

Conversation with a Chatbot before an Online EFL Group Discussion and the Effects on Critical Thinking

Yoshiko Goda*, Masanori Yamada**, Hideya Matsukawa***, Kojiro Hata****
and Seisuke Yasunami*

(Received 08 October 2013 and accepted in revised form 07 April 2014)

Abstract This study investigated learners of English as a foreign language (EFL) who engaged in a group discussion and saw how their discussion was affected by a preceding conversation with a chatbot. The chatbot was designed to reflect the Socratic inquiry method based on Eliza, a computer program developed for psychotherapy. Two case studies were conducted, and 130 university students (Case 1: $n=63$; Case 2: $n=67$) were divided into experimental and control groups and observed. Case 1 served as the pilot study and focused on the effects of a chatbot conversation on the discussion; the critical thinking, satisfaction, and number of conversations in experimental and control groups were analyzed. Case 2 examined the difference in critical thinking pre- and post-discussion in both groups. Case 1 showed that a preceding conversation with a chatbot might lead to an increase in the number of contributions that students made to conversations and could increase the number of conversations in which the students participated. Case 2 results showed that pre-discussion with a chatbot could increase the students' awareness of critical thinking and enable them to form inquiring mindsets.

Keywords: EFL, online discussion, critical thinking, chatbot, pre-discussion activity

1. Introduction

Since the introduction of Eliza, the well-known computer program developed for psychotherapy⁽¹⁾, the effects of a conversational agent or chatbot have been examined in numerous educational settings. Previous research studies investigated the effects of the chatbot on different aspects of education such as decision making in games⁽²⁾ and reflection activities⁽³⁾. Gulz⁽⁴⁾ stated that educational agents, including chatbots, can lead to “increased motivation, increased sense of ease and comfort in a learning environment, stimulation of essential learning behaviors, increased smoothness of information and communication processes, fulfillment of need for personal relationships in learning, and gains in terms of memory, understanding, and problem solving” (p. 315). Velentsianos and Miller⁽⁵⁾ also concluded there was strong evidence for the positive effects of chatbots as

educational agents.

The increased need for communication in the global society requires communication skills, self-regulation, and excellent command of language. Universities must provide individuals with practical opportunities for internationally collaborative work and in-depth discussions to express their opinions among persons from all over the world in synchronous and asynchronous manners. When educators use a learning management system (LMS) or other learning system, written online discussion is beneficial for both students and teachers. LMS allows teachers to divide large class sizes into small groups where students' interactions can be recorded. The teachers can monitor these conversations and give feedback. Students in language courses with large class sizes, then, are given practical opportunities for collaborative learning. In light of this, written online discussion of English as a foreign language (EFL) was the focus of this study.

Two case studies with experimental and control groups were conducted to investigate the effects of a pre-discussion activity wherein a student converses with a chatbot before participating in a group discussion. Effects on the students' satisfaction, critical thinking, and number of conversations in EFL during an online discussion were measured. As a pilot study, Case 1 fo-

*Research Center for Higher Education, Kumamoto University, Japan

**Faculty of Arts and Science, Kyushu University, Japan

***Center for Education in Liberal Arts and Sciences, Osaka University, Japan

****Faculty of Social and Management Studies, Otemae University, Japan

cused on the effects of a chatbot on an online discussion. The students' thinking, satisfaction, and number of conversations following the discussion were analyzed. In Case 1, the critical thinking post-inventory was compared in the experimental and control groups. Case 2 employed the same research approach, but it focused on the change in critical thinking by comparing pre-inventory and post-inventory critical thinking responses between the experimental and control groups.

This study serves as fundamental research for the development of an integrated, outside-the-classroom, social-learning support system, which is the final goal of this project. The system will have two main functions: (1) to help students engage in a collaborative learning activity and (2) to deepen students' cognitive engagement in discussions.

The first function was designed to increase social and cognitive presences in the community of inquiry⁽⁶⁾. The second function was designed to organize learners' ideas and help them prepare for complex arguments. In-depth discussions require critical thinking, which increases the quality of interactions and supports the higher cognitive learning necessary for successful computer-supported collaborative learning (CSCL)⁽⁷⁾.

Furedy and Furedy⁽⁸⁾ suggested that Socratic inquiry might benefit critical thinking. Yang, et al.⁽⁹⁾ found that Socratic questioning had positive effects on the development and maintenance of critical thinking in an online discussion setting. Accordingly, the chatbot was developed based on Eliza's adoption of Socratic dialogue methods. Little research has focused on the application of Socratic dialogue in CSCL, especially in an EFL setting. This study, then, looked at how use of a chatbot affected critical thinking, with the goal of providing evidence to validate one of the system functions.

Chatbots have been used as a pre-discussion facilitator in EFL learning. Jia⁽¹⁰⁾ examined the effects of a context-adaptive chatting partner on EFL learning and stated that the application of an educational agent in an EFL setting might promote the communicative approach of EFL. Jia's focus was on an algorithm to provide reasonable responses to users based on their input. However, this study focused on whether conversing with a chatbot before a discussion could provide the students with the time and opportunity to organize their thoughts in English as well as provide pre-writing practice. Through the chatbot, the students were exposed to English questioning expressions, which they could use

during the discussion. It was suggested that interacting with the chatbot in English before a discussion could help the students communicate.

Facione⁽¹¹⁾ defined critical thinking as "purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based" (p. 2). The same definition was adopted in this research. Critical thinking was defined as a technique of thinking that "entails effective communication and problem solving abilities" (p. 2)⁽¹²⁾. This study also suggested that the increased critical thinking supported by pre-discussion activity would lead to effective communication and higher-order learning.

2. Case 1: Research Methods

Case 1 used a post-test-only quasi-experimental research design. Data were collected during the second semester of 2012 at a university in Japan. The purpose of the case study was to investigate the number of conversations, effects of conversing with a chatbot on critical thinking, and satisfaction with online discussions in EFL.

2.1 Participants

The study included 63 students in two computer-assisted language learning (CALL) courses. All were university sophomores. The courses were held in a computer room, and each student was assigned one computer. All the students had taken the Introduction to Information Technology (IT) course as freshmen, so they were assumed to have the minimum IT skills (i.e., typing skills) required for this research. One of the two classes was designated as the experimental group and the other as the control. Before the discussion, the experimental group conversed with the chatbot, while the control group searched online for information related to the discussion topic.

2.2 Chatbot

The chatbot was a revised version of Eliza that was developed for this study. According to Wallace⁽¹³⁾, a chatbot consists of three modules: a user interface, interpreter, and knowledge base. The present program was

Table 1. Two of the Chatbot’s Conditions and Their Eliza Comments & Reply Options.

#	Condition	Eliza’s Optional Comments & Replies
1	Check for first booting	“Hello, how may I help you? Please let me know your opinion.” “Greetings. Please let me know your opinion.” “Good day. Please let me know your opinion.” “What is on your mind today? Please let me know your opinion.” “Please begin when you are ready.” “Please let me know your opinion.” “Hello, what is your opinion?”
2	Check for “I think” or “I feel”	“Why do you think that?” “Why do you think this is true?” “Is this your idea, or did you hear it from someplace else?” “Have you always felt this way?” “Has your opinion been influenced by something or someone?” “Where did you get that idea?” “What caused you to feel that way?”

defined as a chatbot because it consisted of these three modules. The chatbot was different from the original Eliza in two ways: (1) the elimination of psychotherapy-specific conditions and (2) the application of the algorithm of Socratic questioning. There were 11 conditions to select one of the options as chatbot’s comment or reply based on Socratic questioning⁽¹⁴⁾. Table 1 shows two of the conditions and their alternative questions or comments. One alternative was randomly provided to each student. For convenience, the chatbot in this study was called Eliza.

One comment and reply option for the condition was randomly selected and offered to each student based on a check for each condition, the first and second conversational turns, and the chatbot-checked keywords utilized in the student’s opinion. The conditions were each checked; if a condition was met, one of the options was randomly selected. Keywords chosen by the students were not case-sensitive.

2.3 Research Instruments

This research used two instruments: the critical thinking inventory⁽¹⁵⁾ and a researcher-developed ques-

tionnaire to measure the students’ satisfaction with the discussion activities. The critical thinking inventory was used because of its high reliability and because it had been developed for Japanese university students. It consisted of 33 five-point Likert items and four factors: (1) awareness of critical thinking, (2) inquiring mindset, (3) objectivity, and (4) importance of evidence. The values of Cronbach’s alpha for these four factors were 0.85, 0.82, 0.73, and 0.57, respectively.

The discussion satisfaction was rated on a scale of 1 (not satisfactory) to 4 (very satisfactory) in the questionnaire. For the data analysis, the total points for each student on the critical thinking inventory were used as the critical thinking score, and the Likert score was used as the discussion satisfaction score. The questionnaire for the experimental group also included an open-ended question about the pre-discussion with the chatbot. For the number of conversations, the log from the Blackboard bulletin board system (BBS), a LMS, was analyzed.

2.4 Procedures

One class was designated as the experimental group and the other the control group. The pre-discussion activity was conducted for 10 minutes after the students were given brief instructions. In the experimental group, Eliza was introduced as a learning support tool to help the students think deeply and organize ideas during the pre-discussion. It was defined as a type of artificial intelligence programmed to make discussions more meaningful and provide a higher cognitive-learning experience.

The discussion topic was the essential factors and conditions of an ideal family. The students were given the following instructions: “Are you familiar with ‘Sazae-san’, a Japanese cartoon? Do you think their family is ideal? How does your ideal family look? Please discuss the essential factors and conditions of an ideal family in your group.”

The students in the experimental group were instructed to talk with the chatbot to clarify and deepen their thoughts on the topic. They were also encouraged to use the conversation to develop search terms for gathering Internet evidence to support their opinions. The control group students were asked to list their thoughts, and use them to search the Internet for related information; this was consistent with the traditional pre-discus-

ELIZA: What's on your mind today? Please let me know your opinion.
 YOU: About family.
 ELIZA: I see.
 YOU: I think Sazae's family is good.
 ELIZA: What caused you to feel that way?
 YOU: Because the family members believe in themselves.
 ELIZA: What other reasons might there be?
 YOU: Well, I don't know.
 ELIZA: I'd like to know more about your opinion.
 YOU: I don't like my family so much.
 ELIZA: That is quite interesting.
 YOU: Yeah.
 ELIZA: I'd like to know more about your opinion.
 YOU: My family members aren't friendly, so I can't talk with them.
 ELIZA: I see.
 YOU: Do you have family?
 ELIZA: Why do you think that?
 YOU: I don't know who you are.
 ELIZA: Why do you ask?
 YOU: I'm sorry.
 ELIZA: That's okay. Please continue.

Figure 1. Example of a conversation between a student and the chatbot. ELIZA: Chatbot; YOU: Student.

sion method used in EFL writing classes. Then, the students in both classes were divided into discussion groups of four to six members. All the students were required to discuss the topic with BBS on an LMS. The discussion activities lasted 30 minutes. Afterward, the students in both groups answered the critical thinking inventory and the questionnaire.

2.5 Results

The excerpt (Figure 1) shows parts of actual conversations between the chatbot (Eliza) and the students in the experimental group. In the excerpt, they conversed with Eliza as if the program understood what the students said, although the instructor explained to her students that Eliza was a chatbot developed to help them practice organizing their ideas in English prior to the discussion.

Table 2. Descriptive Statistics and *t*-Test Results.

	Group	<i>n</i>	<i>m</i>	<i>SD</i>	Difference	<i>t</i>	<i>p</i>
Conversation Number	Experimental	31	4.97	2.66	2.53	5.01	0.00
	Control	32	2.44	1.01			
Satisfaction	Experimental	31	3.00	0.63	-0.03	-0.19	0.85
	Control	32	3.03	0.69			
Critical Thinking	Experimental	31	120.90	14.54	0.34	0.10	0.92
	Control	32	120.56	11.45			

The critical thinking, satisfaction scores, and the number of conversations were analyzed using a *t*-test for independent groups. Table 2 shows the descriptive statistics for the three dependent variables and the *t*-test results for each variable. No significant differences were found for the critical thinking and satisfaction scores between the experimental and control groups. The *t*-test results for the four critical thinking factors of the experimental and control groups were not significant. The *t*-scores for factors 1 to 4 were 0.78, 0.97, 0.09, and 0.41, respectively.

Table 2 demonstrates that the experimental group had an average of 4.97 conversations during the discussion activity, while the control group had an average of 2.44. The difference between the two groups was significant ($t[61]=5.01, p<.01$).

Table 3 shows the typical and representative opinions about the pre-discussion activity with Eliza. These were selected from the open-ended question about Eliza in the post-questionnaire. The students had positive and negative opinions about Eliza's usefulness. Some of the students perceived Eliza as a good tool for practice but discussed the need for improvement.

3. Case 2: Research Methods

Case 2 used a pre-test and post-test quasi-experimental research design. It incorporated the same approach as Case 1 in the experimental and control groups. The focus of Case 1 was not critical thinking but overall effect on discussion. Case 2 focused on the change in the learners' critical thinking after the discussion as an effect of the chatbot in the pre-discussion.

3.1 Participants

The initial participants in this research were 77 students taking either of two CALL classes offered in the second semester of 2012. One of the classes was desig-

Table 3. Students’ Opinions about the Prediscussion Activity with Eliza.

Opinion Type	Opinion about Eliza
Positive	It’s useful. I could deepen my opinion. It would be better practice if Eliza responses were a little more like a human’s.
Positive	I think that Eliza was well developed.
Positive & Negative	Some of Eliza’s questions and responses were good for the conversation, but some were unrelated to the previous conversation. I would like to ask Eliza questions as well.
Positive & Negative	Eliza did not respond meaningfully, but it was good practice for conversation in English.
Negative	It was artificial and bleak, then I felt lonesome.
Negative	I didn’t feel it was very useful since Eliza’s questions did not meet my expressions.

nated as the experimental group and the other as the control group. All the students were freshmen at the same university from Case 1. Data were analyzed from 67 students (32 in the experimental group and 35 in the control group) who completed all the research tasks.

3.2 Research Instruments

Case 2 used the same critical thinking inventory as Case 1. Since Case 2 focused on the change in critical thinking, the scoring was done based on the critical thinking factors: (1) awareness of critical thinking, (2) inquiring mindset, (3) objectivity, and (4) importance of evidence.

3.3 Procedures

The students in both the experimental and control groups were required to take the critical thinking inventory (pre-inventory) in the first class. In the next class, the research purpose and procedures were explained to the students.

For the next 10 minutes, the students did the pre-discussion activity. The students in the experimental group conversed with the chatbot to deepen their thoughts, to identify what they already knew and needed to know, and to practice expressing their thoughts and

Table 4. Repeated *t*-Test Results for Experimental and Control Groups.

	Factor	<i>m</i>	<i>SD</i>	<i>t</i>	<i>p</i>
Experimental Group (with Chatbot)	1	1.72	3.37	2.88	0.01
	2	2.06	4.72	-2.47	0.02
	3	0.69	2.1	-1.85	0.07
	4	-0.19	1.8	0.59	0.56
	Factor	<i>m</i>	<i>SD</i>	<i>t</i>	<i>p</i>
Control Group (without Chatbot)	1	-3.31	0.93	-21.03	0.00
	2	3.11	5.39	-3.42	0.00
	3	1.60	3.40	-2.79	0.01
	4	-0.11	2.17	0.31	0.76

Note. *df*=34.

ideas in English. The students in the control group were instructed to list their thoughts and search for related information on the Internet.

After the pre-discussion activities, the students were divided into groups of four to six. They discussed online using the BBS of the LMS. The discussion period lasted 30 minutes. The students in both groups were asked to take the critical thinking inventory (post-inventory) after the discussion.

3.4 Data analysis

The differences between the experimental and control groups for the four factors of critical thinking were analyzed using a *t*-test. The overall pre-inventory results showed no significant difference between the total critical thinking scores of the two groups ($t[65]=-0.85, p=0.40$). The *t*-test results showed that there were no significant differences in factors between the pre- and post-inventory groups.

Table 4 shows the repeated *t*-test results for the four factors in the pre- and post-inventories for both the experimental and control groups. The results show the change in critical thinking. With the chatbot, the experimental group had significant results for factors 1 (awareness of critical thinking) and 2 (inquiring mindset).

Appendix A extracts the significant results from the repeated *t*-test in the two groups for factors 1 and 2. In the experimental group, items 2, 4, and 8 of factor 1 and items 19, 20, and 22 of factor 2 had significantly higher scores in the post-inventory.

In the control group, the changes in factors 1, 2, and 3 (objectivity) were significant. All three *t*-scores were negative, which implies that the post-inventory scores were lower than the pre-inventory scores.

4. Discussion and Future Research

The results of Case 1 suggest that conversing with a chatbot before group discussions may affect the number of conversations of EFL students. Preparation using a chatbot could increase conversations during discussions. An increase in conversations is the first step toward quality interactions, which are essential for activating CSCL.

No significant differences were found between the experimental and control groups in the students' critical thinking and satisfaction. However, both groups scored relatively high in critical thinking and satisfaction. For critical thinking, only post-inventory evaluation was conducted and overall scores were compared, which might lead to no significance between the two groups. The research method and data analyses were not sensitive enough to establish differences. Thus, in Case 2, pre- and post-inventories were conducted, and four factors of critical thinking were analyzed.

Case 2 results showed that pre-discussion activities such as conversing with a chatbot or listing and searching online might affect learners' critical thinking differently. Conversing with the chatbot affected the individual critical thinking factors of "awareness of critical thinking" and "inquiring mindset". Although the results might have been affected by employing Socratic inquiry, the effects of Socratic inquiry were not investigated in this study. As explained in the introduction, Socratic inquiry was used in the chatbot programming because previous research has found positive effects of Socratic inquiry on critical thinking. Therefore, the results were interpreted that a chatbot, which allowed the students to respond by Socratic inquiry, might have a positive influence on their critical thinking.

In the control group, the factors of "awareness of critical thinking", "inquiring mindset", and "objectivity" showed significant negative differences. Listing their thoughts and searching for related information on the Internet as a pre-discussion activity may reduce learners' confidence in some aspects of critical thinking. This pre-discussion activity could also give students a negative perspective on critical thinking.

However, the results of Case 2 showed no significant differences between the experimental and control groups in the post-inventory critical thinking results, which was consistent with the results of Case 1. This may be due to the short duration of the pre-discussion activities. Long-term use of a chatbot might lead to significant effects on students' affection, cognition, and behavior in the context of discussion. Future research should consider both the longitudinal effects and the quality of students' conversations. Also, the results might be affected by the chatbot's development. For example, the chatbot in this study was primitive, so a chatbot with sophisticated artificial intelligence and Socratic inquiry effects should be considered for future research.

Acknowledgement

This work was supported in part by KAKENHI, Grant-in-Aid for Scientific Research (B), 23300304.

References

- (1) Welzenbaum, J.: "ELIZA—A Computer Program for the Study of Natural Language Communication between Man and Machine", *Communications of the ACM*, Vol. 9, No. 1, pp. 36–45 (1966).
- (2) Augello, A., Pilato, G. and Gaglio, S.: "A Conversational Agent to Support Decisions in SimCity Like Games", in *Proc. of the 2009 IEEE International Conference on Semantic Computing*, pp. 367–372 (2009).
- (3) Pon-Barry, H., Clark, B., Schultz, K. et al.: "Contextualizing Reflective Dialogue in a Spoken Conversational Tutor", *Educational Technology and Society*, Vol. 8, No. 4, pp. 42–51 (2005).
- (4) Gulz, A.: "Benefits of Virtual Characters in Computer Based Learning Environments: Claims and Evidence", *International J. of Artificial Intelligence in Education*, Vol. 14, No. 3, pp. 313–334 (2004).
- (5) Velentsianos, G. and Miller, C.: "Conversing with Pedagogical Agents: A Phenomenological Exploration of Interacting with Digital Entities", *British J. of Educational Technology*, Vol. 39, No. 6, pp. 969–986 (2008).
- (6) Garrison, D. R.: *E-Learning in the 21st Century: A Framework for Research and Practice*, Routledge, New York (2011).
- (7) Benbunan-Fich, R., Hiltz, S. R. and Harasim, L.: "The Online Interaction Learning Model: An Integrated Theoretical Framework for Learning Networks", in

- Learning Together Online: Research on Asynchronous Learning Networks, eds. Hiltz, S. R. and Goldman, R., pp. 19–37. Lawrence Erlbaum Associates, Mahwah (2005).
- (8) Furedy, C. and Furedy, J.: “Critical Thinking: Toward Research and Dialogue”, in Using Research to Improve Thinking: New Directions for Teaching and Learning, eds. Donald, J. G. and Sullivan, A. M., pp. 51–69. Jossey-Bass, San Francisco (1985).
- (9) Yang, Y-T. C., Newby, T. J. and Bill, R. L.: “Using Socratic Questioning to Promote Critical Thinking Skills through Asynchronous Discussion Forums in Distance Learning Environments”, American J. of Distance Education, Vol. 19, No. 3, pp.163-181 (2005).
- (10) Jia, J.: “CSIEC (Computer Simulator in Educational Communication): A Virtual Context-Adaptive Chatting Partner for Foreign Language Learners”, in Proc. of the IEEE International Conference on Advanced Learning Technologies (ICALT 2004), 0-7695-2181-9/04 (2004).
- (11) Facione, P. A.: Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction. The Complete American Philosophical Association Delphi Research Report. ERIC Doc. No.: ED 315 423. The California Academic Press, Millbrae (1990).
- (12) Paul, R. and Elder, L.: Critical Thinking: Concept and Tools, The Foundation of Critical Thinking, Tomales, CA (1995).
- (13) Wallace, R.: Be Your Own Bot Master, ALLICE A. I. Foundation, Oakland (2004).
- (14) Intel Corporation: “Designing Effective Projects: Questioning the Socratic Questioning Technique”, <http://www.intel.com/content/dam/www/program/education/us/en/documents/project-design/strategies/dep-question-socratic.pdf> (2007).
- (15) Hirayama, R. and Kusumi, T.: “Effect of Critical Thinking Disposition on Interpretation of Controversial Issues: Evaluating Evidences and Drawing Conclusions”, Jpn. J. of Educational Psychology, Vol. 52, No. 2, pp. 186–198 (2004).

Appendix A. Extract Significant Results of Repeated *t*-Test for Factors 1 and 2 in Case 2.

Factor	Item #	Item	Experimental Group				Control Group			
			<i>m</i>	<i>SD</i>	<i>t</i>	<i>p</i>	<i>m</i>	<i>SD</i>	<i>t</i>	<i>p</i>
1 Awareness of critical thinking	1	I am good at thinking complicated things.	-0.16	0.85	-1.04		-0.40	1.14	-2.07	*
	2	I am good at organizing ideas.	-0.44	1.11	-2.24	*	-0.09	1.20	-0.42	
	4	I can persuade anyone.	-0.56	0.98	-3.24	**	-0.60	0.98	-3.64	**
	5	I am confused when I think complicated things.	0.16	0.63	1.41		0.46	1.17	2.31	*
	6	My friends let me judge since I make a dispassionate decision.	-0.13	0.79	-0.89		-0.54	0.98	-3.28	**
	7	I can concentrate when I solve a problem.	-0.16	0.63	-1.41		-0.57	1.20	-2.83	**
	8	I can continue engaging in a difficult problem.	-0.50	1.11	-2.55	*	-0.14	0.85	-1.00	
	9	I think a thing through step-by-step.	-0.25	0.92	-1.54		-0.40	0.85	-2.79	**
	11	I cannot manage to think about other ideas when I think about one thing.	0.25	1.02	1.39		0.60	1.44	2.47	*
2 Inquiring mind	16	I like a new challenge.	-0.34	1.04	-1.88		-0.37	1.00	-2.19	*
	17	I want to learn about various cultures.	-0.16	0.88	-1.00		-0.46	1.09	-2.47	*
	18	I think it valuable to learn what foreigners think.	-0.16	0.77	-1.15		-0.46	1.24	-2.17	*
	19	I am interested in people with different ideas from mine.	-0.31	0.78	-2.27	*	-0.17	0.75	-1.36	
	20	I want to learn more regardless of topics.	-0.53	1.05	-2.87	**	-0.34	1.03	-1.97	
	22	I am interested in discussions with people having different ideas from mine.	-0.31	0.82	-2.15	*	-0.51	0.78	-3.90	**
	23	I want to ask a question when I am uncertain.	-0.16	0.77	-1.15		-0.43	1.01	-2.51	*

Note. Experimental Group: *n*=32, *df*=31; Control Group: *n*=35, *df*=34. *p*: **p*<.05, ***p*<.01.