Enhancing Post-secondary Writers' Writing Skills with a Chatbot: A Mixed-Method Classroom Study

Michael Pin-Chuan Lin[*] and Daniel Chang

Simon Fraser University, Faculty of Education, Canada // pla85@sfu.ca // dth7@sfu.ca
[*]Corresponding author

ABSTRACT: In the present study, we developed a chatbot that helps teachers to deliver writing instructions. By working with the chatbot, the post-secondary writers developed a thesis statement for their argumentative essay outlines, and the chatbot helped the writers to refine their peer review feedback. We conducted a preliminary analysis of the effect of a chatbot on these writers' writing achievement. We also collected several student testimonials about their chatbot experiences. Several important pedagogical and research implications for chatbot-guided writing instructions and the use of learning technology have been addressed.

Keywords: Chatbot, Intelligent tutoring system (ITS), Learning tool, Writing skills, Classroom study

1. Introduction

Since the past decades, many educators have started to diversify instructions by adopting educational technology in classrooms. Several intelligent tutoring systems (ITS) have been developed, and these ITS are often built upon specific algorithms that offer learners individualized instructions or evaluate students learning products (Kerly, Hall, & Bull, 2007; Ma, Adesope, Nesbit, & Liu, 2014; Rodríguez-Gil, García-Zubia, Orduña, Villar-Martinez, & López-De-Ipiña, 2019; Vanlehn, 2006). However, Murray (1999) pointed out some limitations of ITS, including low fidelity from student perspectives, limited instructional values, lack of student modelling, and limited interactivity. To overcome these limitations, some studies recommend that if a chatbot is programmed on the supplementary side of ITS, it may help to facilitate a real-time dialogue that supports thinking and learning processes. Nevertheless, some research studies have not distinguished the differences between ITS and a chatbot (Kerly et al., 2007; Wang & Petrina, 2013). Traditionally, an ITS often takes over an instructor's role by presenting learning materials and offering feedback to students (Song, Oh, & Rice, 2017). On the other hand, a chatbot is often a supplementary conversational program that interacts with users synchronously, such as human-like conversations, question answering, user support, or tutoring (Abbasi & Kazi, 2014; Clarizia, Colace, Lombardi, Pascale, & Santaniello, 2018; Kerly et al., 2007; Pereira & Díaz, 2018). Especially, Jain, Kumar, Kota, and Patel (2018) defined chatbots as "text-based, turn-based, task-fulfilling programs, embedded within existing platforms" (p. 904).

In educational research, a chatbot has always been implemented with a specific instructional intention, such as promoting class engagement (Kerly et al., 2007) or promoting critical thinking (Goda, Yamada, Matsukawa, Hata, & Yasunami, 2014). For instance, teachers might use a chatbot to promote critical thinking. Goda et al. (2014) carried out two case studies involving a total of 130 university students, divided into two groups, with each participating in two successive class periods. One class period was experimental, where students conversed with an ELIZA-based chatbot, and another class period was a comparison experience where students listed their thoughts and searched relevant information on the Internet. The results from case study 2 revealed a positive impact of the chatbot on students' awareness of critical thinking and inquiring mindset.

Another specific instructional intention might be promoting language learners' conversational skills. Studies in applied linguistics have found that a chatbot might be developed to improve language learners' conversational skills (Fryer & Carpenter, 2006). Particularly, Fryer and Carpenter (2006) mentioned: "there is yet no chatbot designed from the bottom up to meet the needs of FLL (foreign language learners) students" (p. 12). They have suggested how a learner can freely engage conversations with a chatbot, review what has been talked about in a transcript format, and self-analyze the transcription from the interaction. For teachers, a chatbot might also be useful because the conversation, presented in a transcript format, records a student's progression as well as their needs for future learning. Furthermore, when learners are struggling with the learning materials, a chatbot can also offer just-in-time guidance and solve basic problems for students immediately (Fryer & Carpenter, 2006; Pereira & Díaz, 2018).

Upon careful examination of educational technology literature, we have realized that chatbot-guided writing activities are relatively rare and scarce. Mainly, previous research has established a chatbot-led pre-discussion activity to improve students' critical thinking skills (Goda et al., 2014). In our study, a writing chatbot was introduced as part of a university disciplinary writing class activity, acting as a supplementary activity to in-class writing instructions. Acquiring writing skills is not a linear process, yet it is an interactive social process that requires multiple, multichannel input and output between individuals and the chatbot itself. Theoretically speaking, building a chatbot that assists writing instructions might fulfill novice writers' needs to initiate a dialogue when help is needed right away (Cazden, 1988; Edwards & Mercer, 2013).

The current study further advances Goda's et al. study (2014) in a way that the chatbot can help students to generate their thesis statement for an argumentative essay outline. Through working with the chatbot, students will be able to evaluate their ideas for their thesis statements. Overall, the expected outcome of the chatbot is to assist students with two major events in the writing process: drafting a thesis statement for the essay outline and learning to offer peer feedback on the outlines. Data collected as a result of learners' conversation with the chatbot point to the following essential questions:

- When students work with a chatbot for their thesis statements and peer feedback, do these chatbot-led activities enhance their writing achievement?
- How do students perceive the use of chatbot in a university classroom?

These two questions are crucial in understanding how a chatbot can supplement writing instructions, as instructors are not always available when students need help (Song et al., 2017; Xu & Wang, 2006). Based on the social constructivist theory of learning, a conversation is the key to learning (Kalina & Powell, 2009). If so, engaging students in a chatbot-led conversation might support the writing and learning activities. We thus expect that conversations initiated by the chatbot might support student writers in composing their thesis statements and peer feedback.

2. Literature review

2.1. Design framework of Chatbot

The antecedent of a chatbot is ELIZA, a computerized system capable of parsing human's natural language and initiating conversations. Eliza parsed user input, identified key phrases from its backend template and selected corresponding responses (Weizenbaum, 1996). Failure occurs when user input does not exist in its backend text template (Fryer & Carpenter, 2006; Kirakowski, O'Donnell & Yiu, 2009). However, Eliza's programmed scripts and unnatural interaction are the primary design limitations (Goda et al., 2014). Subsequent chatbot research has suggested some important design features, such as field-specific scope control, use of multimedia resources, and fallback response (Jain et al., 2018).

2.1.1. Field specific scope control

A domain-specific chatbot is the most favourable type of chatbot because it minimizes the unrestricted creativity of human language capacity (e.g., Jain et al., 2018; Luger & Sellen, 2016). Especially, Luger and Sellen (2016) have found that a chatbot is better developed within limited scopes, functions and purposes so that users can better facilitate effective interaction with the chatbot and have better ability to fulfill their purposes. For example, Ghose and Barua (2013) developed an FAQ chatbot using AIML (Artificial Intelligence Markup Language). The purpose of the chatbot was to serve as an undergraduate students' advisor, and it was designed for assisting course and admission information retrieval. The research also evaluated the conversation accuracy rate by investigating the student-to-chatbot interaction logs. Results showed the students were more satisfied with the domain-specific chatbot. Their results imply a domain-specific chatbot is more helpful than a chatbot that understands everything. Moreover, a chatbot (Lucy) developed by Wang and Petrina (2013) failed to recognize student input accurately when Lucy was programmed for multiple areas. To solve this issue, Lucy was redesigned and programmed with more specific domains that only handle topics related to tourism. Taken together, designing a domain-specific chatbot seems to increase the accuracy of chatbot's response.

2.1.2. Use of multimedia resources

Combining text and multimedia facilitates the interaction between a chatbot and a user (Jain et al., 2018). A combination of multimedia resources is effective in promoting engagement, especially if the chatbot has a text-to-speech function in the field of education (Fryer & Carpenter, 2006). Furthermore, if a chatbot can begin the conversation by playing a game with a user, then users might be more engaged and motivated to explore the chatbot (Luger & Sellen, 2016). These findings imply that chatbot-guided instructions might consider integrating several multimedia features for students to engage their attention.

2.1.3. Fallback response

Fallback response is a mechanism triggered when a user intent is out of space that may cause a conversation to fail. Kerly et al. (2007) mentioned that a chatbot should deliver an effective conversation in preventing failure. Especially in natural human language, expressions diverse for various purposes. Preventing conversational errors by directing learners to the correct conversation path is critical. Therefore, fallback response should be explicitly addressed to re-engage users into the conversation (Jain et al., 2018).

2.2. The Impact of Chatbot-guided instructions on achievement

Positive effects of chatbot have been reported in research, such as longer memory retention, enhanced critical thinking skills, and improved language use and engagement (Abbasi & Kazi, 2014; Goda et al., 2014; Kerly et al., 2007; Heller, Proctor, Mah, Jewell, & Cheung, 2005; Huang, Hew, & Gonda, 2019; Wang & Petrina, 2013). Abbasi and Kazi (2014) measured the students' learning outcomes and memory retention by comparing the use of a chatbot and the Google search engine. The results showed that the students remembered the responses from the chatbot more than the Google search engine, and they outperformed the students in the Google search engine group. Goda et al. (2014) showed that English as a Foreign Language (EFL) students' critical thinking skills were also enhanced after working with the chatbot, and they were more engaged in learning. Similarly, Wang and Petrina (2013) suggested chatbot is more beneficial to intermediate or lower levels of language learners, as chatbot can be designed to repeat the same materials. These findings suggest that interaction with a chatbot serves great potential for students to engage in the learning process.

Students are motivated when they have someone to talk to during instructions. Studies indicated that a chatbot made the learning task more manageable, and the students were enjoyable by interacting with the chatbot (Heller et al., 2005; Huang, Hew, & Gonda, 2019; Kerly et al., 2007). Furthermore, Huang et al. (2019) designed three chatbots using IBM Watson Assistant (multiple-choice questions, case study, and dictionary FAQs chatbots) to assist with graduate student learning. The design of the chatbots combined video lectures, online quizzes, and answering questions. Although the majority of the students showed neutral and positive experiences with the chatbots, some had negative experiences. Those students felt the chatbot did not speak like a human. Such a result was consistent with Heller et al. (2005), and this indicated that the character of a chatbot is critical. They further noted that natural language understanding (NLU) is a significant limitation of the chatbot because open-ended questions and unstructured problems may confuse the chatbot. Consequently, Kerly et al. (2007) and Wang and Petrina (2013) advised using student-produced data (i.e., dialogues) to refine the chatbot and to overcome NLU limitation. However, their suggestion requires a large amount of student data.

Although a chatbot offers potentials for enhancing student learning, research attempting to implement a chatbot often structures the chatbot with text-based, unnatural scripts. Most importantly, the effects of a chatbot on student writing improvement are still underexplored, because many studies have focused on either language learning in general (Bii, 2013; Goda et al., 2014; Kerly et al., 2007; Wang & Petrina, 2013), or improving students' thinking process (Abbasi & Kazi, 2014; Heller et al., 2005; Huang et al., 2019).

From the technical perspective, based on the literature of the chatbot design (Jain et al., 2018), we grounded our chatbot within Jain's et al. chatbot design recommendations. Thus, we incorporated button clicking, quizzes, and question answering functions and dog pictures to help students write a thesis statement and hope this design will facilitate students' positive learning experiences.

From instructional perspectives, the chatbot was designed primarily for reinforcing process writing instructions (Flower & Hayes, 1987; Graham & Sandmel, 2011). Generating a thesis statement is a key to successful argumentative essays (De Rycker & Ponnudurai, 2011), whereas providing feedback to peers' work might improve the quality of a draft (Guardado & Shi, 2007; Rollinson, 2005). Furthermore, learning to write can be reinforced by social interaction when novice writers begin to engage the conversational process with an agent, such as a chatbot (Cazden, 1987; Kalina & Powell, 2009). Therefore, the instructional features of the chatbot used in the present study mainly reinforce a small area in the writing process, such as helping students to generate a thesis statement.

2.3. Challenges of writing thesis statements

In student writing, there is always a mismatch between their thesis statements and the supporting claims (Cekiso, Tshotsho, & Somniso, 2017; Miller & Pessoa, 2016), or sometimes their thesis statements are absent (Cekiso et al., 2017; Owusu & Yeboah, 2014). For instance, when Miller and Pessoa (2016) investigate history students' difficulties in writing thesis statements, their results indicate that the student thesis statements appear to be too general, lack contextualization, or mismatch between thesis statements and the supporting claims. Similarly, Cekiso et al. (2017) found that first-year foreign multilingual writers present similar coherence problems when they are asked to produce thesis statements on a controversial current event. The coherence problems include an absence of thesis statements, conclusions not related to the thesis statements, and confusing long sentences. Overall, the challenges of writing thesis statements seem to have a strong influence regarding its presence, rhetorical function, location, disciplinary writing practice or genre. Particularly, in writing general argumentative essays, students need to be taught that thesis statements need to be contextualized and positioned right within the introductory paragraph. Therefore, these point to a pedagogical need for a writing activity that has a strong focus on helping writers to come up with thesis statements. If a chatbot can fulfill this need, we think the chatbot will play a strong supplementary role in helping writers to develop a stronger thesis statement needed for their argumentation.

2.4. Peer review instruction

Peer review is an essential part of the writing process. Many writing instructors implement peer review as part of their writing courses. The basis of using peer review in writing instructions has two important theoretical components: process writing instructions and peer learning. In literature, peer review is part of the process writing model approach developed in the 1980s (Flower & Hayes, 1981; Keh, 1990). In the process approach, writing is a multi-staged, multi-drafted process in which students generate different versions of their work based on the feedback they receive from their peers. If necessary, the entire process could be repeated until the draft is ready as a final product.

Several studies indicate that peer review offers instructional benefits, such as reducing instructors' grading load and improving students' writing practices (Cho & Schunn, 2007; Cho & MacArthur, 2010; Cho & Cho, 2011). For instance, writing instructors adopt peer review for assessment purposes, and student writers use peer review as a reference point to improve the quality of their writing. Moreover, several studies suggest that student view peer feedback more positively than instructor feedback because peer feedback provides more extensive detailed views than instructor feedback (Cheng, Liang, & Tsai, 2015; Cho & Cho, 2011; Cho & MacArthur, 2010; Cho & Schunn, 2007; Topping, Smith, Swanson, & Elliot, 2000). Research has also found peer review activity not only enhances students' critical thinking skills but also facilitates social interactions and course engagements among peers (Kulkarni, Kotturi, Bernstein, & Klemmer, 2016). Taken together, these findings not only suggest that incorporating peer review may help to cultivate students' writing skills and reduce instructors' teaching load but also point to a need for educators to think of an innovative way to motivate writers in the peer review process.

3. Overview of the writing Chatbot DD

The learning design of the chatbot DD was based on Jain's et al. (2018) design framework. We developed the chatbot DD using Rasa (version: rasa_core 0.11.12, rasa_core_sdk: 0.11.5, rasa_nlu 0.13.7), an open-source conversational AI framework (Bocklisch, Faulkner, Pawlowski, & Nichol, 2017). We chose Rasa because Rasa emphasizes the needs of non-specialist software developers in the research field (Bocklisch et al., 2017). Figure 1

shows the design of the chatbot DD. The Rasa core module processes dialogues within a domain (the universe where the bot lives in). The Rasa core SDK module includes customizable actions (e.g., bridging the connection between Rasa core and the database). The training data for the chatbot are stored in the Rasa NLU module. The chatbot is then deployed on a web-based server using Chatroom API. The students can access the chatbot through a laptop, cellphone, or tablet, as shown in Figure 2. Following Jain's et al. (2018) suggestion when dealing with a failure conversation, the student can follow the instruction if the conversation with the chatbot encounters an error, as demonstrated in Figure 2.
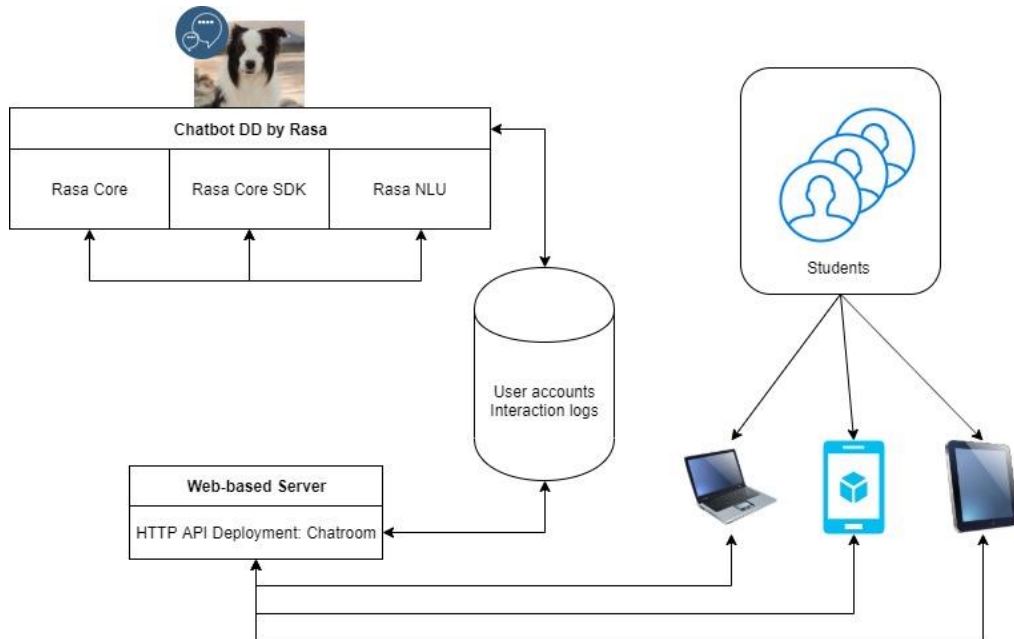

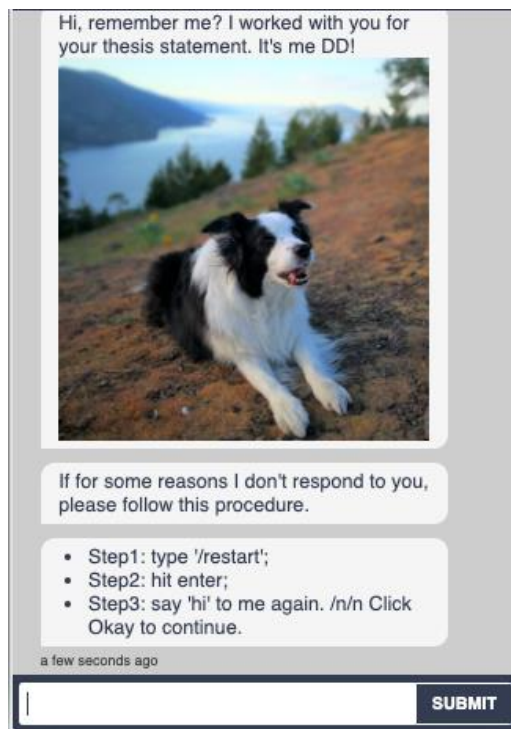*Figure 1.* The framework of the chatbot DD

Figure 2. A web-based chatbot interface

The design of the chatbot DD was consulted with the course instructor and a former graduate-level teaching assistant. The chatbot first greets students and asks the student's identification (ID) number. This ID is stored in the database for future referencing and data analysis. Three design principles are taken from Jain's et al. recommendations (2018): field-specificity, embedded multimedia resources, and fallback response.

There were several challenges when designing the chatbot. One of the major issues is the standability of NLU. Similar phrasing, synonyms, or grammatical errors may confuse the chatbot (Clarizia et al., 2018; Huang et al., 2019; Jain et al., 2018). For example, "I want to improve my thesis statement because my teacher found I made many grammar mistakes and highlighted many errors in my essay. I don't know what to do now." Such a long sentence from a student may cause parsing errors and confusion. Because when the student says, "I want to …" and "I don't know what to do…", the chatbot may not correctly parse the student's sentence and understand whether the student needs help (I want do…) or needs clarification given pre-existing knowledge (I don't know what to do…).

Thus, to minimize the unpredictability of user input and confusion, we structured the chatbot DD by employing a button-clicking function. The fallback response will be triggered when the chatbot DD is unable to recognize a student's typed sentences. The chatbot DD recommends the student follow a certain step to re-start the program, as shown in Figure 2.

Moreover, we tried to structure the chatbot DD to speak the human-like natural language with the use of some dog pictures, which would motivate student learning and facilitate their engagement (Fryer & Carpenter, 2006; Jain et al., 2018). The scripts have been evaluated by three experienced graduate students, who have had post-secondary teaching experience, to ensure the conversation is friendly and natural. Figure 3 shows how the chatbot explained the course concept and the features of a thesis statement to the students. Figure 4 illustrates an example of student interaction. The Yes/No judgement was the concept checking questions (CCQs) after the students received lessons from the chatbot. Students can make their initial learning judgment, followed by confirmation of their judgement from the chatbot (Huang et al., 2019). All the interactions with the chatbot are automatically saved in the Rasa server.
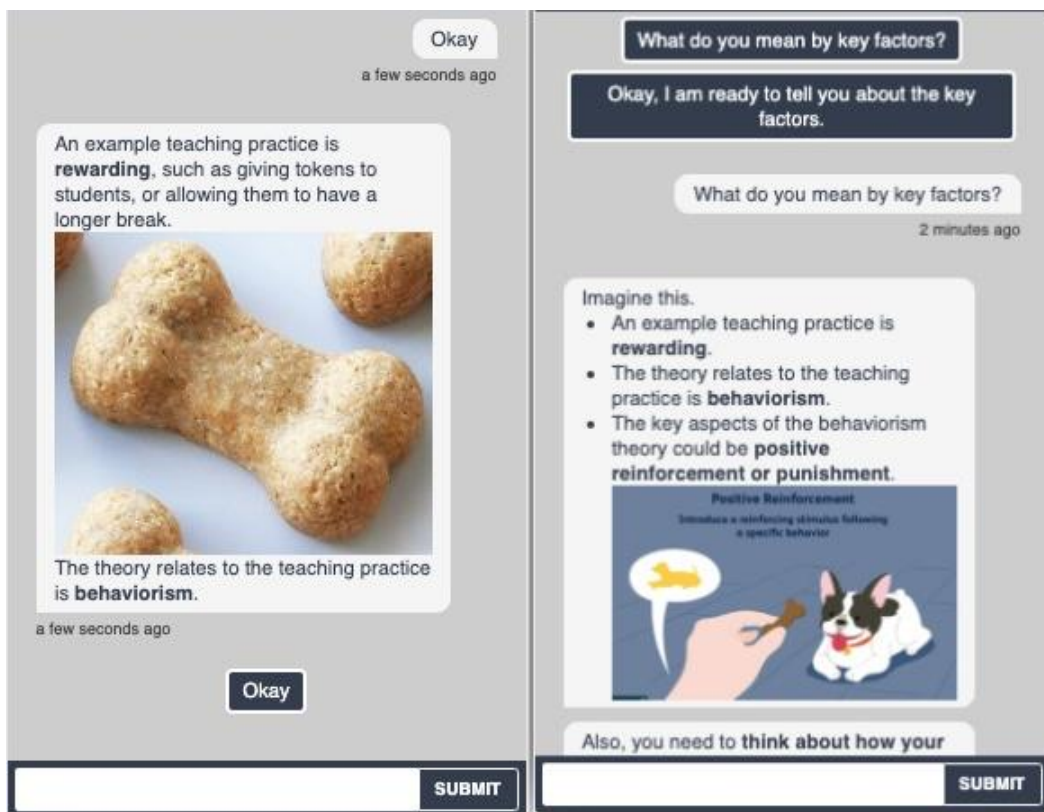

Figure 3. The chatbot DD explained the concepts of thesis statement to the students
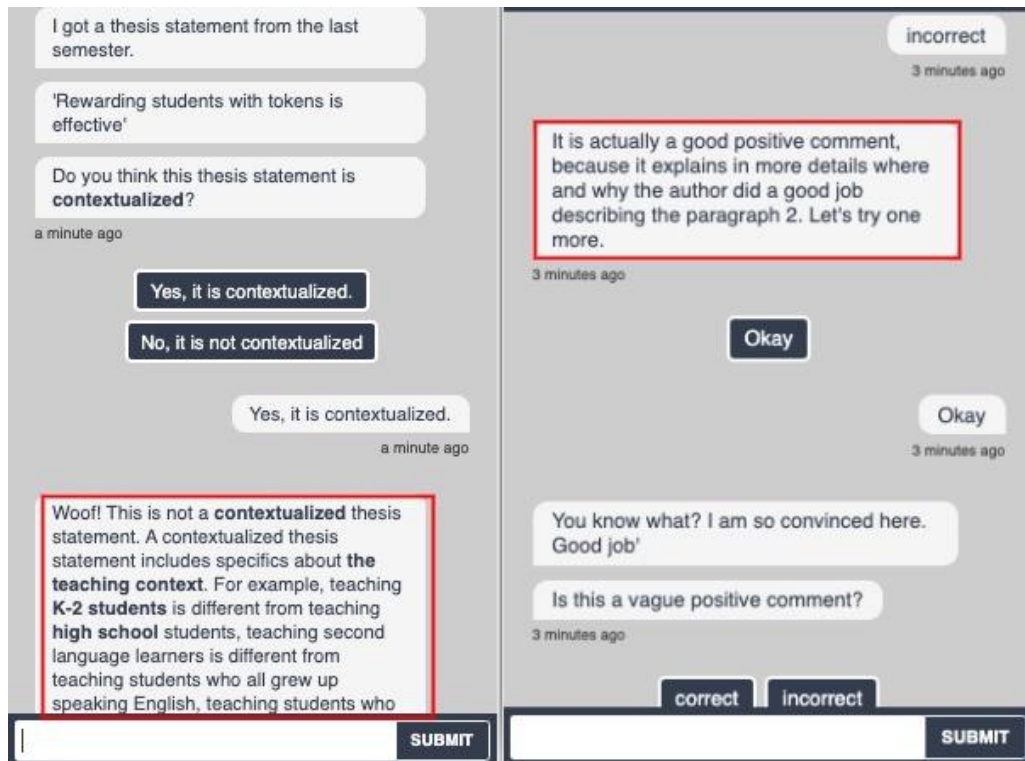
Figure 4. The chatbot tested student understanding with explanations

4. Methods

4.1. Participants

The participants were recruited from a large introductory educational psychology class (ED 100) offered in two consecutive semesters (i.e., Fall 2018 and Spring 2019) at a western Canadian university. The course did not have any prerequisites. The classes offered in two consecutive semesters were identically taught by the same instructor using the same curriculum. At this university, writing was not taught as an independent skill set, yet writing was integrated as part of a subject matter curriculum. A major flaw of this study was that no official measure of writing proficiency was administered before the intervention. The course was elective; every student could enroll if they intended to declare education as the major study subject. We ran an independent sample t-test on the midterm examination, a measure before they turned in their essay outlines. There was no statistically detectable difference between the two semesters. So, this implied that these students started at the same level of disciplinary concept knowledge before they began the essay-outline assignment.

The course had two components. One was a two-hour lecture, where the instructor taught the curriculum to the students, whereas the other was a one-hour tutorial class, where the teaching assistants led the class and answered questions about the course content and assignment expectations. Since this was a large class, there were 11 tutorials in which each tutorial had roughly 15~18 students. A teaching assistant was responsible for roughly 3~4 tutorials. There were three teaching assistants for this class in total. A semester at this university was roughly 13-week long. The students attend the two-hour lecture and one-hour tutorial each week.

The chatbot, which assisted their essay outline writing assignment, was introduced to the tutorial classes in week 6 and week 7 of Spring 2019. In week 6, the students were introduced the chatbot DD, which the chatbot DD will help their generation of thesis statements. During the week 6 tutorial, the students interacted with the chatbot in a computer lab room and came up with a thesis statement for their essay outline assignment. In week 7, the students were expected to bring a draft of their essay outline assignment to the class. The teaching assistants and the researchers led the peer review activity with the chatbot.

There were 190 undergraduate students in the Fall of 2018, and 167 students in the Spring of 2019 class. As mentioned above, in the Fall of 2018, the students were not introduced the chatbot (comparison group). In contrast, in the Spring of 2019, the chatbot was introduced to each tutorial (treatment group). With their written consent, there were 28 students ($n = 28$) from the spring cohort who agreed to fill out the questionnaire about their learning experiences with the chatbot DD. In order to comply with the institutional research ethics obligations, we only selected the students who have granted permission for us to analyze their questionnaire data.

4.2. Instruments

There were three instruments used in this study: the chatbot, an essay outline, and a questionnaire, respectively. First, the chatbot was designed to assist students with improving their thesis statement for the essay outline assignment. The essay outline was a graded component of the course determined by the instructor. The outline served as the planning stage for the students to draft their ideas for the final argumentative essay assignment (Flower & Hayes, 1981). In our study, the essay outline was the measure of student writing achievement. The essay outline contained several major rhetorical features of an argumentative essay, such as a thesis statement, topic sentences, evidence to support the topic sentences, counterarguments, and a conclusion. Each student needed to identify a teaching practice and argued for why the teaching practice of their choice can motivate students to learn, drawing on their course knowledge from the motivational theories of Educational Psychology. Three teaching assistants graded the essay outlines in the course. Before grading, the instructor calibrated the consistency of scoring by hosting a two-hour-long meeting. During the calibration, the teaching assistants were introduced the marking rubric of the entire outline, and then they were assigned one student essay outline for a grading attempt. The instructor repeated the process until the teaching assistants achieved consistency. The grades of the essay outline were used as the quality measure for the students in both experimental and comparison semesters. The marking rubric contained thesis statement (3 marks), arguments (3 marks), and counterarguments, including a rebuttal (4 marks). So, the essay outline was worth 10 points of the course grade. The questionnaire (see Appendix A) was adopted from Schunn, Godley, and DeMartino (2016), Topping et al. (2000), and Torrance, Thomas, and Robinson (1994). The questionnaire measured students' experience with the writing chatbot. Yes/No questions were used to avoid ambiguity in the survey statements and socially desirable responses (Mick, 1996).

4.3. Procedure

Each student in both experimental and comparison semesters was required to submit one essay outline as one of their course assignments. Each student spent two weeks working on their essay outline assignment. The students in the comparison semester wrote the essay outline without interacting with the chatbot. In the experimental semester of weeks 6 and 7, the students interacted with the chatbot to learn to construct a thesis statement for their essay outline, which was a major rhetorical feature of an essay outline. Each tutorial class was fifty minutes. During the first week of the class, drawing discipline-specific examples from Educational Psychology, the chatbot DD introduced the components of a thesis statement and guided the students to write a thesis statement for their essay outline. In the second week of the class, the students reviewed another students' outline by interacting with the chatbot DD. The chatbot DD was programmed with the ability to guide the students in providing effective peer feedback for the outline. The students then submitted their essays in week 13. At the end of the semester, the consent form was distributed to the students in the experimental semester, and they had time to fill out a questionnaire regarding their experiences with the chatbot DD until the semester ended. The questionnaire can be found in Appendix A.

4.4. Data analysis

Two types of data were collected. Qualitative data was the student responses to the open-ended questions of the questionnaire (Q33, Q36, Q37, Q38, and Q39), whereas quantitative data were (1) the student grades from the essay outline assignment and (2) the student responses to the yes/no questions of the questionnaire. Thus, data collected in this study included the score of the essay outline and a questionnaire regarding the experience with the chatbot. In this study, we used mixed methods to examine the data quantitatively and qualitatively. Studies showed that mixed methods provide a holistic and valid view by exploring the in-depth and effects of an innovative tool (i.e., a chatbot) from instructional and student perspectives in the field of educational technology (Creswell & Clark, 2017;

Randolph, 2008). An effective approach to analyze students' open-ended questions is content analysis, as this method uncovers and explores student data to generate inferences about their chatbot learning experience (Patton, 1990; Weber, 1990; Yang, 2010). For the yes/no questions from the questionnaire, we presented descriptive statistics for the student perception of the Chatbot DD. Furthermore, an independent sample t-test was used to examine the difference in writing achievement between the two semesters.

5. Results

5.1. Comparison of the writing achievements between the comparison and experimental semesters

As mentioned in the section of Participants, the students of the two groups were at the same level of disciplinary concept knowledge before they began the essay-outline assignment. Therefore, the outline scores were used to evaluate the effects of the chatbot on students writing achievements.

190 students who were in the comparison semester without using the chatbot and 167 students were in the experimental semester using the chatbot. An independent sample t-test was conducted. The outline scores in the experimental semester ($M = 7.27$, $SD = 2.52$) were statistically better than the scores in the comparison semester ($M = 7.18$, $SD = 2.14$, $t = -0.38$, $p = 0.027^*$).

5.2. Student perceptions towards the Chatbot DD

Table 1 summarizes the results of students' experience with the chatbot DD. More than 80% of the students pointed out the chatbot DD helped them to identify new issues, to improve how to give effective feedback, and to become better reviewers. 78.6% of the students felt that interacting with the chatbot DD was enjoyable. 75% of the students mentioned the chatbot DD enhanced their skills in evaluating a thesis statement.

Table 1. Student experience with the writing chatbot DD

	Identified issues during peer review	Improved peer review feedback	Enjoyable	Became better reviewer	Resolve confusions of instructions	Construct precise thesis statement	Helped me evaluate my thesis statement
Yes	85.71% (24)	82.1% (23)	78.6% (22)	82.1% (23)	64.3% (18)	75.00% (21)	75.00% (21)
No	14.29% (4)	17.9% (5)	21.4% (6)	17.9% (5)	21.4% (6)	25.00% (7)	25.00% (7)
N/A	0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	14.3% (4)	0.0% (0)	0.0% (0)
Total	28	28	28	28	28	28	28

Note. The number within parentheses mean the numbers of students.

Notably, when asking the students in what specific the chatbot DD helped them in writing, seventeen of the students ($n = 17$) indicated the chatbot DD guided them to improve thesis statement, seven of the students ($n = 7$) pointed out the chatbot DD enhanced their skills on giving feedback, one of the students ($n = 1$) felt s/he improved on both skills. Three of the students ($n = 3$) noted the chatbot DD did not help them at all.

The students also suggested some improvements to the chatbot DD as shown in Table 2. Nine students ($n = 9$) recommended the chatbot should respond faster. Seven students ($n = 7$) wanted the chatbot to provide more feedback and examples on the topic they learn. Four students ($n = 4$) noted the explanation on a thesis statement and terms definition could be simplified. Some students had negative opinions about the chatbot because sometimes the response time was slow due to networking or NLU retrieval issues. Thus, they would instead seek help from the teaching assistants or instructor.

Table 2. Student responses to the open-ended questions from the questionnaire

Themes	Findings
Favouring the Chatbot DD's overall feedback and information	*"Detail specific feedback is beneficial"* *"It was really informative"*
User-friendly interface	*"I have the flexibility to skip a lesson or choose a certain topic to learn"* *"It is easy to use"* *"I can just click the buttons instead of typing the answers"* *"It was interactive"* *"It was like chatting with a dog"* *"···talking to a dog made the activity more enjoyable and less stressful"*
Promoting learning	*"The questions DD asked and prompted made me be more critical and reflective, and it benefits in a way being self-regulated"* *"It gave me time to think and reflect on thesis statement and arguments it helped me evaluate other arguments and create an expectation for my paper"*
Positive learning experience	*"Positive experience and greatly assisted in the writing of my thesis statement and peer feedback"* *"It was a relatively positive experience. I enjoyed working with the bot and it helped me create a much better thesis statement."* *"I think it was a fun, interactive way to improve our writing. It was something unique that I had never tried before which caught my attention!"*

6. Discussion

6.1. Do these chatbot-led writing activities enhance the students' writing achievement?

The general outcome of the present study might indicate the potential of using a chatbot as the instructional supplement to teach writing. This novel design of chatbot DD aims to supplement thesis-statement and peer-feedback instructions in our study. To our best knowledge, there has not been literature yet specifically integrating a chatbot in supporting writing instructions. However, there have been some active ITS systems that were developed to teach writing, such as the Writing Pal Intelligent Tutoring System (Roscoe, Allen, Weston, Crossley, & McNamara, 2014) and iStart (McNamara, Levinstein, & Boonthum, 2004). These ITS have been found effective and useful in teaching writing. Future research should carefully operationalize chatbot use in the context of ITS. It is because the development of an educational writing chatbot still bears some difficult technical realities, such as limitations in NLU, which students still cannot freely talk about their writing issues, and the chatbot cannot understand their problems. The NLU limitation is the reason why our chatbot has a structured dialogue. Overall, our finding needs to be interpreted with caution. First, no measure of a pre-test was administered, although the midterm examination implied that the participants were at the same level of the disciplinary concept knowledge before the treatment. Therefore, we reserve this conclusion for extended research. Second, the improved performance might be attributed to the novelty effect (Fryer, Nakao, & Thompson, 2019) – the tendency of increased achievement results from the initial introduction of an innovative technology in which users' interest level is high. In our study, the chatbot was introduced only for two weeks, so it was likely that the students felt very curious, and they showed a strong interest in using the chatbot. Future research might consider observing whether performance still sustained if the chatbot was introduced for a longer period.

6.2. How do students perceive the use of chatbot in a university classroom

When interacting with the chatbot, among those who responded to the questionnaire, 75% of them felt that DD helped them write a precise thesis statement and evaluate the quality of the thesis statement. These findings add

further support to the supplementary role of technology in writing instructions and the importance of conversation when writers are learning to write (Bii, 2013; Kalina & Powell, 2009). Furthermore, consistent with previous research reporting positive student experiences with a chatbot (Fryer & Carpenter, 2006; Goda et al., 2014; Kerly et al., 2007; Luger & Sellen, 2016), we have found that approximately 79% of the students reported an enjoyable experience with the chatbot. Our finding might suggest that integrating a chatbot with writing instruction might improve student learning-to-write engagement.

Asking students to review their peers' essays requires extensive instructional explanations (Cho & Schunn, 2007; Li, Liu, & Steckelberg, 2010; Yang, 2011). In our study, 85% of the students reported the chatbot helped them identify their writing issues; 82% of the students felt they become a better reviewer with improved peer feedback quality, and 64% of the students mentioned the chatbot helped them to resolve confusions of the peer review instructions. Therefore, using a chatbot on the side of a peer review activity might solve instructional confusion and benefit students in becoming better peer reviewers.

On the design and development side of the chatbot, it seems the students enjoyed learning with the chatbot because the chatbot has multimedia features, such as using an animal dog picture, friendly language, and button clicking. From the open-ended survey data, some students liked the button clicking feature as it made the chatbot user-friendly. Furthermore, some students liked the chatbot was embodied by a dog picture, which made the overall learning experience more entertaining. These findings are consistent with previous research in educational chatbot design that directive, friendly language and the button clicking function are crucial in engaging student learning and minimize the difficulties of learning tasks (Heller et al., 2005; Huang et al., 2019; Jain et al., 2018; Kerly et al., 2007).

7. Conclusion and limitations

In conclusion, a writing chatbot was introduced as part of a disciplinary writing class. Although some positive student testimonials and improved essay outline performance were reported, the present study has some limitations. Particularly, the writing proficiency of the participating students was not well-controlled. In the present study, administering a pre-test was not administered. Future research examining the effect of a chatbot on writing achievement need to take students' writing proficiency into account for analysis. Secondly, the improved performance of the essay outline might be due to the novelty effect. Future researchers might extend the use of chatbot over a longer period to see if performance and student interest levels still sustain (Fryer et al., 2019). Despite several methodological concerns, we argue that chatbot has by far used as a supplement instead of a standalone ITS because of its NLU constraint (Heller et al., 2005).

Also, we developed the chatbot based on the existing design framework, including external database storage, conversation failure mechanism, images and quizzes combination, delivering effective and natural conversation, and user-friendly interface. What advances the current research will be collecting and establishing a large-scale repository of student writing issues and its solutions to develop a classification system that correctly identifies a student's learning dialogue (Huang et al., 2019; Jain et al., 2018; Wang & Petrina, 2013). Future chatbot designers and researchers are recommended to overcome the limitation of NLU, so students can freely talk to a chatbot without encountering conversational errors, or fallback.

Acknowledgement

References

Abbasi, S., & Kazi, H. (2014). Measuring effectiveness of learning chatbot systems on student's learning outcome and memory retention. *Asian Journal of Applied Science and Engineering*, *3*(2), 251-260. doi:10.15590/ajase/2014/v3i7/53576

Bii, P. (2013). Chatbot technology: A Possible means of unlocking student potential to learn how to learn. *Educational Research*, *4*(2), 218-221.

Bocklisch, T., Faulkner, J., Pawlowski, N., & Nichol, A. (2017). *Rasa: Open source language understanding and dialogue management*. arXiv:1712.05181 [cs.CL]

Cazden, C. B. (1988). *Classroom discourse: The language of teaching and learning*. Portsmouth, NH: Heinemann.

Cekiso, M., Tshotsho, B., & Somniso, M. (2017). Exploring first-year university students' challenges with coherence writing strategies in essay writing in a South African university. *International Journal of Educational Sciences*, *12*(3), 241-246. doi:10.1080/09751122.2016.11890431

Cheng, K. H., Liang, J. C., & Tsai, C. C. (2015). Examining the role of feedback messages in undergraduate students' writing performance during an online peer assessment activity. *The Internet and Higher Education*, *25*, 78-84. doi:10.1016/j.iheduc.2015.02.001

Cho, Y. H., & Cho, K. (2011). Peer reviewers learn from giving comments. *Instructional Science*, *39*(5), 629-643. doi:10.1007/s11251-010-9146-1

Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction*, *20*(4), 328-338. doi:10.1016/j.learninstruc.2009.08.006

Cho, K., & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline: A Web-based reciprocal peer review system. *Computers & Education*, *48*(3), 409-426. doi:10.1016/j.compedu.2005.02.004

Clarizia, F., Colace, F., Lombardi, M., Pascale, F., & Santaniello, D. (2018). Chatbot: An Education support system for student. In *International Symposium on Cyberspace Safety and Security* (pp. 291-302). Amalfi, Italy: Springer. doi:10.1007/978-3-030-01689-0_23

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage publications.

De Rycker, A., & Ponnudurai, P. (2011). The Effect of online reading on argumentative essay writing quality. *GEMA Online® Journal of Language Studies*, *11*(3), 147-162.

Edwards, D., & Mercer, N. (2013). *Common Knowledge (Routledge Revivals): The Development of understanding in the classroom*. London, United Kingdom: Routledge.

Flower, L., & Hayes, J. R. (1981). A Cognitive process theory of writing. *College composition and communication*, *32*(4), 365-387. doi:10.2307/356600

Fryer, L., & Carpenter, R. (2006). Bots as language learning tools. Language learning and technology. *Language Learning & Technology*, *10*(3), 8-14. doi:10125/44068

Fryer, L. K., Nakao, K., & Thompson, A. (2019). Chatbot learning partners: Connecting learning experiences, interest and competence. *Computers in Human Behavior*, *93*, 279-289. doi:10.1016/j.chb.2018.12.023

Goda, Y., Yamada, M., Matsukawa, H., Hata, K., & Yasunami, S. (2014). Conversation with a chatbot before an online EFL group discussion and the effects on critical thinking. *The Journal of Information and Systems in Education*, *13*(1), 1-7. doi:10.12937/ejsise.13.1

Ghose, S., & Barua, J. J. (2013). Toward the implementation of a topic specific dialogue based natural language chatbot as an undergraduate advisor. In *2013 International Conference on Informatics, Electronics and Vision (ICIEV)* (pp. 1-5). Dhaka, Bangladesh: IEEE. doi:10.1109/ICIEV.2013.6572650

Graham, S., & Sandmel, K. (2011). The Process writing approach: A Meta-analysis. *The Journal of Educational Research*, *104*(6), 396-407 doi:10.1080/00220671.2010.488703

Guardado, M., & Shi, L. (2007). ESL students' experiences of online peer feedback. *Computers and Composition*, *24*(4), 443-461. doi:10.1016/j.compcom.2007.03.002

Heller, B., Proctor, M., Mah, D., Jewell, L., & Cheung, B. (2005, June). Freudbot: An Investigation of chatbot technology in distance education. In *EdMedia+ Innovate Learning* (pp. 3913-3918). Waynesville, NC: Association for the Advancement of Computing in Education (AACE).

Huang, W., Hew, K. F., & Gonda, D. E. (2019). Designing and evaluating three chatbot-enhanced activities for a flipped graduate course. *International Journal of Mechanical Engineering and Robotics Research*, *8*(5). doi:10.18178/ijmerr

Jain, M., Kumar, P., Kota, R., & Patel, S. N. (2018). Evaluating and informing the design of chatbots. In *Proceedings of the 2018 on Designing Interactive Systems Conference 2018* (pp. 895-906). Hong Kong, China: ACM. doi:10.1145/3196709.3196735

Kalina, C., & Powell, K. C. (2009). Cognitive and social constructivism: Developing tools for an effective classroom. *Education*, *130*(2), 241-250.

Keh, C. L. (1990). Feedback in the writing process: A Model and methods for implementation. *ELT Journal*, *44*(4), 294–304. doi: 10.1093/elt/44.4.294

Kerly, A., Hall, P., & Bull, S. (2007). Bringing chatbots into education: Towards natural language negotiation of open learner models. *Knowledge-Based Systems*, *20*(2), 177-185. doi:10.1016/j.knosys.2006.11.014

Kirakowski, J., O'Donnell, P., & Yiu, A. (2009). Establishing the hallmarks of a convincing chatbot-human dialogue. *Human-Computer Interaction*, 49-56. doi:10.5772/7741

Kulkarni, C., Kotturi, Y., Bernstein, M. S., & Klemmer, S. (2016). Designing scalable and sustainable peer interactions online. In *Design Thinking Research* (pp. 237-273). doi:10.1007/978-3-319-40382-3_14

Li, L., Liu, X., & Steckelberg, A. L. (2010). Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British journal of educational technology*, *41*(3), 525-536. doi:10.1111/j.1467-8535.2009.00968.x

Luger, E., & Sellen, A. (2016). Like having a really bad PA: the gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 5286-5297). San Jose, CA: ACM. doi:10.1145/2858036.2858288

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A Meta-analysis. *Journal of educational psychology*, *106*(4), 901-908. doi:10.1037/a0037123

McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy training for active reading and thinking. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 222-233. doi:10.3758/BF03195567

Mick, D. G. (1996). Are studies of dark side variables confounded by socially desirable responding? The Case of materialism. *Journal of consumer research*, *23*(2), 106-119. doi:10.1086/209470

Miller, R. T., & Pessoa, S. (2016). Where's your thesis statement and what happened to your topic sentences? Identifying organizational challenges in undergraduate student argumentative writing. *TESOL Journal*, *7*(4), 847-873. doi:10.1002/tesj.248

Murray, T. (1999). Authoring intelligent tutoring systems: An Analysis of the state of the art. *International Journal of Artificial Intelligence in Education*, *10*, 98-129

Owusu, E., & Adade-Yeboah, A. (2014). Thesis statement: A Vital element in expository essays. *Journal of Language Teaching & Research*, *5*(1), 56-62.

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage.

Pereira, J., & Díaz, Ó. (2018). Chatbot dimensions that matter: Lessons from the trenches. In *International Conference on Web Engineering* (pp. 129-135). Cáceres, Spain: Springer. doi:10.1007/978-3-319-91662-0_9

Randolph, J. J. (2008). *Multidisciplinary methods in educational technology research and development*. Hämeenlinna, Finland: HAMK Press/Justus Randolph.

Rodríguez-Gil, L., García-Zubia, J., Orduña, P., Villar-Martinez, A., & López-De-Ipiña, D. (2019). New approach for conversational agent definition by non-programmers: A Visual domain-specific language. *IEEE Access, 7*, 5262-5276. doi:10.1109/ACCESS.2018.2883500

Rollinson, P. (2005). Using peer feedback in the ESL writing class. *ELT journal*, *59*(1), 23-30. doi:10.1093/elt/cci003

Roscoe, R. D., Allen, L. K., Weston, J. L., Crossley, S. A., & McNamara, D. S. (2014). The Writing Pal intelligent tutoring system: Usability testing and development. *Computers and Composition*, *34*, 39-59. doi:10.1016/j.compcom.2014.09.002

Schunn, C., Godley, A., & DeMartino, S. (2016). The Reliability and validity of peer review of writing in high school AP English classes. *Journal of Adolescent & Adult Literacy*, *60*(1), 13-23. doi:10.1002/jaal.525

Song, D., Oh, E. Y., & Rice, M. (2017). Interacting with a conversational agent system for educational purposes in online courses. In *2017 10th international conference on human system interactions (HSI)* (pp. 78-82). Ulsan, South Korea: IEEE. doi:10.1109/HSI.2017.8005002

Topping, K. J., Smith, E. F., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment & evaluation in higher education*, *25*(2), 149-169. doi:10.1080/713611428

Torrance, M., Thomas, G. V., & Robinson, E. J. (1994). The Writing strategies of graduate research students in the social sciences. *Higher education*, *27*(3), 379-392. doi:10.1007/BF03179901

Vanlehn, K. (2006). The Behavior of tutoring systems. *International journal of artificial intelligence in education*, *16*(3), 227-265.

Wang, Y. F., & Petrina, S. (2013). Using learning analytics to understand the design of an intelligent language tutor–Chatbot Lucy. *International Journal of Advanced Computer Science and Applications*, *4*(11), 124-134. doi:10.14569/IJACSA.2013.041117

Weber, R. P. (1990). *Basic content analysis* (2nd ed.). Newbury Park, CA: Sage.

Weizenbaum, J. (1966). ELIZA-a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36-45. doi:10.1145/365153.365168

Xu, D., & Wang, H. (2006). Intelligent agent supported personalization for virtual learning environments. *Decision Support Systems*, *42*(2), 825-843. doi:10.1016/j.dss.2005.05.033

Yang, Y. F. (2010). Students' reflection on online self-correction and peer review to improve writing. *Computers & Education*, *55*(3), 1202-1210. doi:10.1016/j.compedu.2010.05.017

Yang, Y. F. (2011). A Reciprocal peer review system to support college students' writing. *British Journal of Educational Technology*, *42*(4), 687-700. doi:10.1111/j.1467-8535.2010.01059.x

Appendix A. Questionnaire regarding student experience with the chatbot

1. Student ID: _____
2. Name: _____
3. Email: _____
4. Major: _____
5. Academic residency: _____
6. EAL: _____
7. In the past 6 months, which of these forms of writing have you engaged in? (Please select all that apply):
   a. Plans and notes (taking notes in class)
   b. Reports and assigned work (writing emails, cover letters for jobs)
   c. Writing for publication (e.g., writing a book or blog posts)
   d. Research term paper within a course (e.g., Literature review for PHIL 120; persuasive paper for FALx99)
   e. High school essays or provincial exams (e.g., English 12 or equivalent)
   f. Other (please specify): _____
8. When you write, what strategies do you always adopt or use. (choose up to 3):
   a. Brainstorming
   b. Taking notes from research sources
   c. Mindmapping
   d. Ordering notes
   e. Making an outline
   f. Drafting
   g. Revising
   h. Sharing ideas with a friend and receiving feedback
   i. Other
9. When you revise your paper, what are your goals? (choose up to 3)
   a. Improving clarity
   b. Improving style
   c. Developing content
   d. Correcting errors
   e. Rearranging the text
   f. Reducing length

10.   Generally speaking, at what point do you like to start writing?
11.   Structuring my arguments for my term paper is relatively easy for me (Y/N)
12.   I would describe myself as a poor writer (Y/N)
13.   I find writing a thesis statement is difficult (Y/N)
14.   I worry a lot about whether the grammar and spelling is correct for my thesis statement (Y/N)
15.   I worry that my difficulty with writing will jeopardize completing my essay (Y/N)
16.   I am not good at coming up with a thesis statement (Y/N)
17.   I gain a great deal of pleasure from writing (Y/N)
18.   I find writing a frustrating process (Y/N)
19.   I worry the clarity of my thesis statement will affect my paper grade (Y/N)
20.   I find the process of writing highly stressful (Y/N)
21.   I find writing a thesis statement pretty frustrating (Y/N)
22.   I find the process of coming up with a thesis statement quite stressful (Y/N)
23.   The easiest part of the writing process is producing a plan (Y/N)
24.   Structuring my arguments to form a well-structured thesis statement is relatively easy for me (Y/N)
25.   I find writing hard work (Y/N)
26.   I worry a lot about whether my grammar and spelling are correct (Y/N)
27.   Working with DD the Thesis Bot helped me construct my thesis statement more precisely (Y/N)
28.   Working with DD helped me evaluate my own thesis statement (Y/N)
29.   The bot DD helps me identify new issues with my peer's dialectical map (Y/N)
30.   The bot DD helps improve my feedback for my peer's dialectical map (Y/N)
31.   What specifically did DD help me? (Open-ended)
32.   When I worked with DD, he helped me change the quality of my review (Y/N)
33.   Tell me how did (or did not) your review change because you engaged with the bot (Open-ended)
34.   Working with DD the Peer Review/Thesis Bot was enjoyable (Y/N)
35.   Working with DD taught me how to be a better reviewer of my peer's work (Y/N)
36.   Did you find working with DD the Peer Review/Thesis Bot helped you resolve some confusions from the instructions? If yes, what specifically was resolved? (Open-ended)
37.   Describe the experience you have had in the peer-review/thesis chatbot (Open-ended)
38.   What did you like about the DD chatbot? (Open-ended)
39.   What suggestions would you provide to make the chatbot more effective? (Open-ended)