How L2 Learners' Critical Thinking Ability Influences Their L2 Performance: A Statistical Approach

Shin'ichiro ISHIKAWA

Institute of Promotion of Higher Education, Graduate School of Cross-Cultural Studies
Kobe University
Kobe, Japan
iskwshin@gmail.com

Abstract— **This paper analyzed scores of a critical thinking (CT) skills test and standard L2 proficiency test of seventy-four Japanese college students. The quantitative analyses revealed that (1) learners' performance on five aspects of CT ability greatly varied, (2) each of the five aspects was largely independent of the others, (3) the integrative CT skill contributed to L2 listening and overall proficiency to some extent, (4) the influence of learners' L2 proficiency levels was limited, and (5) CT ability could be related to male and science students relatively more strongly than to female and business students.**

Keywords—critical thinking, L2 English proficiency, statistical modeling

I. INTRODUCTION

Critical thinking (CT) has been widely regarded as one of the skills that should be appropriately developed in school education. For example, the National Institute for Educational Policy Research of Japan (2013) has recently introduced the concept of "the 21st century skill," which includes basic language/math/information skills, thinking abilities, and practice abilities. Concerning thinking abilities, the Institute explicitly mentions CT as one of its constituents (pp. 26–27) [1].

Reflecting such a recent trend, many TESOL practitioners have considered the possibility of incorporating CT skills training into L2 teaching. However, the correlation between CT ability and L2 English proficiency has not been fully elucidated. Therefore, the current study aims to clarify the relationship between five aspects of CT ability (inference, recognition of assumptions, deduction, interpretation, evaluation of arguments) and L2 skills of Japanese L2 English learners.

CT is generally defined as any type of "rational / logical / intellectual thought" executed through such cognitive processes as listening, reading, discussing, and stating one's opinions (Kusumi, 1996) [2]. It is also summarized as "the ability to consider a range of information derived from many different sources, to process this information in a creative and logical manner, challenging it, analysing it and arriving at considered conclusions which can be defended and justified" (Moon, 2008, p. 21) [3].

Tanaka and Kusumi (2007) list eight types of CT-related cognitive activities: (1) to clarify the problem to be discussed, (2) to distinguish between facts and values, (3) to distinguish between claims based on evidence and those without evidence, (4) to be careful about overgeneralization from a few samples, (5) to be careful about a subjective bias included in a claim, (6) to avoid emotion-based inference, (7) to think about the other possible options, and (8) to think logically [4].

These subskills can be integrated into several higher-order concepts. Binkley et al. (2012) regards CT as a composite of knowledge, skills, and attitudes / values / ethics [5], while Ennis (1987) suggests that CT comprises abilities and tendencies [6], which are paraphrased as skills and attitudes by Kusumi (2005) [7]. It is notable that CT in a broader sense has aspects of both ability (skill/practice) and attitude (tendency/knowledge).

Concerning the relationship between CT attitude and L2 performance, Ishikawa (2016) conducted a questionnaire survey of 169 Japanese college students at the A2 to B2 lower level and compared students' CT attitude scores with their TOEIC (Test of English for International Communication) scores [8]. In this study, the author gave students eighteen question items, which were classified into four types of CT attitude (orientations toward (i) logical thinking, (ii) intellectual exploration, (iii) objectivity, and (iv) evidence). The data analysis showed that (1) the correlation r between overall CT attitude scores and the TOEIC (sub)scores were 0.16 (listening), 0.19 (reading), and 0.24 (total), (2) the correlations among four CT types were 0.23 to 0.51, (3) the correlations tended to be stronger for male students, science students, and students at a higher proficiency band than for female students, humanities students, and students at a lower proficiency band, and (4) orientation toward intellectual exploration, which are assessed by the statements such as "I think it is important to understand how foreigners fell and think" and "I think it is important to continue to study throughout one's lifetime," were related to L2 proficiency more strongly than other types of CT attitude.

Concerning the relationship between CT ability (practical use of CT skills) and L2 performance, Ghanizadeh, Pour, and Hosseini (2017) compared the scores of 107 Iranian students on a CT skills test (inference-making and evaluation of arguments

only) and the ILETS Reading, and reported that (1) the test scores were significantly correlated (r = 0.341 between inference-making and IELTS, and 0.304 between evaluation of arguments and IELTS), (2) inference was a more powerful predictor, and (3) the regression model explained approximately 10% of the IELTS scores [9].

Junining (2013) analyzed the relationship between intelligence (IQ), CT skills in reading, and L2 proficiency of 35 pre-advanced Indonesian learners of English and found that (1) IQ was correlated neither with CT nor with L2 proficiency, while (2) CT skills were correlated with L2 proficiency. Thus, the author claims that CT skills should be taught in (L2) classes [10].

Although these studies have hinted at the possible contribution of CT ability to L2 proficiency, whether this is applicable to Japanese learners of English has not yet been clarified. Therefore, the current study focuses on the scores of Japanese college students on a CT skills test and an English proficiency test.

II. LITERATURE

CT attitude is assessed with questionnaires (subjects are usually required to judge to which extent given statements apply to themselves on 1–5 or 1–7-point Likert scales), while CT ability is usually measured by some type of CT skills test. The table below summarizes the major CT skills tests developed to date (Liu, Frankel, & Roohr, 2014 [11]).

TABLE I.   ELEMENTS MEASURED IN MAJOR CT SKILLS TESTS

CT Skills Tests	Measured Elements
California CT Disposition Inventory	(a) truth-seeking, (b) open-mindedness, (c) analyticity, (d) systematicity, (e) confidence in reasoning, (f) inquisitiveness, (g) maturity of judgment (Facione, Facione, & Sanchez, 1994 [12])
California CT Skill Test	(a) analysis, (b) evaluation, (c) inference, (d) deduction, (e) induction, (f) overall reasoning skills (Facione, 1990 [13])
Cornell Critical Thinking Test (Level Z for Grades 11–12+)	(a) induction, (b) deduction, (c) credibility, (d) identification of assumptions, (e) semantics, (f) definition, (g) prediction in planning experiments (The Critical Thinking Co., 2014 [14])
ETS Proficiency Profile: CT	(a) distinguish between rhetoric and argumentation in a piece of nonfiction prose, (b) recognize assumptions and the best hypothesis to account for information presented, (c) infer and interpret a relationship between variables, (d) draw valid conclusions based on information presented (ETS, 2010 [15])
Watson–Glaser CT Appraisal: Standard Form	(a) inference, (b) recognition of assumptions, (c) deduction, (d) interpretation, (e) evaluation of arguments (Watson & Glaser, 2008 [16], Watson & Glaser, 2008 [17])
Watson–Glaser CT Appraisal: Short Form/ WG2	(a) recognize assumptions, (b) evaluate arguments, (c) draw conclusions (Watson & Glaser, 2010 [18]).

As suggested in the table above, there exist varied viewpoints of what CT entails. Thus, different tests pay attention to different elements, and in some cases the same tests have come to measure different factors through revisions. For example, Watson-Glaser CT Appraisal (WG), one of the most widely administered CT skills tests, traditionally measured five elements, but recently it has been suggested that three elements (inference, deduction, and evaluation of arguments) might be integrated into a new single element of "drawing conclusions." The developers of WG have recently proposed the RED (i.e., recognition of assumptions, evaluation of arguments, and drawing conclusions) model (Watson & Glaser, 2012, p. 6) [19].

Thus the question arises, to which extent does CT ability contribute to learners' L2 proficiency? The Pearson's correlation r values between CT ability and L2 verbal skills have been reported to be 0.43–0.48 for 60 near-native Malay college students (Moore, 1995 [20]), 0.81 for 83 advanced Iranian college students (Fahim et al., 2010 [21]), and 0.238–0.288 for 275 Malay college students (Rashid & Hashim, 2008 [22]).

Reported strengths of correlations greatly differ between studies and whether CT ability really influences L2 performance still remains uncertain. CT has traditionally been discussed in relation to the students' grades in school subjects, and therefore research on the relation between CT ability and L2 performance is still "in its infancy stage" (Rashid & Hashim, 2008).

III. RESEARCH DESIGN AND METHOD

A. Aim and RQs

Focusing on Japanese learners of English, the current study aims to clarify how learners' CT ability is constructed and how it influences their L2 performance. Five research questions (RQs) are to be discussed: RQ1: To which extent can Japanese college students manage five aspects of CT? Are there some aspects at which they are particularly good or poor? (Learners' Performance in Five Aspects of CT Ability); RQ2: How are different aspects of CT ability interrelated? And how can they be integrated? (Relationship and Statistical Integration of Five Aspects of CT Ability); RQ3: To what extent do different aspects of CT ability and the integrated CT ability contribute to learners' L2 performance? (Contribution of CT Ability to L2 Performance); RQ4: Does the degree of contribution of the integrative CT ability to L2 performance change according to students' proficiency bands? (Influence of L2 Proficiency Level on Contribution of CT Ability to L2 Performance); RQ5: How are students' personal attributes, such as sex and academic major, related to their CT ability and L2 proficiency? (Influence of Personal Attributes)

B. Subjects

Seventy-four freshmen at a research-oriented national university in Japan participated in the experiment. Half of them major in business and economics (B/E_M), and the remaining half in architecture (Arc_M). Their mean score on the TOEIC test is 527.5. They are classified into four proficiency bands: A2_lower (< 500 on TOEIC; N = 27), A2_upper (500+; N = 19), B1_lower (550+; N = 20), and B1_upper (670+; N = 8). Subjects have reportedly received no organized instruction in CT sills.

C. Instruments and Procedure

Subjects took the TOEIC in April 2017, and a CT ability test in June 2017. TOEIC is an English proficiency test developed by Educational Testing Service (ETS). It includes 100 listening questions and 100 reading questions. The test time is two hours total (45 minutes for listening sections and 75 minutes for reading sections.) Scores are statistically adjusted and given between 10 and 990. (5 and 495 for each section).

The CT ability test administered in the study is an excerpt from the Watson–Glaser Critical Thinking Appraisal (WG) UK version (Watson & Glaser, 2002 [23]). Instructions and questions were translated into Japanese, and when needed wordings in the original were also changed to make them more natural in a Japanese context (e.g., "North England" and "South England" in the original were replaced by "Prefecture A" and "Prefecture B."). The number of question items were limited to twenty total: seven items in inference, four items in recognition of assumptions, and three items each in deduction, interpretation, and evaluation of arguments. In the inference section, for example, subjects were given a short passage, and after reading it were required to judge the validity of given statements and choose from "true," "partly true," "partly false," "false," and "have insufficient data." Subjects were given twenty minutes to answer twenty questions.

D. Methods of Data Analysis

For RQ1, we compare the correct response ratios in five aspects of CT ability. For example, if a learner answers three questions correctly in the inference section, the correct response ratio is calculated as 42.9%. We also discuss coefficients of variance (CV), which are obtained by dividing standard deviations by means, to discuss the degree of variance for each of the five aspects of CT ability.

For RQ2, we first pay attention to correlations between scores of different aspects of CT ability. Next, we conduct a cluster analysis to see how the five aspects can be clustered. The initial distance is defined as the square root of $(2-2r)$ and the distance after aggravation is calculated with the Ward Method. We then conduct principal component analysis (PCA) to summarize five kinds of scores in a single score. We adjust the obtained primary PC (PC1) values so that they are all positive using the formula PC1*10+50, which we call an integrative CT ability score.

For RQ3, we examine correlations between CT ability scores and TOEIC scores (listening, reading, and total), then we conduct multiple regression modeling to see the relationship between CT ability scores as independent variables and L2 proficiency scores as dependent variables. Also, we conduct cluster analysis to see which aspects are clustered together with proficiency-related variables.

For RQ4, based on the learners' TOEIC scores, we classify learner data into four groups: A2_lower (<500; N = 27), A2_upper (500+; N = 19), B1_lower (550+; N = 20), and B1_upper (670+; N = 8). We then conduct linear regression modeling on the learners' integrative CT ability scores as independent variables and their total TOEIC scores as dependent variables.

Finally, for RQ5, we first compare CT scores between male students and female students, and between business/economics students and architecture students using Welch's t-test. Also, effect sizes (defined as Cohen's d) are discussed. We then conduct correspondence analysis to visualize the relationships among variables. Correspondence analysis identifies a few (usually two) axes classifying the item category data most effectively and visualizes the relationships among item categories on a two-dimensional scatter plot.

IV. RESULTS AND DISCUSSIONS

A. RQ1 Learners' Performance on Five Aspects of CT Ability

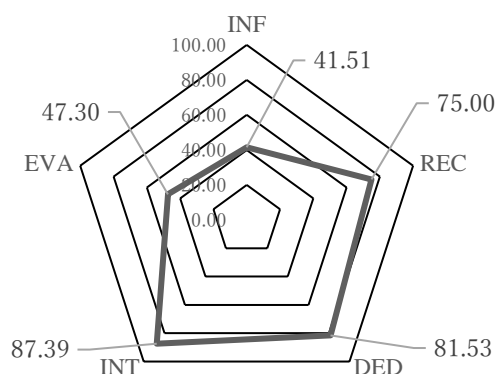Ratios of correct responses in five sections are shown in the figure below:

Fig. 1.   Correct response ratios (%) of five sspects of CT ability

The figure shows that learners' performances on five aspects of CT greatly vary. It seems that they are relatively better at interpretation (87.39%), deduction (81.53%), and recognition of assumptions (75.00%), while their performances on inference (41.51%) and evaluation of arguments (47.30%) are lower than 50%.

Coefficient of variance (CV) is the lowest for recognition of assumptions (0.17), which is followed by interpretation (0.21), deduction (0.24), and inference (0.39), and is the highest for evaluation of arguments (0.64). It seems that learners' cognitive performances tend to be more limited and at the same time more unstable for evaluation of arguments as well as for inference.

B. RQ2 Relationship and Statistical Integration of Five Aspects of CT Ability

Pearson's correlation r values are shown in the table below:

TABLE II.   CORRELATIONS BETWEEN ASPECTS OF CT ABILITY

	INF	REC	DED	INT	EVA
INF	1.00				
REC	0.19	1.00			
DED	0.17	0.00	1.00		
INT	0.05	0.00	-0.02	1.00	
EVA	0.13	0.09	-0.05	-0.04	1.00

$^{**}p < .01$; $^{*}p < .05$; $^{+}p < .10$

The correlations between the five aspects were not statistically significant, which suggests that each aspect is rather independent, at least for Japanese learners of English. Regarding three elements (inference, deduction, and evaluation of arguments) as a new single aspect of CT does not necessarily seem appropriate for the current data.

Independence of individual aspects of CT ability is also suggested in the tree diagram obtained from the cluster analysis.
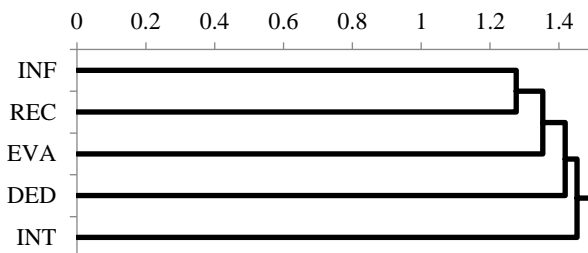

Fig. 2. Tree diagram of the five aspects of CT ability

Aggregation of variables hardly occurs at an earlier stage, suggesting that there are no significant groupings or clusters of any of the five variables.

Next, we conducted PCA to summarize five aspects of CT ability scores into a new single value. Although we obtained three principal components (PC) whose eigenvalues were higher than 1.0, only the primary principal component (PC1) had positive coefficients for all five variables, which we regard as an integrative CT ability score.
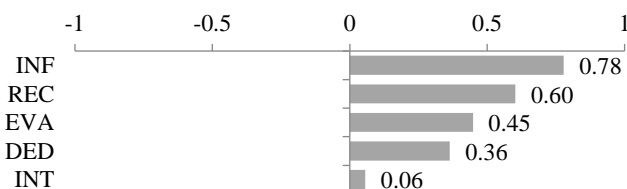

Fig. 3. The contributions of the five aspects of CT ability (PC1)

As shown in the figure above, the largest load is on inference, followed by recognition of assumptions, evaluation of arguments, and deduction. Intuition is hardly included in PC1. Although the five aspects are all exclusive and independent, intuition is particularly deviant among them.

C. RQ3 Contribution of CT Ability to L2 Performance

The analysis showed correlations between aspects of CT ability and L2 performance. L, R, and LR in the table represent listening score, reading score, and total score respectively.

TABLE III.   CORRELATIONS BETWEEN ASPECTS OF CT ABILITY AND L2 LISTENING AND READING SKILLS

	INF	REC	DED	INT	EVA	CTA	Av
L	**0.22**[+]	0.17	0.12	0.06	0.17	**0.31**[**]	0.18
R	0.03	0.19	-0.04	0.14	**0.20**[+]	0.17	0.11
LR	0.14	**0.21**[+]	0.04	0.12	**0.21**[+]	**0.27**[*]	0.17

[**] $p < .01$; [*] $p < .05$; [+] $p < .10$

First, the correlations were somewhat stronger than those reported in Ishikawa (2016), where the correlations between CT attitude and L2 performance were 0.16 (listening), 0.19 (reading), and 0.24 (total). We can assume that CT ability may be related to L2 performance more directly than CT attitude.

Second, many of the correlations were not significant. Significant tendencies of correlations ($p < .10$) were observed only between (i) inference and listening skill, (ii) recognition of assumptions and total skill, (iii) evaluation of arguments and reading as well as total skill, and (iv) integrative CT ability and listening as well as total skill. It is noteworthy that integrative CT ability influences L2 proficiency to some extent. In addition, we note that inference, namely the ability to "draw some conclusions from certain observed or supposed facts," plays a certain role in listening, which requires listeners not only to perceive the oral input superficially but also to infer the meaning embedded in it, while evaluation of arguments, namely the ability to distinguish between strong claims and weak claims, plays a role in reading, which requires readers to read between the lines and carefully examine the value of the written input.

Third, correlations were shown to be relatively stronger with listening skills than with reading skills. Interestingly, this is opposite to the tendency reported in Ishikawa (2016), where correlations were 0.161 with listening skill and 0.192 with reading skill. Presumably, CT attitude mainly concerns careful examination of the information and therefore influences L2 reading, while CT ability mainly concerns real-time online processing of the information and therefore influences L2 listening.

Is it possible for us to estimate learners' L2 proficiency by their CT ability? The figure below illustrates the relationship between integrative CT ability scores (X axis) and overall L2 proficiency scores (Y axis) of subjects.
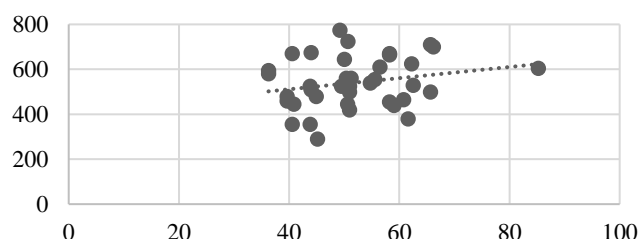

Fig. 4. Scatter plot of the integrative CT ability scores (X) and the TOEIC total (listening and reading) scores (Y)

This regression equation is obtained:

$$y = 2.37x + 408.95 \ (R^2 = 0.07)$$

Although we may observe a somewhat increasing trend in the figure, as shown in the squared-r value, which represents how much of the variation of a dependent variable can be explained by a linear equation model, estimation of the overall L2 proficiency solely by learners' CT ability does not seem possible at a practical level. It should be noted that the squared-r value obtained here is close to that reported in previous studies (e.g., Ghanizadeh, Pour, & Hosseini, 2017).

D. RQ4 Influence of L2 Proficiency Level on Contribution of CT Ability to L2 Performance

Classifying learners into four proficiency bands, we conducted linear regression modeling and obtained formulas: A2_lower: $y = -0.83x + 462.53$ ($R^2 = 0.03$); A2_upper: $y = 0.29x + 509.36$ ($R^2 = 0.03$); B1_lower: $y = 0.01x + 598.62$ ($R^2 < 0.001$); and B1_upper: $y = 0.41x + 677.39$ ($R^2 = 0.01$). The R^2 are all low and a stable tendency was not observed.

E. RQ5 Influence of Personal Attributes

First, we compared the mean scores of the five aspects of CT ability of male students ($N = 51$) and female students ($N = 23$).

TABLE IV.   MEAN SCORES OF FIVE ASPECTS OF CT ABILITY OF MALE AND FEMALE STUDENTS

	INF	REC	DED	INT	EVA	CTA
Male	2.96	3.04	2.49	2.71	1.37	50.90
Female	2.78	2.91	2.35	2.43	1.52	48.11
p	.50	.38	.29	.06+	.49	.34

** $p < .01$; * $p < .05$; + $p < .10$

Except for evaluation of arguments, male students seem to have performed better on all the CT subskills. However, significant differences were not observed, though a significant tendency of difference was suggested in interpretation ($t = 1.91$; $df = 36.97$; Cohen's $d = 0.52$). Unlike the findings in Ishikawa (2016), sex difference was not confirmed in execution of CT skill. Barjesteh and Vaseghi (2012) conducted CT strategy training on 240 Iranian students, who included both sexes and both high and low proficiency learners, and found that (1) use of CT strategies affected learners' reading comprehension scores, (2) the training was effective not only for high proficiency learners but also for low proficiency learners, and (3) sex did not affect use of CT strategies [24]. In addition, Watson and Glaser (2012) also emphasize that there are "only small or negligible differences between men and women in terms of average scores" (p. 40) and conclude that there exists no evidence showing gender bias in the Watson-Glaser CT Appraisal. CT ability seems to be a construct largely independent from other factors such as sex and L2 proficiency.

Next, we compared the mean scores of business/economics students ($N = 35$) and architecture students ($N = 39$).

TABLE V.   MEAN SCORES OF THE FIVE ASPECTS OF CT ABILITY OF MALE AND FEMALE STUDENTS

	INF	REC	DED	INT	EVA	CTA
B/E	2.83	2.89	2.37	2.71	1.37	47.85
Arch	2.97	3.10	2.51	2.54	1.46	51.93
p	.55	.08+	.30	.16	.67	.13

** $p < .01$; * $p < .05$; + $p < .10$

Except on interpretation, architecture students seem to outperform economics students. However, once again significant differences were not observed, though a significant tendency of difference was suggested in recognition of assumptions ($t = 1.78$; $df = 63.53$; Cohen's $d = 0.43$). Unlike the findings in Ishikawa (2016), which suggested science students' outperformance of humanities students, the influence of academic majors was not seen here. These results may suggest an interesting discrepancy between CT attitude and CT ability. On the level of attitude, different types of learners show different orientations, while on the level of ability, namely, actual performance, such a difference often tends to be levelled.

Finally, we have a look at the interplay of personal attributes, aspects of CT ability, and L2 proficiency. The figure below is a scatter plot obtained from correspondence analysis. S1−74 represent individual subjects. The contribution of Z1 (horizontal axis) is 48.4% and the contribution of Z2 (vertical axis) is 27.3%, suggesting that the variables are primarily separated on the horizontal axis into the left and right halves. In the right half, reading proficiency, female students, and business/economics major are positioned together, suggesting a close relationship among them, while in the left half, listening proficiency, male students, architecture students, and aspects of CT ability are positioned together. This suggests that male students and architecture students are relatively more strongly related in execution of CT. Concerning the vertical axis, meanwhile, we see a simple division between L2 proficiency in the upper half and CT ability and learner attributes in the lower half.
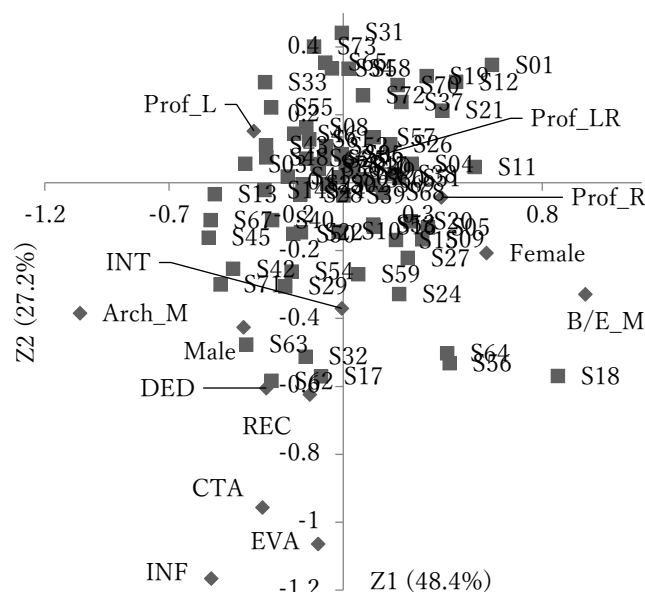

Fig. 5.   Scatter plot of personal attributes, CT ability, and L2 proficiency

V. CONCLUSION

In the current study, we analyzed the results of a CT skills test and standard L2 proficiency test of 74 Japanese college students. First, concerning RQ1 (Learners' Performance in Five Aspects of CT Ability), we found that correct response rates greatly varied (41.51 to 87.39%) and tended to be higher for interpretation, deduction, and recognition of assumptions, while lower for inference and evaluation of arguments.

Concerning RQ2 (Relationship and Statistical Integration of Five Aspects of CT Ability), it was found that correlations were low and each aspect had a unique status. Among the five aspects, inference influences the integrative CT ability most strongly, while interpretation is the most deviant from it.

Then, concerning RQ3 (Contribution of CT Ability to L2 Performance), we showed that the correlations between CT ability and L2 performance were a little stronger than those between CT attitude and L2 performance. Also, we confirmed a significant tendency for the integrative CT ability to influence listening skill as well as overall L2 proficiency ($p < .1$). The fact that CT ability influences listening skill rather than reading skill seems to suggest that CT ability, namely, actual use of CT skill, is closely related to the online information processing required in listening. However, estimating learners' L2 proficiency solely from their CT ability was found to be hardly possible.

Concerning RQ4 (Influence of L2 Proficiency Level on Contribution of CT Ability to L2 Performance), we could not find a clear influence of proficiency, for the equation models were not significant for any of the proficiency groups.

Finally, concerning RQ5 (Influence of Personal Attributes), we did not find significant differences in CT ability scores between male/female students, or between business/ architecture students. However, correspondence analysis showed the possibility of a connection between aspects of CT ability, L2 listening skill, male students, and architecture students.

These findings, especially those of a certain level of contribution of the integrative CT ability to L2 listening and overall proficiency, are noteworthy in that they provide some empirical support for introduction of CT skills training in L2 classes. However, correlations between CT ability and L2 performance are often low and insignificant, and a pedagogical need to deal with CT in L2 classes rather than in other subject classes still remains unclear. Although there is no doubt that educators need to develop students' CT ability as well as their creativity, communication skills, and cooperativeness, we need further research to discuss how and to what extent we should deal with CT in L2 classes.

REFERENCES

[1] National Institute for Educational Policy Research, Japan, *Basic Principle for Curriculum Development to Foster Potentials and Abilities Required in the Changing Society*. Report, 5. Tokyo: National Institute for Educational Policy Research, Japan, 2013. [*Shakai no Henka ni Taio suru Shishitsu ya Noryoku wo Ikusei suru Kyoiku Katei Hensei no Kihon Genri*].

[2] T. Kusumi, "Functional inference and critical thinking," in *Thought: Cognitive Psychology Series Vol. 4*, S. Ichikawa Ed. Tokyo: University of Tokyo Press, 1996, pp. 37-60. ["Kinoteki suiron to hihanteki shiko" in *Shiko: Ninchi Shinrigaku 4*]

[3] J. Moon, *Critical thinking: An exploration of theory and practice*. London, UK: Routledge, 2008.

[4] Y. Tanaka, and T. Kusumi, "Effect of goal and context on judging the use of critical thinking," in *The Japanese Journal of Educational Psychology*, vol. 55, no. 4, pp. 514-525, 2007. ["Hihanteki shiko no shiyo handan ni oyobosu mokuhyo to bummyaku no koka"]

[5] M. Binkley, O, Erstad, J. Herman, S. Raizen, M. Ripley, M. Miller-Ricci, et al., "Defining twenty-first century skills," in *Assessment and Teaching of 21st Century Skills*, P. Griffin, V. McCawa, and E. Care, Eds. pp. 17-66. 2012, New York: Springer.

[6] R. H. Ennis, "A taxonomy of critical thinking dispositions and abilities," in *Teaching Thinking Skills: Theory and Practice*, J. Baron and R. Sternberg, Eds. New York: W. H. Freeman, 1987, pp.9-26.

[7] T. Kusumi, "Measurement of ability and attitude of critical thinking," 2005 Activity Report of Dept. of Educational Measurement and Curriculum Development, Graduation School of Education, University of Tokyo, vol. 1, pp. 103-120. ["Hihanteki shiko no noryoku to taido no sokutei"]

[8] S. Ishikawa, "Attitude of critical thinking and L2 perception: Influence of Learners' sexes, majors and L2 proficiency levels," in *Language Studies and Quantitative Approach*, Y. Ishikawa, S. Ishikawa, Y. Shimizu, T. Tabata, K. Cho, and T. Maeda, Eds. Tokyo: Kinseido, 2016, pp. 130-144. ["Hihanteki shiko taido to L2 rikairyoku-gakushusha no seibetsu: Semmonkei, shujukudo reberu no eikyo," in *Gengo Kenkyu to Keiryoteki Apurochi*]

[9] A. Ghanizadeh, A.V. Pour, and A. Hosseini, "IELTS academic reading achievement: The contribution of inference-making and evaluation of arguments," *European Journal of English Language Teaching*, vol. 2, no. 2, pp. 1-21, 2017.

[10] E. Junining, "The relationship between intelligence, critical thinking skills in reading and EFL proficiency among EFL learners," Post-Graduate Dissertation, Universitas Negeri Malang, 2013.

[11] O. L. Liu, L. Frankel, and K. C. Roohr, "Assessing critical thinking in higher education: Current state and directions for next-generation assessment, ETS Research Report, no. 14-10, pp. 1-23, 2014.

[12] N. C. Facione, P. A. Facione, and C. A. Sanchez, "Critical thinking disposition as a measure of competent clinical judgment: The development of the California Critical Thinking Disposition Inventory," *Journal of Nursing Education*, vol. 33, no. 8, pp. 345–350, 1994.

[13] P. A. Facione, The California Critical Thinking Skills Test-college level. Technical report #2. Factors predictive of CT skills. Millbrae, CA: California Academic Press, 1990.

[14] The Critical Thinking Co., *Cornell Critical Thinking Test level Z*, 2014. Retrieved from http://www.criticalthinking.com/cornell-critical-thinking-test-level-z.html

[15] Educational Testing Service, *ETS Proficiency Profile User's Guide*. Princeton, NJ: Author, 2010.

[16] G. Watson, and E. M. Glaser, *Watson–Glaser Critical Thinking Appraisal, Forms A and B Manual*, Upper Saddle River, NJ: Pearson Education, 2008.

[17] G. Watson, and E. M. Glaser, *Watson–Glaser Critical Thinking Appraisal, Short Form Manual*, Upper Saddle River, NJ: Pearson Education, 2008.

[18] G. Watson, and E. M. Glaser, *Watson–Glaser II Critical Thinking Appraisal: Technical Manual and User's Guide*, San Antonio, TX: NCS Pearson, 2010.

[19] G. Watson, and E. M. Glaser, Watson-Glaser Critical Thinking Appraisal, User-Guide and Technical Manual: UK Supervised and Unsupervised Versions 2012, San Antonio, TX: NCS Pearson, 2012.

[20] R. A. Moore. "The relationship between critical thinking, global English language proficiency, writing, and academic development for 60 Malaysian second language learners," Ph.D. Dissertation, Bloomington: Indiana University, 1995.

[21] M. Fahim, A. Tabataba, M. Bagherkazemi, A. Tabataba, M. Alemi, "The relationship between test takers' critical thinking ability and their performance on the reading section of TOEFL," *Journal of Language Teaching and Research*, vol. 1, no. 6, pp. 830-837, November, 2010.

[22] R. A. Rashid, and R. A. Hashim, "The relationship between critical thinking and language proficiency of Malaysian undergraduates," *Proceedings of the EDU-COM 2008 International Conference. Sustainability in Higher Education: Directions for Change*, Edith Cowan University, Perth Western Australia, 19-21 November 2008.

[23] G. Watson, and E. Glaser, Watson-Glaser Critical Thinking Appraisal: UK Edition, Practice Test, London, UK: Pearson Assessment, 2002.

[24] H. Barjesteh, and R. Vaseghi, "Critical thinking: A reading strategy in developing English reading comprehension performance," *Sheikhbahaee EFL Journal*, vol. 1, no. 2, pp. 21-34, August, 2012.