, although not high, students skills related to logic in written argumentation were not especially low either, indicating this skill is present or has been previously developed to some extent in Chinese university students. Pei et al.'s study however, did not focus on FCT or emphasise the fair inclusion of opposing ideas.

Intellectual accuracy

There was no statistically significant change in "accuracy," although the descriptive statistics from the RBA present a slight increase from the pretest, M = 3.19 to the posttest, M = 3.50, which is consistent with data from the content analysis. This finding may reflect some challenges. For instance, the difficulty faced by learners to express accurate thinking from a fairminded perspective in written argumentation using EFL, or the potential ineffectiveness of the teaching program for strengthening this standard, perhaps, given the limited time

available. Course length and learning time are considered to play a central role in cultivating CT, as Nejmaoui (2018), Dwyer et al., (2015), and Williams and Lahman (2011) have observed.

It should be noted, as well, that these results were generated in a context where the topics in the post-test writing task were not covered during instruction or prepared in advance. This made it more challenging for the participants to perform the tasks with higher accuracy. Performances are likely higher in contexts where argumentative writing tasks are developed with topics that were somewhat prepared beforehand, especially in terms of finding sources of information to support claims—or to develop evidence-based argumentation—and broaden perspectives.

In addition, the non-statistical difference reported here reaffirms somewhat the tendency found by Kuhn (1991) that most people reason using "pseudoevidence" (p.44) as opposed to "genuine, high-quality evidence" (pp. 94, 232), resulting in poor integration of counter arguments or alternative perspectives, and weak recognition of knowledge and beliefs as tentative and subject to evaluation and change (pp. 94, 183, 264-266). Note that a score of 3 on the RBA 5-point scale represents a weak disposition to include opposing views, yet strong skills for defending one single position. This includes skillfully providing reasons for the sake of winning arguments or defending one's own position (Paul & Elder, 2014, p. 2; Waller, 2012; Walters, 1994).

Nevertheless, while this group of students' fairminded accuracy generally did not receive favorable evaluations, these tendencies are observed across cultural and educational contexts, as a general human trait (Kahneman, 2011, pp. 209-220; Halpern, 2014, p.41; Paul & Elder, 2012, p. 227 and 2014, p. 203) and may not necessarily be specific to Chinese university students. For instance, Kahneman (2011) explains that a persuasive argumentation can be elaborated with highly coherent or logical reasons but with narrow, weak, or false evidence causing an illusion of truth or "cognitive illusion" (p. 212). He suggests that this is a natural tendency of the human mind (system 1) which, among other aspects, is characterized for making quick associations based on previously known beliefs or knowledge that "elaborate accounts of the reality" or create "a very good story," though not necessarily a true story, resulting in a tendency toward intellectual inaccuracy (pp. 209-220).

Yet in other studies, cultural aspects are relevant to the findings discussed here. For example, Ku et al. (2014) reported a tendency toward disregarding the inclusion of counter-reasons (75% of participants) and rebuttals (92%) in a written argumentation task by Hong Kong Chinese undergraduates after being exposed to an authority figure (in the experimental group). Likewise, Chan et al. (2011) found a similar trend among Chinese students. Echoing Ho's (1994) assertion, Chan et al. (2011) and Ku et al. (2014) suggest that Confucian cultures, where authority

is venerated, may influence beliefs about knowledge to negate counterarguments and two-sided reasoning. This characterizes "naïve belief" associated with a perception of the nature of knowledge as certain and absolute (Chan et al., 2011). Similarly, Dong (2015) asserts that dogmatic tendencies toward truth and knowledge persist in Chinese society (p. 361), and thinking from multiple perspectives is not fostered in the education system (p. 356), as Zhang (2016) concurs. These views could be interpreted to corroborate the findings of this study to some extent regarding Chinese university students' tendencies concerning fairminded accuracy.

Depth of thought

No statistically significant gain is observed on learners' depth of thought, while participants' ratings from the RBA show an increase of 0.53 scale-points in the posttest (pretest M = 3.08, posttest M = 3.61). Conversely, the content analysis reflects a slight decrease in the posttest from the pretest (graphic 3). This implies that the training design was unable to produce significant changes in participants' depth of thought. Again, limited training duration or lack of targeted practice may be the causes. Similarly, the use of EFL instead of L1 and the challenge of performing a task spontaneously (without previous preparation) could have limited a deeper argumentation. The results may also reflect some challenges in the participants' natural abilities to support or justify a

claim within well-organized lines of vertical reasoning. For example, Kuhn's study (1991) found that most participants tended to provide superficial responses when prompted to develop an argument (pp. 44-45, 210). Nevertheless, in the present study, participants' average rating of M = 3.35 (RBA), is closer to 3, which in this rubric scale represents skillful depth of thought without fairmindedness; namely, "weak-sense critical thinking".

This result coincides somewhat with Pei's et al (2017), where participants' "profundity" also obtained an average rating of M = 3.35 in a 1-to-5 scale. Again, the rubric's constructs in that assessment did not focus on FCT. This implies that research participants' depth of thought in both studies with Chinese university students is neither null nor strong. Although the increase on the RBA of 0.53 in the present study shows that the teaching program may have had some positive effect on the depth of thought standard, more time or more specific focus on this aspect could have improved this further.

Breadth of thought

Components of breadth of thought are observable within the other three thinking standards as it can be considered an essential quality to fairmindedness (Paul & Elder, 2007; 2012, pp. 1-6). Nevertheless, according to the rubric's description (see Appendix A) mastery of this standard puts emphasis not only on the inclusion of an opposing

perspective, but on the recognition of at least one merit in it, or the recognition of one demerit in the arguments used to justify the supported position. This needs to be interpreted critically, since an opposing position can be constructed as a "straw-man argument," and thus may in fact foment selfish thinking, sophistry, or "weak-sense critical thinking" (Paul & Elder, 2012; pp. xxvii, 2-4; Higgins, 2014).

The results from the content analysis and the descriptive statistics show a narrow rise regarding participants' breadth of thought displayed on their essays in the posttest, while the Wilcoxon test indicates no statistical significance between the pretest and posttest results. Potential causes again could include ineffectiveness of the teaching program, insufficient time, or difficulties faced by participants related to the use of EFL instead of L1 and the challenge of performing a task spontaneously (without previous preparation), as noted earlier.

As for training time, Liu and Stapleton (2014) proved the efficacy of a 12-week intervention for improving students' essays counterargument and rebuttal integration that contrasts the seven-week intervention here. Meanwhile, as mentioned in the discussion under "accuracy," certain literature points to inherent human tendencies (Kahneman, 2011; Lee, 2018; Halpern, 2014, p. 41; Paul & Elder, 2014, p. 280) while others identify cultural/contextual and pedagogical factors specific to Chinese students (Chan, Ho, & Ku, 2011; Tan, 2017; Dong, 2015; Zhang, 2016; Davies

& Barnett, 2015, p. 1; Fung & Howe, 2014; Ku, Lai, & Hau, 2014; Liu & Stapleton, 2014) which could be at play here in this modest performance. Eventually, the limitations of Chinese university students concerning integration of counterarguments or breadth of thought deployment observed in the present study are consistent with prior research (e.g., Chan et al., 2011; Ku et al., 2014; Liu & Stapleton, 2014; Pei et al., 2017).

Practical Implications

The findings suggest that the instructional practices used in this study may foster thinking standards such as logic, and with more training time, intellectual accuracy, breadth, and depth of thought. Under this condition, practitioners may confidently utilize techniques in CT education such as argument mapping and Socratic questioning, paired with formative feedback to boost these intellectual standards. In particular, argument mapping was highly effective in developing analysis, evaluation, and synthesis based on argumentation. Besides, rubric-based instruction-assessment promotes self and collaborative assessment, facilitating learners' engagement in autonomous learning and reflective thinking that could be focused on CT/FCT development.

Limitations and future research

As a quasi-experimental, one-group, pretest-posttest exploratory research design with 12 participants and no control group, generalizability is beyond the scope of this study. The rubric's reliability coefficients are practically low, and triangulation with the content analysis system may still be imprecise due to small sample size, insufficient rater training, the inherent complexity in assessing thinking; or the unreliable nature of non-standardized open-ended tests, as some authors have noted (e.g., Bell et al., 2013; Doherty et al., 2019; Insight Assessment, 2019; Jonsson, 2014; Liu et al., 2014; Rezaei & Lovorn, 2010).

Moreover, the rubric tool designed for this study may be wordy and impractical for other contexts, while extensive descriptions of performance are needed in rubric design to capture thinking quality with fidelity (Dwyer et al., 2015; Jonsson & Svingby, 2007; Landis et al., 2007, pp. 141–142; Tiruneh et al., 2016). Furthermore, there may be particular thinking characteristics displayed on the argumentation tasks not exactly captured or represented in any of the categories throughout this rubric tool. Finally, the type of questions used as writing prompts presented a challenge for assessment, since some questions or topics call for two-sided reasoning in a way that is clearer than others, affecting participants' performance.

Further research may consider refinement of the rubric tool, subjecting it to a more thorough validation process.

Furthermore, any replication of the teaching intervention and assessment processes in future research should consider the inclusion of a larger sample size and a control group. A longer training period is also recommended to see potential gains, particularly with regard to accuracy and breadth of thought. Rubric-based assessment is time-consuming; thus larger scale studies need to consider the inclusion of a team of well-trained raters. Finally, comparative studies may be interesting to conduct to see if there are differences among groups from different contexts and connections between contextual/cultural factors and traits such as intellectual accuracy, breadth of thought or epistemological development.

Conclusion

Participants' performance regarding accuracy, depth, and breadth of thought displayed in the argumentative writing tasks using EFL tended toward weak-sense CT before and after the teaching intervention. The teaching program did not produce statistically significant changes, either as a whole, or in three out of the four thinking standards explored. However, there was a significant improvement in logic after the intervention from level 3 to level 4 (rounded score). This gain means that compared with the pretest, participants in the posttest adopted a more fairminded logical argumentation, generally including opposing ideas rather than neglecting them. Besides, this logical quality reflects better-elaborated arguments in terms of the presentation of reasons in a

coherent, organized way, which may be attributed to the effects of the teaching intervention. Meanwhile, participants' thinking accuracy and breadth of thought reflect a tendency toward persuasive one-sided thinking using informal evidence to back up positions or judgments before complex social issues. Embedding fairminded critical thinking in argumentative writing is more cognitively demanding than writing common argumentation tasks. Therefore, using EFL instead of L1 and performing tasks spontaneously (without previous preparation) pose particular challenges and could have affected performance with the four standards. Nevertheless, the limitations observed in intellectual accuracy and breadth of thought under this fairminded argumentation approach could also be attributed to inherent human flaws in reasoning influenced by the culture or context, requiring explicit training and commitment from both individuals and educators. Hence, endeavors to systematically address these aspects via training or formal education require institutional support and a long-term commitment.

References

AACU. (n.d.). Critical Thinking VALUE Rubric. *Association of American Colleges & Universities*. http://www.aacu.org/value/rubrics/critical-thinking

Abrami, P., Bernard, R., Borokhovski, E., Waddington, D., Wade, A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of Educational Research*, 85(2), 275–314.

Alwehaibi, H. (2012). Novel program to promote critical thinking among higher education students: Empirical study from Saudi Arabia. *Asian Social Science*, 8(11), 193–204. https://doi.org/10.5539/ass.v8n11p193

Bell, A., Mladenovic, R. & Price, M. (2013). Students'perceptions of the usefulness of marking guides, grade descriptors and annotated exemplars. *Assessment & Evaluation in Higher Education*, 38(7), 769–788.

Cargas, S., Williams, S., & Rosenberg, M. (2017). An approach to teaching critical thinking across disciplines using performance tasks with a common rubric. *Thinking Skills and Creativity*, 26, 24–37. https://doi.org/10.1016/j.tsc.2017.05.005

Chan, N., Ho, I. T., & Ku, K. Y. L. (2011). Epistemic beliefs and critical thinking of Chinese students. *Learning and Individual Differences*, 21(1), 67–77. https://doi.org/10.1016/j.lindif.2010.11.001

Chen, L. (2017). Understanding critical thinking in Chinese sociocultural contexts: A case study in a Chinese college. *Thinking Skills and Creativity*, 24,

140–151. https://doi.org/10.1016/j.tsc.2017.02.015

Davies, M., & Barnett, R. (2015). The palgrave handbook of critical thinking in higher education. In M. David & R. Brannet (Eds.), *The Palgrave Handbook of Critical Thinking in Higher Education*. https://doi.org/10.1057/9781137378057

Dewey, J. (1933). *How we think: A restatement of the relations of reflective thinking to the educative process*. Lexington, MA: Heath.

Doherty, J. H., Wenderoth, M. P., Merrill, J. E., Urban-lurain, M., & Haudek, K. C. (2019). Deconstruction of Holistic Rubrics into Analytic Rubrics for Large-Scale Assessments of Students' Reasoning of Complex Science Concepts. *Practical Assessment, Research, and Evaluation*, 24(7). https://doi.org/DOI: https://doi.org/10.7275/9h7f-mp76

Dong, Y. (2015). Critical thinking education with Chinese characteristics. In D. & R. Barnett (Ed.), *The Palgrave handbook of critical thinking in higher education* (pp. 351–368). New York: Springer.

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2010). The evaluation of argument mapping as a learning tool: Comparing the effects of map reading versus text reading on comprehension and recall of arguments. *Thinking Skills & Creativity*, 5(1), 16–22.

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2015). The effects of argument mapping-infused critical thinking instruction on reflective judgement

performance. *Thinking Skills and Creativity*, 16, 11–26. https://doi.org/10.1016/j. tsc.2014.12.002

Elder, L., & Cosgrove, R. (2019). Critical societies: Thoughts from the past. Retrieved February 13, 2020, from *The Foundation for Critical Thinking website*: http://www.criticalthinking.org/pages/critical-societies-thoughts-from-the-past/762

Elder, L., & Paul, R. (2013). *A glossary of critical thinking terms & concepts: The critical analystic vocabulary of the english language*. Tomales, CA: Foundation for Critical Thinking.

Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educational Leadership*, 43(2), 44–48. https://doi.org/10.3102/0034654310376953

Facione, P. A. (1990). Critical Thinking : A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction . *The California Academic Press*, 423(c), 1–19. https://doi.org/10.1016/j.tsc.2009.07.002

Facione, P. A. (2015). Critical Thinking: What It Is and Why It Counts. *In Insight Assessment*. https://www.insightassessment.com/Resources/Importance-of-Critical-Thinking/Critical-Thinking-What-It-Is-and-Why-It-Counts

Facione, P., & Gittens, C. A. (2016). *Think Critically* (3rd ed.). Pearson Education.

Finken, M. M. (1992). Guidelines for Scoring Illinois Critical Thinking *Essay Tests*. http://www.criticalthinking.net/IllCTEssayTestFinken-Ennis12-1993LowR.pdf.

Fisher, A. (1991). Testing Fairmindedness. *Informal Logic*, 13(1). https://doi.org/10.22329/il.v13i1.2549

Fisher, A. (2004). *The Logic of Real Arguments* (2nd ed.). Cambridge: Cambridge University Press.

Fisher, A. (2011). *Critical Thinking and Introduction* (2nd ed.). Cambridge: Cambridge University Press.

Freire, P. (2005). Pedagogy of the oppressed - 30th anniversary edition. *Pedagogy of the oppressed* (1970) (30th ed.). https://doi.org/10.1215/00382876-1472612

Fung, D., & Howe, C. (2014). Group work and the learning of critical thinking in the Hong Kong secondary liberal studies curriculum. *Cambridge Journal of Education*, 44(2), 245–270. https://doi.org/10.1080/0305764X.2014.897685

Gall, M., Gall, J. P., & Borg, W. R. (2015). *Applying educational research: How to read, do, and use research to solve problems and practice* (7th ed.). Hoboken: Pearson Education.

Halpern, D. F. (2014). *Thought and Knowledge: An Introduction to Critical Thinking* (5th ed.). New York: Taylor & Francis.

Hamp-Lyons, L. (1991). *Assessing second language writing in academic contexts*. NJ: Ablex: Norwood.

Hernández, J. (2016, July 30). Study Finds Chinese Students Excel in Critical Thinking. Until College. *The New York Times*.https://nyti.ms/2apJpjj ASIA

Higgins, S. (2014). Critical thinking for 21st-century education: A cyber-tooth curriculum? *Prospects*, 44(4), 559–574. https://doi.org/10.1007/s11125-014-9323-0

Ho, D. Y. F. (1994). Filial piety, authoritarian moralism, and cognitive conservatism in Chinese societies. *Genetic, Social, and General Psychology Monographs*, 120, 347–365.

Insight Assessment. (2019). Can critical thinking be assessed with rubrics? Retrieved February 9, 2020, from *Insight Assessment a division of California Academic Press* website: http://www.insightassessment.com/Resources/Can-critical-thinking-be-assessed-with-rubrics?

Jonsson, A. (2014). Rubrics as a way of providng transparency in assessment. *Assessment & Evaluation in Higher Education*, 39(7), 840–852.

Jonsson, Anders, & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130–144. https://doi.org/10.1016/j.edurev.2007.05.002

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Kelly, A. V. (2004). *The Curriculum: theory and practice* (5th ed.). London: SAGE.

Ku, K. Y. L., Ho, I. T., Hau, K. T., & Lai, E. C. M. (2013). Integrating direct and inquiry-based instruction in the teaching of critical thinking: an intervention study. *Instructional Science*, 42(2), 251–269. https://doi.org/10.1007/s11251-013-9279-0

Ku, K. Y. L., Lai, E. C. M., & Hau, K. T. (2014). Epistemological beliefs and the effect of authority on argument-counterargument integration: An experiment. *Thinking Skills and Creativity,* 13, 67–79. https://doi.org/10.1016/j.tsc.2014.03.004

Kuhn, D. (1991). *The skills of argument*. Cambridge University Press.

Kuhn, D. (2007). Is direct instruction an answer to the right question? *Educational Psychologist*, 42(2), 109-113. https://doi.org/10.1080/00461520701263376.

Landis, M. Swain, K. Friehe, M. Coufal, K. (2007). Evaluating Critical Thinking in Class and Online: Comparison of the Newman Method and the Facione Rubric. *Communication Disorders Quarterly*, 28(3), 135–143.

Lee, Y. L. (2018). Nurturing critical thinking for implementation beyond the classroom : Implications from social psychological theories of behavior

change. *Thinking Skills and Creativity*, 27, 139–146. https://doi.org/10.1016/j. tsc.2018.02.003

Li, M. (2012). History and Reflection of Researches on Thinking Skills of English Major Students in China. *Newsletter for Critical and Creative Thinking Education* 2012, 9, 6–17.

Lipman, M. (2003). *Thinking in education*. Cambridge, UK: Cambridge University Press.

Liu, F., & Stapleton, P. (2014). Counterargumentation and the cultivation of critical thinking in argumentative writing: Investigating washback from a high-stakes test. *System*, 45(1), 117–128. https://doi.org/10.1016/j.system.2014.05.005

Liu, O. L., Frankel, L., & Roohr, K. C. (2014). Assessing Critical Thinking in Higher Education: Current State and Directions for Next-Generation Assessment. *ETS Research Report Series*, 2014(1), 1–23. https://doi.org/10.1002/ets2.12009

Livingston, S. A. (2009). Constructed-Response Test Questions: Why We Use Them; How We Score Them. *ETS - R & D Connections*, 11(11), 1–8. Retrieved from http://144.81.87.152/Media/Research/pdf/RD_Connections11.pdf

Nejmaoui, N. (2018). Improving EFL Learners' Critical Thinking Skills in Argumentative Writing. *English Language Teaching,* 12(1), 98. https://doi. org/10.5539/elt.v12n1p98

Newman, D. (1995). A Content Analysis Method to Measure Critical Thinking in Face-to-Face and Computer Supported Group Learning. *Interpersonal Computing and Technology,* 3(2), 56–77.

Nold, H. (2017). Using Critical Thinking Teaching Methods to Increase Student Success: An Action Research Project. *International Journal of Teaching and Learning in Higher Education*, 29(1), 17–32.

Norris, S. P. (1992). Testing for the Disposition to Think Critically. *Informal Logic*, 14(2). https://doi.org/10.22329/il.v14i2.2538

Paton, M. (2005). Is critical analysis foreign to Chinese students? In E. Manalo & G. Wong-Toi (Eds.), *Communication Skills in University Education: The International Dimension* (pp. 1–11). Auckland, New Zealand: Pearson Education New Zealand Limited.

Paul, R., & Elder, L. (2005). Estándares de Competencia para el Pensamiento Crítico. In F*undación para el pensamiento crítico*.

Paul, R., & Elder, L. (2007). *Consequential Validity: Using A ssessment to Drive Instruction*.

Paul, R., & Elder, L. (2012). *Critical Thinking: Tools for taking Charge of Your Learning and Your Life* (3rd ed.). Boston: Pearson Education.

Paul, R., & Elder, L. (2014). *Critical Thinking: Tools for Taking Charge of Your Professional and Personal Life* (2nd ed.). New Jersey: Pearson Education.

Pearson. (2012). Watson-Glazer - *Critical Thinking Appraisal User-Guide and Technical Manual*. https://doi.org/http://dx.doi.org/10.1787/9789264259119-en

Pei, Z., Zheng, C., Zhang, M., & Liu, F. (2017). Critical Thinking and Argumentative Writing: Inspecting the Association among EFL Learners in China. *English Language Teaching*, 10(10), 31. https://doi.org/10.5539/elt.v10n10p31

Popham, W. J. (2003). *Test Better, Teach Better: The Instructional Role of Assessment*. Retrieved from http://books.google.com/books?id=J_N3wnkm5dwC&pgis=1

Qian, J. (2015). *A study of critical thinking's impact on English majors' argumentative writing.* Unpublished thesis. Central China Normal University. Wuhan, China.

Rainey, H. P. (1926). A Test of Fair-Mindedness. *The School Review,* 34(3), 233–235. Retrieved from http://www.jstor.org/stable/1079493

Rezaei, A. R., & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing*, 15(1), 18–39. https://doi.org/10.1016/j.asw.2010.01.003

Santos, L. F. (2020). Critical thinking perspectives across contexts and curricula: Dominant, neglected, and complementing dimensions. T*hinking Skills and Creativity*, 35. https://doi.org/10.1016/j.tsc.2019.100610

Shaheen, N. (2016). International students' critical thinking–related problem areas: UK university teachers' perspectives. *Journal of Research in International Education*, 15(1), 18–31. https://doi.org/10.1177/1475240916635895

Smith, T. E., Rama, P. S., & Helms, J. R. (2018). Teaching critical thinking in a GE class: A flipped model. *Thinking Skills and Creativity*, 28(February), 73–83. https://doi.org/10.1016/j.tsc.2018.02.010

Stanovich, K. E. (2011). *Rationality and the Reflective Mind*. New York: Oxford University Press.

Tan, C. (2017). Teaching critical thinking: Cultural challenges and strategies in Singapore. *British Educational Research Journal,* 43(5), 988–1002. https://doi.org/10.1002/berj.3295

Tankersley, K. (2007). *Tests that teach: using standardized tests to improve instruction*.http://books.google.com/books?id=kOE2p9CKgH4C&pgis=1

Tian, J., & Low, G. D. (2011). Critical thinking and Chinese university students: A review of the evidence. Language, *Culture and Curriculum,* 24(1), 61–76. https://doi.org/10.1080/07908318.2010.546400

Tiruneh, D. T., Weldeslassie, A. G., Kassa, A., Tefera, Z., De Cock, M., & Elen, J. (2016). Systematic design of a learning environment for domain-specific and domain-general critical thinking skills. *Educational Technology Research and Development*, 64(3), 481–505. https://doi.org/10.1007/s11423-015-9417-2

van Helvoort, J., Brand-Gruwel, S., Huysmans, F., & Sjoer, E. (2017). Reliability and validity test of a Scoring Rubric for Information Literacy. *Journal of Documentation*, 73(2), 305–316. https://doi.org/10.1108/JD-05-2016-0066

Waller, B. N. (2012). *Critical thinking: Consider the verdict* (6th ed.). Upper Saddle River, NJ: Pearson.

Walters, K. (1994). *Re-thinking reason: New perspectives in critical thinking. Albany*, NY: SUNY Press.

Watson, G. B. (1925). *The measurement of fair-mindedness*.

Williams, L., & Lahman, M. (2011). Online discussion, student engagement, and critical thinking. *Journal of Political Science Education*, 7(2), 143–162. https://doi.org/10.1080/15512169.2011.564919

Xiaobei, W. U. (2017). Are chinese students good at critical thinking:a preliminary analysis and evaluation of a recent study abroad and the report on it concerning this issue. *Journal of Chongqing University of Technology*.

Zhang, T. (2016). Why do Chinese postgraduates struggle with critical thinking? Some clues from the higher education curriculum in China. *Journal of Further and Higher Education*, 41(6), 857–871. https://doi.org/10.1080/0309877X.2016.1206857Appendix

Appendix A. Analytic rubric design for assessing fairminded argumentation in timed writing tasks

Levels / Standards	[5] Skilful argumentation & fairmindedness	[4] Average argumentation & fairmindedness	[3] Skilful argumentation & weak fairmindedness	[2] Average argumentation & weak fairmindedness	[1] Ineffective argumentation
Logic • Coherent discourse organization around a clear thesis. • Fairmindedness is shown by planned integration of opposing views.	Opposing views are integrated into the overall argument structure, and all the following criteria are met: a) A clear general argument structure, with separated elements, consisting of: (i) introduction (including a thesis[2]), (ii) at least 2 body paragraphs[3] (including supporting arguments and reasons), (iii) Conclusion. b) Cohesion can be observed between the above elements through cause/ effect relationships. c) Clear paragraph structure, consisting of: (i) an argument supporting the thesis (ii) reasons (iii) evidence, examples or explanations.	The essay structure includes (i) a thesis (ii) supporting arguments (iii) reasons and evidence, and opposing views are present in at least one chain of reasoning. However, one or more of the following issues may be present: a) Links between composition elements are not necessarily complete or clear. b) A supporting argument is incomplete, or is not clearly connected to the thesis. c) Paragraphs include points that do not relate to the thesis, are incomplete, or repeat previously stated points.	Opposing views are not considered in any chain of reasoning. However, all the rules of logic for strong argumentation (a-c) are met: a) A clear general argument structure (introduction, body, conclusion). b) Cohesion observed between the above elements through cause/ effect relationships. c) Clear paragraph structure, consisting of: (i) an argument supporting the thesis (ii) reasons (iii) evidence, examples or explanations.	Opposing views are given no consideration in any chain of reasoning. Additionally, some parts may have the following issues: a) Links between composition elements are not necessarily complete or clear. b) A supporting argument is incomplete, or does not clearly support the thesis. c) Paragraphs include points that do not relate to the thesis, are incomplete, or repeat previously stated points.	Characterised by more than one of the following issues: a) Absence of a clear thesis. b) Lack of essay structure: bullet points or writing that has otherwise not been organised into paragraphs. c) Absence of supporting arguments or reasons to qualify the thesis. d) Elements of the argument are not distinguishable, or have not been organised in a coherent way.
Accuracy • Arguments supported with reasonable and reliable evidence. • Fairmindedness is shown by suspending judgement when evidence is insufficient, avoiding absolutism, and showing how opposing views are supported.	Includes at least one fully elaborated opposing view, and meets the following criteria: a) Includes more than one supporting argument for the thesis, and justifies each of these with equal weight. b) Uses reliable evidence[4] to justify supporting arguments. In some cases, uses non-controversial informal evidence[5]. c) Uses tentative language to avoid certainty and recognize when evidence is refutable or not universally valid. d) Reasons and evidence directly support the thesis without any weaknesses in their logic.	Includes at least one opposing view in any chain of reasoning, and meets the following criteria: a) Includes more than one supporting argument for the thesis, but not with equal weight. b) Uses reliable evidence[4] to justify one supporting argument c) Poorly considered or inconsistent use of tentative or absolutist language in relation to validity of sources of evidence. d) Some reasons or evidence do not directly support the thesis or support the thesis weakly, or have minor flaws in their logic.	Fair consideration of opposing views is omitted or intentionally downplayed. However, the following criteria for skilful argumentation are met: a) Uses two or more chains of reasoning to support the thesis. b) Uses reliable evidence to justify supporting arguments in favour of one position. In some cases, uses non-controversial informal evidence[5]. c) Some tentative language may be used in relation to evidence. d) Reasons and evidence directly support the thesis without any weaknesses in their logic.	Fair consideration of opposing views is omitted or intentionally downplayed. Additionally, the following issues may be found: a) Uses no more than two[3] chains of reasoning to support the thesis. b) Defends one side using mostly unreliable evidence. c) Reliance on personal or otherwise anecdotal evidence or generalisations. d) Lack of tentative language use in relation to evidence. e) One or more supporting argument may be unclear or incomplete or have minor flaws in their logic.	Characterised by the following logic issues: a) Lack of clear connections between arguments. b) No clearly observable argument elements (thesis, supporting ideas & reasons). And/or exhibits any of the following issues with presented evidence: a) Displays simplistic thinking through use of generalizations. b) Reasons presented to support arguments are vague, short, or incomplete and based on unreliable sources, or are redundant or contradictory to the thesis.

Depth • Claims supported by a clear vertical chain of reasoning;[1] (i.e. claim + reason + evidence / explanation). • Fairmindedness is shown by providing the same depth of reasoning to opposing views, as has been applied to their own reasons.	**Includes at least one fully elaborated opposing view, and meets the following criteria:** a) Includes two or more chains of reasoning, to support the thesis in a comprehensive, multi-faceted way. b) Reasons used to justify all supporting arguments[6] have three levels of depth (from general to specific or from specific to general) and effectively explain why or how reasons are significant to the thesis. c) These chains of reasoning have no weaknesses to their logic.	**Includes at least one opposing view in any chain of reasoning (not necessarily fully elaborated), and meets the following criteria:** a) Includes more than one chain of reasoning to support the thesis. b) Reasons used to justify supporting ideas have three levels of depth (from general to specific or from general to specific) and effectively explain why or how reasons are significant to the thesis. c) These chains of reasoning have no weaknesses to their logic.	**Fair consideration of opposing views is absent from the whole essay. However, the following criteria for skilful argumentation are met:** a) Presents no more than two chains of reasoning to support the thesis. b) Reasons used to justify supporting ideas reach two levels of depth in at least one chain of reasoning (In the other chains of reasoning, the significance of reasons may not be sufficiently explained, or only have one level of depth. c) Minor flaws in logic can be present in one chain of reasoning.	**Fair consideration of opposing views is absent from the whole essay. Additionally, the following criteria are met:** a) Presents no more than two chains of reasoning to support the thesis. b) Reasons used to justify supporting ideas reach two levels of depth in at least one chain of reasoning (In the other chains of reasoning, the significance of reasons may not be effectively explained, or only have one level of depth. c) Minor flaws in logic can be present in one chain of reasoning.	**Characterised by the following general issues:** a) No clear connection between ideas. b) No observable depth to arguments. **Specifically:** c) Presents no more than one complete reason to support the thesis, with no more than one level of depth. d) Ideas or statements that attempt to support the thesis may appear contradictory or redundant, or simply rephrase it.
Breadth • Supporting an argument from several angles, showing a range of reasons. • Fairmindedness is shown by giving equal consideration to both sides of an argument.	**Meets both of the following criteria:** a) Includes at least one fully elaborated opposing view, fully elaborated and treated with the same weight, clarity and logic as the author's own supporting arguments, although one may be more strongly supported than the other(s). b) Presents a discussion that includes three or more supporting arguments[6] and elaborates each of these with the same weight, clarity and logic.	**Includes at least one opposing view in any chain of reasoning (not necessarily fully elaborated), and meets one of the following criteria:** a) Presents a discussion that includes at least two supporting arguments, although one may be more strongly supported than the other(s). b) Presents a discussion that includes two supporting arguments, but reasons are unclear, disconnected, incomplete or have weaknesses to their logic.	**Excludes any fair presence of any opposing view, but meets both of the following criteria:** a) Presents a discussion that includes three or more elaborated supporting arguments. b) Justifies each chain of reasoning with the same weight, clarity, and logic. **As well as either of the following criteria:** c) Presents one side of the issue with extensive justification to support it, while the merits of opposing views are not recognized. d) Presents two sides of the issue but clearly downplays one of the sides, emphasizing its demerits without recognizing its merits.	**Excludes any fair presence of an opposing view, and displays the following characteristics:** a) Presents a discussion that includes no more than one supporting argument. b) If two supporting arguments are presented, they do not have equal weight, clarity, or logic. In other words, one supporting argument is too short, limited, or may exhibit problems of clarity or logic.	**Makes no mention of any opposing viewpoint, and is characterised by the following issues:** a) Presents no more than one supporting argument. b) Lack of valid reasons. Displays simplistic thinking through generalizations, or otherwise unjustified claims. c) If an opposing view is present, then it has serious problems with logic (contradictory or redundant).

Notes:

[1] Opposing views in this rubric is equivalent to rival sides, counterarguments or pros & cons.

[2] Thesis equates to position, conclusion, or main idea.

[3] The number of paragraphs can vary when using the rubric tool in other circumstances (e.g., longer writing tasks, etc.)

[4] Reliable evidence in this rubric refers to citation of research findings from reputed academic literature.

[5] Informal evidence refers to common sense, circumstantial evidence, personal experience. It may also refer to self-evident factual or historical references (it is subjected to discussion during instruction).

[6] Supporting arguments and "chain of reasoning" in this rubric refer to non-opposing sides or angles that add to the argument strength

Appendix B. *Indicators of positive and negative CT traits for content analysis*

ACCURACY	CODES
Positive	
Thesis/conclusion stated in a self-evident tone or evidence with tentative language	[A1+]
Idea or statement justified with formal evidence, otherwise with self-evident reasons	[A2+]
Controversial ideas treated with tentative language.	[A3+]
Negative	
Thesis/conclusion No justified	[A1-]
Statement/claim without valid justification.	[A2-]
Controversial idea justified with certainty and unreliable evidence/weak reasons.	[A3-]
DEPTH	
Positive	
Two-level depth cause/reason	[D1+]
Three-level depth cause/reason	[D2+]
Negative	
Thesis statement or topic sentence with only one supporting reason (non-root cause/reason).	[D1-]
Thesis statement or topic sentence with no supporting reason (non-root cause/reason).	[D2-]
BREADTH	
Positive	
Opposing/rival idea, counterargument, or pro & con idea included and fully elaborated.	[B1+]
Opposing/rival idea or pro & con discussion included not fully elaborated.	[B2+]
Second/Third chain of reasoning fully elaborated	[B1+]
Negative	
Absence of opposing/rival ideas, counterargument, or pro & con ideas.	[B1-]
Opposing/rival idea, counterargument included as a "straw-man argument"	[B2-]
Composition/essay with only one chain of reasoning elaborated or a second one not fully elaborated.	[B3-]
LOGIC	
Positive	
Thesis/conclusion responds directly to the question prompt	[L1+]
Coherent paragraph (in relation to the thesis statement)	[L2+]
Paragraph meets cohesion (cohesion within ideas in the paragraph)/minor issues can be admitted (minor loosen/disconnected/incomplete ideas)	[LC3+]
Negative	
Thesis/conclusion does not respond to the question	[L1-]
Paragraph/topic sentence disconnected from the thesis/proposition (coherence issue).	[L2-]
Paragraph cohesion with major issues	[L3-]

Appendix C. *Numerical and graphical normality analysis*

Table C1

Tests of Normality

| | Shapiro-Wilk | | |
	Statistic	df	Sig.
Logic_Test1	.855	12	.042
Accuracy_Test1	.751	12	.003
Depth_Test1	.827	12	.019
Breadth_Test1	.832	12	.022
Logic_Test2	.934	12	.422
Accuracy_Test2	.915	12	.248
Depth_Test2	.777	12	.005
Breadth_Test2	.944	12	.547

Graphic C1.
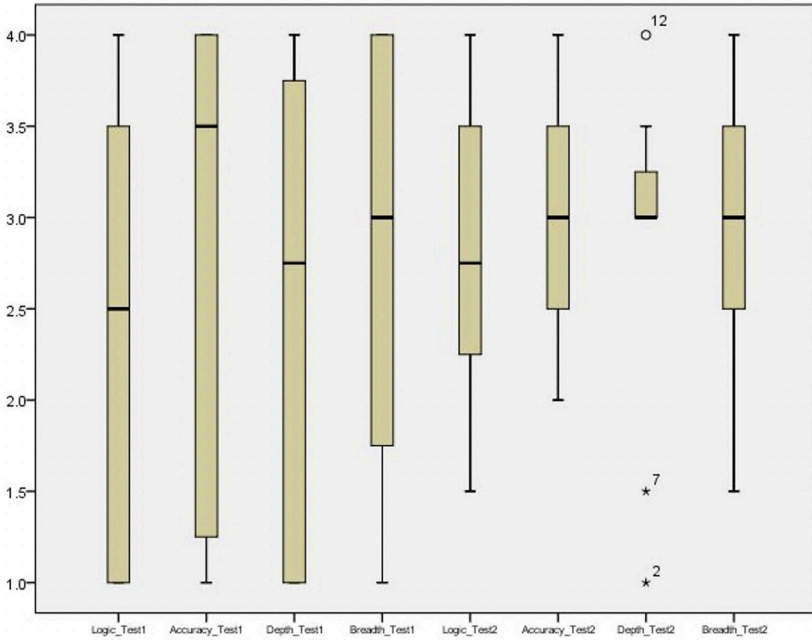
Appendix D. Meaningful themes defined in a participative process through Freire's posing methodology

1. How can the misunderstandings around the HK-China mainland conflict be best solved?
2. Is "misunderstanding" the most important cause of chaos and conflict between HK-Mainland China? If not, which one is the most important one?
3. How can climate change challenge be best addressed?
4. What is the best way to solve the problems of people in situation of emotional and economic vulnerability? 3
5. How can selfishness, unfairness, lack of empathy, and materialistic attitudes in people/society be best addressed?
6. Is selfishness, unfairness, lack of empathy and materialistic attitudes the most serious problems in society?
7. What is the best thing we can do to improve wellbeing of humanity?
8. What is the best way to improve morality in our society?
9. Is lack of morality the most serious problem that affects our society? If not which is the most serious one? 5
10. What is the best way to improve Democracy in our society?
11. Is "lack of democracy" the most serious problem that affects our society? If not which is the most serious one?
12. What is the most important feature that good politics should posses for the benefit of society?
13. How can violence in cases such as HK best be addressed/solved?
14. Which environmental pollutant is the most serious? The one which needs to be solved first?
15. How can we eliminate social unfairness/violence?
16. How can resources be better distributed?
17. Is resources distribution the most serious problem in society that should addressed to solve economic problems of people?
18. What would you do if meeting in a conflict between law and morality?
19. What is the most important aspect that can contribute to cooperative relationships?
20. How does "greed" affect us and society?
21. Is "greed" the most serious problem in society that should be addressed? 2
22. How does "ego" affect us and society?
23. Is Ego the most serious problem in society that should be addressed? If not which is the most serious one?
24. How can we improve "understanding" in society? 1
25. Should the education system promote "understanding" as the first educational priority among other things such as math, English, etc.?
26. What is the most remarkable act of love a person could make? 4

27. What is the most serious problem of our world?

28. Is morality the most serious problem in our society? Why? If not, which one is the most serious?

29. What kind of education is required to improve people's moral?

30. What is the best way to educate people in values?

31. How can poor people's living environment be best solved?

32. What public policies are not working well?

33. What are the most effective public policies that can be adopted by the government to improve the most serious problems of our society?

34. What does being well-educated mean?

35. Those people that harm society do that because they are not well-educated? Is it right? Is this the most important reason?

36. Improving the quality of education for everybody is the only way for making our society better?

37. How does greed affect society?

38. How can war be stopped?

39. What is the best way to reduce environmental pollution?

40. Why many people are in poverty?

41. What is the worst consequence of poverty?

42. How to reduce/eliminate poverty?

43. What's the problem with international order (sociopolitical system in the world)?

44. What is the ideal international order (sociopolitical system) that people want for a fair society?

45. Are "little fresh meat" (male idols who are beautiful and delicate) too feminine to have positive influence on Chinese society?

46. What is the best way to improve social justice and equality?

47. Is the lack of recognition of social injustice and inequality in citizens the most challenging problem for social justice and equality? If not, which one is the most challenging one?

48. Does politics limit the freedom of human beings?

49. What is the best way for us to address the climate change problem?

50. Is it a very good thing for human beings to be free of illness even to live in the world forever?

51. How can war be stopped?

52. What is better, equality for everybody or freedom to achieve outstanding privileges based on individual merits/effort?

53. Creation/new inventions promote ambition and a materialistic life and is not good for the whole society. Is that true?

54. What is better, cooperation or competition?

55. Can one person survive in society if he/she only has a caring heart for others but is not clever enough?

56. What is better, a caring heart or a brilliant mind?