Computers & Education 179 (2022) 104395

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

FAQ chatbot and inclusive learning in massive open online courses

Songhee Han (한송희)[a,*], Min Kyung Lee (이민경)[b]

[a] Curriculum and Instruction, The University of Texas at Austin, USA
[b] School of Information, The University of Texas at Austin, USA

ARTICLE INFO

Keywords:
Chatbot
Massive open online courses
Human and computer interaction
Inclusive learning environment
Accessibility

ABSTRACT

Recognizing the research gap involving the lack of equity considerations in new technology implementation, this study compares students' learning experiences when using an FAQ chatbot with using an FAQ webpage. We trained a natural language processing–based chatbot utilizing content from an FAQ webpage and deployed it in two journalism massive open online courses (MOOCs) with 46 students and compared their experiences with 74 students' experiences with the FAQ webpage as a baseline. There were equal numbers of male and female students, their ages ranged from 18 to 65+, and they hailed from 45 unique countries. Considering the importance of supporting students with an inclusive Q&A experience before implementing any new technology into real-world operation, this study investigates students' disparate Q&A experiences by measuring their intention to use the interface as well as perceived Q&A service quality, enjoyment, and barriers utilizing a between-subjects online experiment. The results indicate that the students preferred an FAQ webpage over an FAQ chatbot, and the chatbot users experienced a higher magnitude of barriers compared to the webpage users. For the chatbot users, we found that region and native language factors influenced their Q&A experiences significantly. We discussed the meaning of the students' disparate experiences from multiple perspectives—namely, human-computer interaction, MOOC context, and technologies as social practice aspects. Lastly, we determined how feasible it is to provide an inclusive learning experience for the MOOC population with the FAQ chatbot, based on the contextualized meaning of MOOC inclusiveness in current literature. This study suggests multi-faceted aspects to consider when adopting new technologies in MOOCs to provide an inclusive learning experience, and underscores the need for more active research in chatbot use to serve diverse student needs in MOOCs.

1. Introduction

Frequently asked questions (FAQ) webpages are widely used by massive open online course (MOOC) providers to bridge common knowledge gaps among students in an accessible manner. However, it is common to see students become frustrated with FAQ webpages and turn to alternatives such as writing emails or making queries on forums; then inevitably many of those students complain about the lack of responses to their queries from MOOC providers (Chatterjee et al., 2020). Also, sometimes a text-heavy FAQ webpage creates information overload for students who simply cannot find what they are looking for despite the answers being on the page (Shaw, 2012). As a result, FAQ webpages often increase students' dissatisfaction with MOOC providers' Q&A services. Ideally, MOOC providers would hire more staff to improve their Q&A service and produce more prompt responses for their students via emails or

* Corresponding author. Curriculum and Instruction, The University of Texas at Austin, Austin, TX, 78712, USA.
E-mail address: song9@utexas.edu (S. Han).

https://doi.org/10.1016/j.compedu.2021.104395
Received 7 June 2021; Received in revised form 28 November 2021; Accepted 29 November 2021
Available online 1 December 2021
0360-1315/

forum posts; however, this requires them to invest their limited resources in costly staff recruitment, retention, and quality assurance (Chatterjee et al., 2020).

To solve this problem, many websites are using natural language processing–based chatbots instead of FAQ webpages to provide more prompt responses to individual queries. Chatbots are software programs that communicate with users through natural language interaction interfaces (Rubin et al., 2010; Shawar & Atwell, 2007). Although chatbots would not solve the fundamental problem of FAQ-based Q&A services not being able to answer every possible question, they can provide human-like responses through dialogue-based interactions, which feel more immediate and customized compared to webpages. Since multiple previous studies indicated the effectiveness of chatbot adoption in online learning spaces (e.g., Huang et al., 2019; Kloos et al., 2018) and others reported higher quality and enjoyment levels in course evaluations through a chatbot than a traditional web-based service (Wambsganss et al., 2020), adopting an FAQ chatbot in MOOCs seems to be a cost-effective imperative considering the organization's and students' needs.

However, MOOC providers must look beyond novelty and effectiveness, and fully consider the potential social repercussions upon any new technology implementation. As Spector (2020) argued, educational entities need to focus on all students, and not just a small number of outstanding achievers with sufficient accessibility, when they implement any new technologies. Considering the distinctively broad spectrum in a MOOC student population, providing an inclusive learning environment becomes extremely important. Therefore, MOOC providers should make sure any demographic factors do not create an inequitable learning experience upon making any technological changes.

In this regard, we first examine the demographic factors reported in the current literature creating different learning experiences for MOOC students in this study. By doing so, we conceptualize the definition of "inclusive learning" in MOOCs, which may be different from what is generally referred to as "inclusive" in society. Next, we investigate whether an FAQ chatbot in MOOCs that we trained would create any more negatively disparate experiences for the students in courses compared to using a regular FAQ webpage. Considering that this point is crucial before implementing any technology into real-world operation on a full scale to provide an inclusive learning experience through MOOCs, this study delves into students' initial openness (also referred to in this study as a degree of intention to use the assigned interface) as well as the Q&A service's quality, enjoyment, and perceived barriers for both the FAQ chatbot user-student group (CG) and the FAQ webpage user-student group (WG). Upon finding disparate learning experiences for the CG compared to the WG in any of the four aspects mentioned earlier, we investigate four demographic factors and how they influenced the different learning experiences for the CG. Lastly, we determine how feasible it is to provide an inclusive learning experience for the population it serves with the FAQ chatbot and share our lessons learned.

In summary, the research questions (RQs) of this study are:

1. What were the students' unique experiences of using an FAQ chatbot compared to using an FAQ webpage in massive open online courses targeting journalism professionals regarding their intention, quality, enjoyment, and barrier levels when using the chatbot?
2. Did demographic factors including age, gender, region, and native language influence the chatbot group's unique experiences, and if so, how?
3. From an inclusive learning perspective in a massive open online course context, how feasible is it to provide an inclusive learning experience with the chatbot for the majority of the students, and what lessons can we learn from this examination?

2. Literature review

2.1. Inclusiveness (accessibility) in MOOCs

Clarifying key concepts is the first step when designing a technology–enhanced learning environment to make the most of the new technology while avoiding disruption of students' learning (Boyle & Ravenscroft, 2012). Because stakeholders often define "inclusiveness" in different ways, calibrating the definition to the specific context of students' needs, experiences, and preferences is crucial in providing an accessibility-guaranteed learning environment (Coughlan et al., 2019). According to Iniesto et al. (2019), there are four key aspects of MOOC accessibility. The first aspect is about promoting the use of inclusive learning designs, such as universal design, to protect the inherent learning rights of individuals with cognitive disabilities. The second aspect is about technical elements that relate to Web Content Accessibility Guidelines. This regulation details a wide range of recommendations for making web content more accessible, especially for people with disabilities. The third aspect is about the user experience ensuring that activities inside a MOOC are feasible for students through a proper user interface and pedagogical design. The fourth aspect is about an overall quality that concerns multiple components in a course, including staff and student support, curriculum design, course design, delivery, and assessment.

The first two aspects place great emphasis on equity for people with disabilities. However, there is no actual way for MOOC providers to measure their impact of observing these two aspects because the Americans with Disabilities Act of 1990 prevents providers from asking about students' disabilities when they register for courses. Therefore, when they try to enhance the accessibility of their courses, they are usually focused on the latter two aspects that are related to the scrutinization of accessibility from a user experience perspective. Ferguson (2019) mentioned this focus as "accessibility more generally in educational technology" (p. 44), and Coughlan et al. (2019) explained that the accessibility in MOOCs indicates "open access," meaning "the removal of entry requirements and the flexibility provided by support for study at variable levels of intensity, part time, and at a distance" (p. 52). Based on this definition, they argued that teaching thousands of students in a course could be counted as increasing accessibility; however, MOOC providers need to be cautious not to create barriers for some groups of students with their pedagogical and technical design. Therefore, they emphasized that MOOC accessibility research should explore "how a particular technology or service can be designed to enable or exclude particular users" (p. 54). Similarly, Lambert (2020) suggested that future research should address accessibility issues stemming from students' factors, including language differences and technological availabilities in MOOCs while addressing the lack of inclusive education guides in current MOOC studies.

There are two definitions of inclusiveness in MOOC studies in the current literature—one is focused on accessibility for students with disabilities, and the other is more focused on the general student population and concerns barriers that students might experience through certain technology implementations or services. Because of the lack of data regarding students with disabilities in MOOCs, and the main purpose of this study being the examination of the general feasibility of FAQ chatbots, we follow the latter definition of inclusiveness in MOOCs for this study: guaranteeing open access and supporting flexibility for everyone who wishes to learn, taking advantage of the flexible feature of the courses. Following the caution from Coughlan et al. (2019), we acknowledge that chatbots may influence certain students differently, so we investigate the possible different learning experiences in using a chatbot according to their demographic factors that we can detect from the survey results in this study.

2.2. Technologies as social practice and crucial demographic factors in MOOCs

Suchman et al. (1999) asserted that "technologies can be assessed only in their relations to the sites of their production and use" (p. 404) and proposed to regard technologies as social practices. This perspective of technology use is widely accepted in educational technology studies as well. Selwyn (2007) stated that educational technology research should view technology use in a social context so that researchers see its influence relating to the participants and environment, instead of treating technology as a fixed entity. Similarly, Bower (2019) insisted that technology-mediated learning studies need to frame the technology use in a sociopolitical context and recognize how various personal beliefs and practices influence its use. He emphasized the importance of adopting a critical perspective in connecting students' multifaceted learning environment and the substantial influence of technology use based on technology-mediated learning theory. He assumed technology has no ultimate intentional agency, but the context of its use determines the influence.

In MOOC studies, researchers also have reported that students' learning contexts are influential in their learning, and their learning environment is highly dependent on some demographic factors, including age, gender, region, and native language. For example, Kaveri et al. (2015) found that age and gender factors are significant indicators that distinguish a MOOC user from a non-user in India. From their survey dataset, they found out that the likelihood of signing up for MOOCs increased with the respondent's age, and males tended to sign up for MOOCs at twice the rate of females. De Souza and Perry (2021) also highlighted women's much lower course enrollment rate compared to men's, although the proportions of course completion between women and men were equivalent in their study. In addition, Bayeck et al. (2018) indicated that gender influences students' perception of the importance of single-gender grouping in MOOCs, and reported that more female students did not prefer single-gender grouping in courses. Furthermore, they discovered that students' views differed according to their regions. Regarding the interplay of region and native language factors, Littlejohn and Hood (2018) noted the challenge of guaranteeing equal access in countries using multiple official languages with diverse ethnic groups, such as in India and Zimbabwe. They cautioned that people in ethnic minority groups might encounter discrimination and become disadvantaged by unequal access to educational opportunities due to their language factors in a country.

In summary, the current literature reveals several crucial demographic factors that impact students' learning experience in MOOCs. From the "technologies as social practice" perspective, this study aims to investigate whether any of these demographic factors influenced the CG's unique experience compared to the WG's to determine the feasibility of chatbot implementation to provide an inclusive learning environment.

2.3. Chatbot use in an educational context

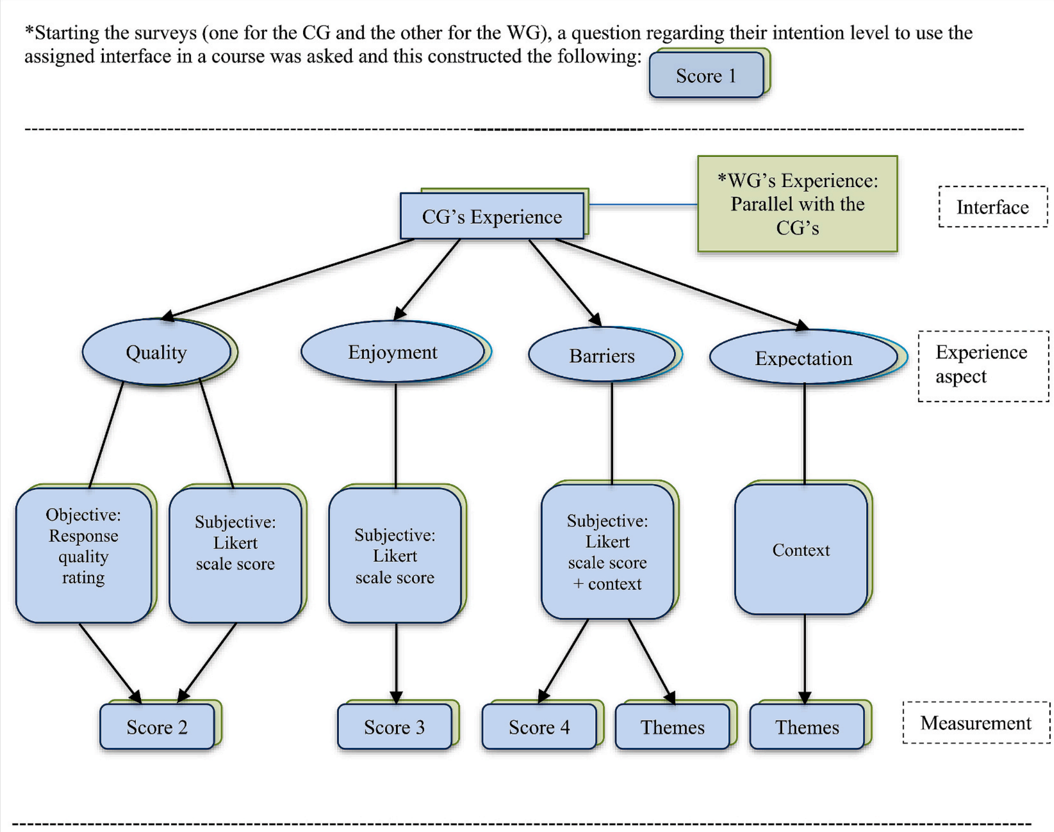
ELIZA—developed by Joseph Weizenbaum in 1966 to simulate a therapist role—was the first chatbot in history utilized in a practical context. Ever since ELIZA appeared, there were continuous research efforts to utilize chatbots with an educational purpose. So far, the major stream of chatbot-related educational research is focused on chatbot development itself rather than its use. For example, Sandoval (2018) designed and implemented a syllabus FAQ chatbot and described the background of its development, explicitly including the organizational needs and the chatbot's performance enhancement processes. However, since there were no case study findings included in the article nor any follow-up study conducted on this chatbot's actual performance in real life, it is hard to know its real usage in an online learning context. Later, there was a research effort to test multiple types of chatbots' performance in enhancing social presence and enforcing knowledge gains (Huang et al., 2019) and exploring various designs of voice-based chatbots (Kloos et al., 2018) in online learning; however, the research emphasis was on the new chatbot design development.

As the use of chatbots in online learning moves to a maturation phase, research focusing more on its use in supporting learning has been conducted. In this regard, Song et al. (2019) compared learning management system data and chatbot log data and analyzed the relationship between students' online learning participation and interaction with the chatbots. Recently, Winkler et al. (2020) investigated the possibility of using a voice-based scaffolding chatbot as an "assistant instructor" in video-based online courses. Even though it was an experimental study in a lab situation, the result indicated the voice-based chatbot's full potential in facilitating meaningful learning in online instructional videos.

In short, the current trend of chatbot research in online learning is moving from development toward examining its uses. The latest research focus is on proving the effectiveness of chatbots in facilitating students' authentic learning by supplementing the teacher's limited guiding role in online learning, especially in an asynchronous mode. However, despite the current advancement of chatbot adoption in online education, there is an obvious lack of studies focused on student-chatbot interaction examining their varying learning experience triggered by a chatbot as new technology adoption.

3. Conceptual framework for data collection

To answer RQ1, we constructed four measurements for students' Q&A experience with FAQs, as Fig. 1 describes. First, we constructed one score for the quality aspect (Score 2) by normalizing the subjective measurement and objective ratings and combining the two measurements with equal weights to generate one final value per each student's response. Second, we constructed a measurement



Analysis

Data	Data pre-processing	Technique	Method
Score 2	1/2: Normalized Text Similarity Score 1/2: Normalized self-reported Q&A interaction quality	Mann-Whitney U test	Compare Score 2_{CG} and Score 2_{WG}
Scores 1, 3, and 4	Self-reported intention (1), Q&A enjoyment (3), barrier (4) levels	Mann-Whitney U test	Compare Score N_{CG} and Score N_{WG}
Demographic factors	Age (3 ranges), gender (3 choices), region (6 categories), native language (2 divisions)	Kruskal-Wallis test (age, gender, region) Mann-Whitney U test (native language)	Test four individual factors for the CG
Barriers/future expectation themes	No data pre-processing conducted	Qualitative open-coding followed by thematic analysis	Open-code the open-ended responses and find themes

Fig. 1. Conceptual framework for data collection with data analysis components.

for each student's levels of intention to use the assigned interface (Score 1), perceived enjoyment (Score 3), and barriers (Score 4) based on their Likert scale responses. Finally, we calculated these four scores per student, compared the scores per group, and then examined the possible unique experience of the CG as a whole.

To answer RQ2, we investigated whether and how any of the demographic factors (age, gender, region, native language) that current literature indicates have influence on students' experience in MOOCs actually triggered students' disparate Q&A service experiences in the CG.

Lastly, to answer RQ3, we reviewed the findings from the previous two RQs and examined their implications in the MOOCs context. The themes that emerged from the survey open-ended responses to the perceived barriers and future expectations questions were analyzed to acquire lessons from this case for possible future FAQ chatbot implementation in MOOCs. Fig. 1 represents how the four scores and two kinds of themes were constructed per group and how those were analyzed using specific tests and methods.

4. Methods

4.1. Research design: Mixed-methods

This mixed-methods study has three phases: development, user testing–oriented survey, and evaluation. In the development phase, we trained a chatbot using FAQ webpage content as a training set on the Dialogflow platform. Once sufficiently trained, the chatbot was deployed on a website enabling the study participants to interact with it. As for the user testing–oriented survey, we randomly divided the total students of two courses—who were identified as active in the courses from the learning management system log data—into two groups, WG and CG, and asked them to participate in an assigned user testing–oriented survey. The survey questions consisted of five categories in the following order: level of intention to use the assigned interface, user-testing, post-user testing, expectations for the future feature, and demographics. The questions regarding demographics and some regarding post-user testing were designed to collect quantitative data, and the others involving post-user testing and expectations for the future features were designed to collect qualitative data. Finally, in the evaluation phase, we examined the differences in students' experiences with the Q&A service according to their groups and their implications by analyzing both quantitative and qualitative data from an inclusive learning perspective in MOOCs that we conceptualized. As a result, the conclusion of this examination determined the feasibility of chatbot implementation and produced lessons for possible chatbot implementation in other MOOC spaces.

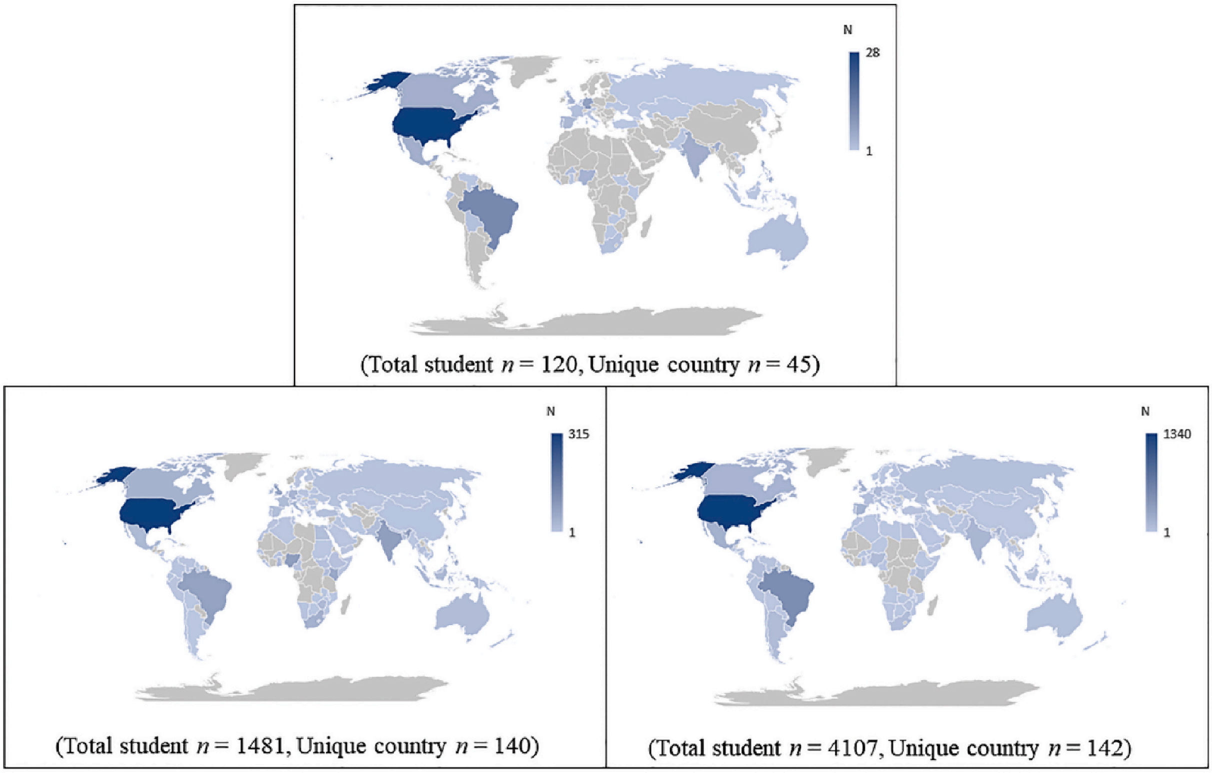


Fig. 2. Global distributions of the survey respondents (top), Course 1 (bottom-left), and Course 2 (bottom-right) students. Grey parts indicate there were no course participants from those areas.

4.2. Setting: Research site, courses, and participants

The research site is located in the southwestern United States and provides MOOCs for journalists' professional development. This site launches 15 courses on average per year, attracting 1000 to 6000 students per course for the past three years. The two courses in this study were delivered in English and lasted for four weeks, from February to May 2021: a health journalism course that attracted 1340 students from 140 unique countries, and a newsletter strategy course that attracted 4107 students from 142 unique countries. Considering that most of the courses this site provides attract an average of 2000 students from 140 unique countries, the student population in these two courses is representative of the site as a whole.

The active students who made progress in the two courses were asked to participate in this study after receiving approval from an institutional review board. These students were randomly assigned to either the CG or the WG and were given the same kinds of survey questions explained in *4.3. Measures, instruments, and analyses* section. A total of 120 students participated in this study, including equal numbers of male and female students ($n = 59$ each), 1 non-binary student, and 1 student who did not comment on gender. Their ages ranged from 18 to 65+, and they hailed from 45 unique countries. To show participant representativeness, Fig. 2 illustrates the survey respondents' worldwide distribution and their distribution across the two courses. The survey participant recruitment rate per group and course is shown in Table 1.

4.3. Measures, instruments, and analyses

Two surveys were constructed to collect the data: one for the CG and the other for the WG. These surveys were created using the Likert scale and open-ended questions. The first question in the surveys asked how much the respondent intended to use the assigned interface. Since developing a chatbot is a time- and resource-consuming process, it would not be prudent for a MOOC provider to invest its limited resources in developing one if the CG's response to this question was significantly low compared to the WG's. We asked this question prior to the start of user testing to avoid any possible prejudice from post-user testing and to gather data about the respondents' baseline preference for their assigned interface. Each response was measured based on a 5-point Likert scale (1: Never intend to use to 5: Very likely to use, with 3 being a neutral statement). This measurement produced Score 1 (i.e., level of intention to use the assigned interface) for each respondent.

Next, we asked both groups three identical user testing questions. We guided the survey participants to visit either a webpage containing the FAQ chatbot or the FAQ webpage and interact with the assigned interface. During their interaction respondents were asked to find answers to three questions: (1) "What are the course activities?" (2) "How can you receive a certificate of completion?" and (3) "When do you meet synchronously in the course?" The participants were asked to fill open-ended text boxes with the answers to these questions, and then the raters (one coauthor and one graduate student collaborator) rated each response according to the rubrics (see Table 2) that were developed in advance. We found that 5% of the total ratings were inconsistent between raters (Cronbach's $\alpha = 0.91$). Therefore, the raters went through a subsequent review session to adjust the inconsistent ratings with unanimous consent. This final rating score per response was normalized and constitutes half of Score 2, representing the objective aspect of Q&A service quality.

Followed by the user testing, post-user testing questions were presented in the surveys. Both groups were asked the same number of questions ($n = 9$) with similar content. The only difference between groups in the post-user testing questions was a result of reflecting on each group's different testing interface, and this did not change the focus of the questions (e.g., "The chatbot made me find the answers easily" for the CG was equivalent to "The webpage made me find the answers easily" for the WG). Each of the three questions were asked to measure the self-reported (subjective) level of Q&A quality, enjoyment, and barriers in using the assigned interface in the post-user testing section of the surveys. Each response was measured based on a 5-point Likert scale (1: Strongly disagree to 5: Strongly agree, with 3 being a neutral statement). Score 1 was constructed by the Likert scale–based response score of one question. The first three questions were normalized and constituted the other half of Score 2. Scores 3 and 4 were constructed by finding the mean of the Likert scale–based response scores for the three questions per student.

To investigate the students' perceived barrier themes, we included one optional open-ended question asking them to elaborate on any challenges or concerns that they encountered while using the assigned interfaces. Half of the total responses to this question were synchronously open-coded by two coders (one coauthor and one graduate collaborator), and the rest were coded independently by the same coders. After finding out that 10% of the total open-codes produced independently were inconsistent from one coder to the other (Cronbach's $\alpha = 0.85$), the coders had an intensive reviewing session together and made changes upon unanimous consent. The final consistent open-codes were used in the thematic analysis. To analyze the final optional open-ended responses regarding future expectations for the assigned interface, the same coders open-coded the responses together because of the manageable total response

Table 1
Survey participant recruitment rate per group and course.

Category	Group	
	CG	WG
Course 1	22	38
Course 2	24	36
Number of survey recruit emails sent	725	725
Recruit success rate	6.34%	10.21%

Table 2
Scoring rubrics for user-testing responses.

Item 1 Find the answer for "What are the course activities?"					
Score	0	1	2	3	4
Included keyword number	None	One	Two	Three	Four
Keyword	(a) quiz, (b) reading, (c) forum or discussion, (d) video or lecture				
Item 2 Find the answer for "What do you do to receive a certificate of completion?"					
Score	0	1	2	3	4
Included keyword number	None	One	Two	Three	Four
Keyword	(a) above 70 quiz score, (b) forum or discussion, (c) apply, (d) fee				
Item 3 Find the answer for "When do you meet synchronously in the course?"					
Score	0	2		4	
Included keyword	None of the keywords	Course flexibility-related terms		Never	

number ($n = 17$). These responses were categorized under two labels—easy or hard to fulfill—and utilized in assessing the chatbot's general feasibility for MOOC spaces.

Regarding the Likert scale–based post-user testing question items, we followed the instruments from Wambsganss et al. (2020) with Likert scale questions but adapted the questions to our study context. That is, we measured the participants' perceived Q&A service quality with the FAQ chatbot and the FAQ webpage by using one subjective measurement (self-reported Q&A interaction quality by a participant) and one objective measurement (raters' ratings of user testing answers). To measure the self-reported Q&A service quality, we asked students to respond to the following: "The chatbot/webpage helped me to find the answers easily," "I would prefer using the chatbot/webpage as an interface to find answers to my questions about courses compared to using others," and "The chatbot/webpage was useful in finding answers to the questions." We also measured the perceived level of enjoyment by asking the students to respond to the following, per Kim et al. (2019) and Wambsganss et al. (2020): "It was fun to use the chatbot/webpage," "I am satisfied with the chatbot/webpage," and "The process of using the chatbot/webpage was pleasant." Furthermore, we measured the perceived level of barriers by asking the students to respond to these statements: "I had trouble using the chatbot/webpage," "I was concerned about my personal information while using the chatbot/webpage," and "Using a chatbot/webpage seemed to involve high risk" (Kasilingam, 2020). Reflecting on the literature of public misconception about artificial intelligence technology (De Saint Laurent, 2018), we asked perceived privacy risk–related questions with a question about the difficulty level in use when measuring the perceived level of barriers. As De Saint Laurent (2018), Searle (1980), and Turing (2009) indicated, people often assume the same process happened when coming across the same results from humans and machines. Hence, there was a possibility that a student might have mistakenly perceived the FAQ chatbot in this study as an assistant chatbot connected to a personal information database instead of being restricted to FAQs only. In this case, we acknowledged that the student might have experienced unnecessarily increased barriers in using the chatbot, so we decided to include a perceived privacy risk aspect to the list of potential barriers and investigated its influence.

The last part of the surveys consisted of demographics and expectation questions for future features. These were largely based on the instruments used in previous studies with a similar student population (Liu et al., 2019). The survey participants' age (1: 18–44; 2: 45–64; 3: 65+; three age ranges from Howden & Meyer, 2011), gender (1: male; 2: female; 3: non-binary; 4: prefer not to say), region (1: Africa; 2: Asia; 3: Europe; 4: North America; 5: Oceania; 6: South America; six categories from National Geographic, n.d.), and native language (English/non-English) were collected and one optional open-ended question was asked to elaborate their future expectations for the assigned interface.

5. Results

Upon confirming that the data were not normally distributed, we employed nonparametric tests to conduct inferential statistical analyses. We used the Mann-Whitney *U*U test to examine significant mean differences between the groups. In this way, we compared the mean ranks of the four scores per group and investigated the possible unique experiences of CG, pursuing RQ1. Upon finding the unique experiences of CG regarding the four scores, we examined the influence of students' demographic factors in causing those unique experiences in the CG by operating Kruskal-Wallis tests—for age, gender, and region factors—and the Mann-Whitney *U* test—for the native language factor, pursuing RQ2. Finally, we determined the feasibility of the chatbot implementation based on the findings from the first two RQs while investigating RQ3 and elaborated on the lessons learned from this examination based on the themes that emerged from the open-ended responses. We used a significance level of 0.05 for all statistical tests.

5.1. RQ1: Unique experiences of using an FAQ chatbot compared to using an FAQ webpage

According to the Mann-Whitney *U*U test results, Score 1—the level of intention to use the assigned interface—was significantly lower in the CG ($Mdn = 3.00$, $n = 46$) than in the WG ($Mdn = 4.00$, $n = 74$), $U = 1225.50$, $z = -2.73$, $p = .006$, with a small effect size $r = 0.25$. On the contrary, Score 4—the perceived barrier level—was significantly greater in the CG ($Mdn = 2.50$) compared to the WG ($Mdn = 1.67$), $U = 863.00$, $z = -4.60$, $p < .001$, with a medium effect size $r = 0.45$. However, the differences between groups in Scores 2 (Q&A quality) and 3 (enjoyment) turned out to be statistically insignificant, especially the statistical difference in Score 2 between groups ($p = .963$). The statistical difference in Score 3 between groups was close to the significance level; however, it still had no statistical significance ($p = .067$). To show the significant differences in the distributions of Score 1 and Score 4 between groups, we created comparative frequency histograms in Fig. 3.

5.2. RQ2: Demographic factors' influence on the chatbot group's unique experiences

Considering the CG's unique experience regarding Score 1 and Score 4 compared to the WG's, we examined the influence of students' demographic factors on these two scores among the CG. All of the demographic factors we examined in this study—age ($p = .827$), gender ($p = .970$), region ($p = .463$), and native language ($p = .951$)—were statistically insignificant in creating a different Score 1 (intention level) in the CG.

Regarding Score 4, Kruskal-Wallis and Mann-Whitney *UU* test results revealed that region and native language factors were statistically significant in creating the differences in the CG. A Kruskal-Wallis test revealed a statistically significant difference in the perceived barrier levels across the six regional conditions (Africa, Asia, Europe, North America, Oceania, and South America), $\chi 2$ (5, $N = 46$) = 12.47, $p = .029$. The barrier level was the highest in Asia ($Mdn = 3.33$, $n = 11$) and Oceania conditions ($Mdn = 3.33$, $n = 1$), and the ones in Europe ($Mdn = 2.00$, $n = 7$) and North America conditions ($Mdn = 2.00$, $n = 10$) were the lowest, with the ones in Africa ($Mdn = 2.67$, $n = 7$) and South America ($Mdn = 2.17$, $n = 10$) in between. To summarize and compare the disparate perceived barrier levels per region, we plotted them in Fig. 4. In addition, a Mann-Whitney *UU* test revealed a statistically significant higher level of the perceived barrier among non-native English users ($Mdn = 3.00$, $n = 29$) compared to native English users in the CG ($Mdn = 2.00$, $n = 17$), $U = 156.50$, $z = -2.06$, $p = .039$, with a medium effect size $r = 0.30$. Fig. 5 shows the statistically significant differences in the distributions of Score 4 (perceived barrier level) between the binary language groups in comparative frequency histograms. However, age ($p = .348$) and gender ($p = .208$) factors turned out to be insignificant in creating different perceived barrier levels in the CG.

5.3. RQ3: Feasibility of the chatbot from an inclusive learning perspective and the lessons learned

Pursuing RQ1, we found out that the CG tended to possess a lower intention level to use the FAQ chatbot compared to the WG's intention level to use the FAQ webpage. Regarding the perceived barrier levels, the CG tended to possess a higher perceived barrier level with the FAQ chatbot compared to its counterpart's perceived barrier level with the FAQ webpage. When we investigated the demographic factors' influence in creating this disparate experience in the CG guided by RQ2, we found that none of the demographic factors had significance in creating different intention levels to use the FAQ chatbot, whereas region and native language factors had statistical significance in creating higher perceived barrier levels in using the FAQ chatbot for certain student groups. In particular, the finding showing non-native English users' aggravated challenge level with the chatbot, even though their levels of Q&A service quality and perceived enjoyment were statistically equivalent to their native English classmates' levels, indicates that there are unknown reasons for them to feel more challenges (see Fig. 6). Therefore, we determined that it is not feasible to provide an inclusive learning experience for the majority of students with the current chatbot mainly due to the higher perceived barrier levels in using the chatbot for the students in Asia and Oceania (26.10% of the total CG), and for the non-native English users (63.04% of the total CG). We also considered the lower intention to use the FAQ chatbot compared to the intention to use the FAQ webpage in this assessment.

However, the qualitative data we gathered about the survey participants' perceived barrier elaborations and future expectations in using the assigned interfaces guided us to see what and how to improve both interfaces as well as the overall future direction for Q&A services in MOOCs. The lessons we learned from the participants' open-ended responses are as follows.

5.3.1. Low quality of Q&A content for both interfaces

We received 37 unique responses from the survey participants to the open-ended question about their perceived barrier level elaboration. These unique responses generated 48 open-codes ($n = 30$ from the CG; $n = 18$ from the WG) that we categorized into seven different themes, represented in Fig. 7. The most dominant theme regarding barriers for both the CG and the WG was the limited number of questions successfully answered. Almost half of the total responses from the CG mentioned an unsatisfactory Q&A experience via the chatbot as a barrier (e.g., "I asked the chatbot about the courses you have. But the chatbot didn't understand my

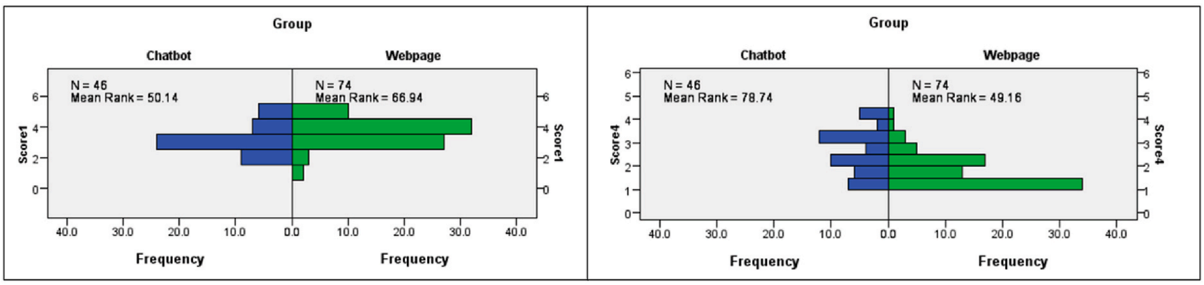


Fig. 3. Distribution of Score 1—level of intention to use the assigned interface (left)—and Score 4—level of perceived barriers in using the assigned interface (right)—between chatbot and webpage groups. Score 1(intention level) and Score 4 (barrier level) ranges from 1.00 (low) to 5.00 (high). Frequency shows the number of students per score range. As Score 1 is constructed from a single question item, each score range directly corresponds to each of the five Likert scale response values. On the contrary, Score 4 is constructed as a mean value of three question items, so the results are mostly decimal numbers. Therefore, more granular score ranges (1–1.5, 1.5–2, 2–2.5, etc.) are utilized for Score 4.

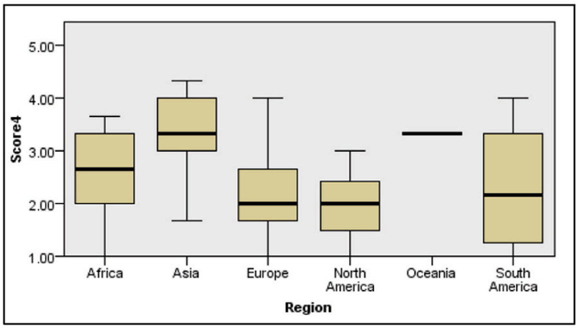


Fig. 4. Distribution of Score 4—level of perceived barriers in using the chatbot—per six regional conditions. Score 4 (barrier level) ranges from 1.00 (low) to 5.00 (high).

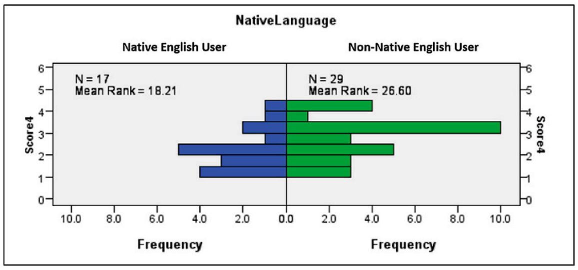


Fig. 5. Distribution of Score 4—level of perceived barriers in using the chatbot—between binary language groups. Score 4 (barrier level) ranges from 1.00 (low) to 5.00 (high). Frequency shows the number of students per score range, and the number of bars for this score was determined by the dataset characteristics in the same manner of creating Fig. 3.

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Score1 is the same across categories of NativeLanguage.	Independent-Samples Mann-Whitney U Test	.951	Retain the null hypothesis.
2	The distribution of Score2 is the same across categories of NativeLanguage.	Independent-Samples Mann-Whitney U Test	.368	Retain the null hypothesis.
3	The distribution of Score3 is the same across categories of NativeLanguage.	Independent-Samples Mann-Whitney U Test	.909	Retain the null hypothesis.
4	The distribution of Score4 is the same across categories of NativeLanguage.	Independent-Samples Mann-Whitney U Test	.039	Reject the null hypothesis.

Fig. 6. Hypothesis test summary of native language factors with Scores 1, 2, 3, and 4. Score 1 is the intention to use the chatbot level, Score 2 is the Q&A quality level, Score 3 is the enjoyment level, and Score 4 is the perceived barrier level.

question. It seems the chatbot still needs to learn."). The same portion of WG respondents also expressed their complaints about the fact that they could not find the answer to the questions they had on the webpage (e.g., "Many questions do not apply to mine."). Considering that the Q&A webpage content was the training set for the chatbot, meaning the chatbot was limited to providing only information contained on the webpage, this finding suggests that the problem lies with the Q&A content itself, which needs to be further amended to satisfy the students' demands.

5.3.2. Stronger necessity of chatbot interface with refined Q&A content

The survey participants' future expectation responses ($n = 22$) revealed that students want to receive a simple answer consisting of one or two sentences per finely assorted question. For example, some students wanted to know the amount of certificate fee only—

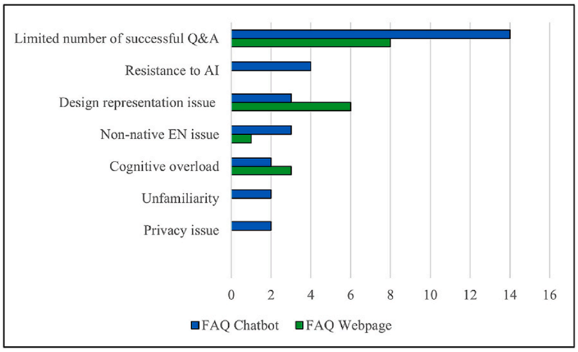


Fig. 7. Perceived barrier themes per interface. The x-axis values represent the numbers of open-codes that emerged per theme.

$30.00—instead of receiving verbose information about all the general requirements for applying for a certificate, which is how fee information on the FAQ webpage content is organized. They also wanted to see additional questions added to the FAQ for both the chatbot and webpage, such as how to check a course's total enrollment and future course information. These demands are easy to meet for the MOOC provider if they make more entries of focused questions with succinct answers. However, this approach poses a possibility of increasing students' cognitive overload for FAQ webpage users who responded that there were too many questions and it was too difficult to locate what they were looking for and identified this problem as a barrier in using the webpage (i.e., cognitive overload), not to mention some WG respondents complaining about the fatigue stemming from the information overload by saying "text heavy, had to seek out information, didn't like it at all." For the CG, the cognitive overload theme included unpleasant experiences coming from overly long responses to their questions caused by the same Q&A content as the webpage (e.g., "The biggest difficulty was a long response. No difference from the FAQ page. I had to read too much. I would rather want short answers."). Therefore, the potential solution of increasing the volume of refined Q&A content may be more suited to a chatbot interface rather than a webpage interface. To clarify the cognitive overload theme, we would like to mention the definition of the design representation issue theme here. For the CG, this theme included responses about the chatbot menu design (e.g., "Send a message button—little triangle at the end of the message—is too light. Please make it darker."). For the WG, this theme was about the visual components of the webpage, including font size, color, and spacing of the text, which did not create cognitive overload according to the students. Regardless of the interface, all of the responses regarding excessive item numbers of Q&As or complaints about text heaviness fell into the cognitive overload theme.

5.3.3. Resistance to using AI-based technology and other negative barrier themes from the CG

Some CG participants expressed their strong resistance to using any AI-based technologies, including chatbots (e.g., "I believe it's extremely strange to have a bot trying to interact as if it was human. A bot is not a person and should not act or be treated like that … some people don't like to 'talk' to bots.") This "resistance to AI" theme was the second largest barrier theme that emerged from the CG, followed by the "design representation issue" including cultural differences in the chatbot's interface visual design components—menu location and color—and "non-native English user issue," which includes responses such as "Since I am not a native English speaker, I found it a bit hard to communicate with a chatbot programmed to answer in specific ways to specific questions." One noteworthy finding about this barrier theme is that we received very similar responses from the WG as well (e.g., "I would be concerned that an international crowd not accustomed to US online classes might not know the terms synchronous and asynchronous."). This is an understandable result considering that both interfaces produced the same answers to questions; however, the heavier emphasis on this theme in the CG's survey result (10% of the total CG) indicates non-native English users' negative experience with the chatbot compared to the WG (5.56% of the total WG). Additionally, "unfamiliarity" (e.g., "This was my first time of using chatbot in my life so that it was not easy.") and "privacy issue" (e.g., "asking personal information should not be mandatory unless really necessary.") were the unique barrier themes for the CG.

6. Discussion

This study examines students' unique experiences of using an FAQ chatbot as compared to an FAQ webpage, and investigates whether and how demographic factors influenced the chatbot users' unique Q&A service experience. In this section, we discuss the implications of this study's findings, followed by limitations and future research directions.

6.1. Intention and perceived barrier levels

Even though this study reveals statistically significant differences in the levels of intention and perceived barriers among the CG compared to the WG, it also confirms that the CG's lower intention to use the chatbot is not associated with any demographic factors of the participants we investigated—age, gender, region, and native language. Considering that the intention level question is presented at the beginning of the survey before the CG engages with the chatbot, the lower intention level to use the chatbot originates from the participants' personal biases, not from the chatbot itself. Therefore, we may assume that other environmental factors contributed to the reluctance to use a chatbot. The lower survey participant recruitment rate of the CG compared to the WG (see Table 2), where the WG's recruitment success rate was 1.61 times higher than the CG's despite recruitment emails being sent to the same number of eligible students in the courses based on the random group assignments, also indicates there are some unknown barrier factors at work for the CG. Additionally, we identified unique barrier themes from the CG, including "resistance to AI," "unfamiliarity," and "privacy issues," which hint at what the barrier factors might be.

Multiple previous studies suggest that the main obstacles in replacing chatbots with traditional service interfaces are rooted in the nature of AI, which demands time to enhance its performance through multiple sequences of trial and error based on training sets (Di Gaetano & Diliberto, 2018; Robino, 2018; Valtolina et al., 2020). However, some studies also reported the reluctance toward chatbot use originated by users (Jenkins et al., 2007) and explored possible user factors in creating the resistance to use a chatbot, such as privacy concerns (Valtolina et al., 2020). In the current study, the quantitative (i.e., lower intention level to use the chatbot) and qualitative findings of certain CG's unique barrier themes (i.e., resistance to AI, unfamiliarity) align with the findings of Jenkins et al. (2007) in terms of confirming users' resistance to using new technology such as chatbots. Moreover, the CG's unique perceived barrier theme about privacy issues indicates users' concerns for privacy, which corresponds to previous study findings with chatbots (Rhee & Choi, 2020; Shumanov & Johnson, 2021). The study contributes to the current literature by confirming that the intention to use a chatbot is lower than traditional interfaces in MOOC contexts, and this result is most likely coming from user factors possibly including concerns for privacy and general unfamiliarity with the technology—themes that are rarely explored in current MOOC studies.

6.2. Quality and enjoyment

This study contributes to the field by reporting findings that indicate the levels of Q&A service quality and enjoyment in using the chatbot are not statistically different from using a traditional FAQ webpage; this differs from other studies that reported enhanced quality and enjoyment through chatbot interfaces compared to traditional web-based surveys in course evaluation (Kim et al., 2019; Wambsganss et al., 2020). The current study's inconsistent finding indicates that more domain-specific and users' environmental factor–related research is necessary to further our insights on how to implement new technologies while properly considering their contextual uses.

6.3. Multifaceted approach to inclusive learning with new technologies in MOOCs

In a nutshell, we determined that it is infeasible to guarantee an inclusive learning environment for the majority of students with a chatbot in MOOCs largely due to the statistically significant higher perceived barrier level for some groups of students by region—in Asia and Oceania—and by native language—non-native English users. Because the current MOOC study literature suggests the importance of a fair and accessible learning experience regardless of students' demographic factors, which include their regions and native languages, an exacerbated learning experience in the CG with a higher barrier level for a certain group of students is a definite deal-breaker. However, our qualitative analysis suggests that the overall quality of Q&A content for both interfaces is equally unsatisfactory for the students, so the higher perceived barrier level for the CG might be due to user factors including privacy concerns and unfamiliarity with chatbots. Other than these two factors, some human and computer interaction studies suggest that expectancy violation effects (Burgoon & Walther, 1990; Cappella & Greene, 1982; Kahneman & Miller, 1986) could play a role in increasing users' negative experience with a chatbot by setting a high expectation that the chatbot should perform as a human-like being (Go & Sundar, 2019); this could be a future research direction to investigate the higher perceived barrier level for the CG. Therefore, implementing a new technology such as a chatbot for an inclusive learning experience in MOOCs requires an understanding of the relationship between individuals' psychological, attitudinal, and behavioral responses.

Given the domain of chatbot implementation, it is also necessary to give the contextual meaning of inclusiveness in MOOCs an immense amount of attention. To investigate demographic factors' influence on the CG's unique experience, the current study examined age, gender, region, and native language factors, which existing literature has reported as critical factors in creating disparate learning experiences in MOOCs. As a result, region and native language conditions have been confirmed as significant in creating the different levels of perceived barriers in the CG. This finding confirms some literature reporting the importance of region and native language factors in MOOC learning (e.g., Littlejohn & Hood, 2018); however, it contradicts other studies reporting the influence of age and gender factors. Therefore, additional research is required to resolve this inconsistency.

Lastly but equally as important as the two previous aspects—human and computer interaction and MOOC context—the aspect of technologies existing within a social context needs to be considered to provide an inclusive learning experience with new technology such as a chatbot in MOOCs. As Smutny and Schreiberova (2020) indicated, educational chatbot applications are still at a nascent stage, and there is much research that needs to be conducted to better support students' learning on both macro and micro levels. On the macro level, the use of chatbots should be determined after a series of thorough selection and review processes to provide a sociopolitically fair and accessible learning space for all students. To support students' learning at the micro level, there is an active current line of research investigating making chatbots more personalized to better support individual students' learning by matching users' personalities with coinciding chatbot personalities (Rhee & Choi, 2020; Shumanov & Johnson, 2021). As the authors of these studies suggest, the social impact of making chatbots more personalized needs to be carefully considered, and any ethical considerations, including how to collect, manage, and use personal data for chatbot use, need to be addressed from a critical perspective focusing on its influence on the users and their environment based on robust social agreement.

6.4. Limitations and future research

One main limitation of this study is the number of CG participants ($n = 46$). Although we recruited CG participants from the same population as the WG ($n = 725$) under the same condition, a much smaller number of students joined this study compared to the WG ($n = 74$). This diminished recruitment rate implies that some barriers unique to participating in the CG are present, giving us valuable insights as covered in the discussion section. However, it would have been better if we had more responses to validate the implications more robustly. Additionally, the chatbot log data revealed that the CG participants' attitudes toward the chatbot were quite varied in terms of perceiving it as human, non-human, or something in between. Considering the ample literature reporting the importance of attitude toward an agent in creating positive behavioral intentions, including continuous use of the agent (Go & Sundar, 2019; Kent & Taylor, 2002; Seltzer & Mitrook, 2007), connecting the findings from this study with the chatbot log data analysis could provide insights for improving the chatbot to support students' personalized learning. Furthermore, we propose to divide students' perceived barriers aspect into three categories—visual design (e.g., color, location), difficulty in use (e.g., language, mental effort, response mechanism), and privacy-themed (e.g., personal data, security system)—in future studies. We expect that this research direction would provide more useful chatbot response design points, enhancing usability in developing FAQ chatbots that are different from assistant-type chatbots.

7. Conclusion

Automated technologies are becoming attractive solutions in MOOCs to ease instructors' work in answering repetitive questions from a massive number of students so that they can spend more time on students' cognitive scaffolding (Poquet et al., 2020). Considering that MOOC students often report they are not sufficiently guided by instructors or course providers in how to navigate their course activities (Julia & Marco, 2021), providing an FAQ chatbot could improve students' learning experience by adding one more tool for student support. However, it is important to thoroughly examine whether this new technology could benefit the majority of the students from an inclusive learning perspective before adoption, mainly due to the unique contextual meaning of inclusiveness in MOOCs.

Statistically speaking, the current study shows that users have a lower intention to use an FAQ chatbot compared to an FAQ webpage, and students experience a higher magnitude of barriers when using the chatbot in MOOC spaces. Nevertheless, the levels of service quality and enjoyment are the same across the two interfaces when the Q&A content for both interfaces is the same. When using an FAQ chatbot, region and language factors are influential in creating a disparate learning experience among the students. However, our qualitative analyses suggest the prevalence of additional factors behind the scenes that make students reluctant to use a chatbot in the first place. This finding indicates that a multifaceted approach needs to be taken when introducing new technology such as chatbots, which show potential in providing customized responses that promote inclusive learning when trained properly based on high-quality refined training sets reflecting the students' needs. Taken together, this study suggests what aspects to consider when adopting new technologies in MOOCs to provide an inclusive learning experience and underscores the need for more active research in chatbot use to serve diverse student needs in MOOCs.

Credit author statement

Songhee Han: Conceptualization, Methodology, Investigation, Data curation, Software, Formal analysis, Visualization, Validation, Writing original draft, Writing – review & editing. Min Kyung Lee: Methodology, Investigation, Formal analysis, Resources, Supervision, Writing – review & editing.

References

Bayeck, R. Y., Hristova, A., Jablokow, K. W., & Bonafini, F. (2018). Exploring the relevance of single-gender group formation: What we learn from a massive open online course (MOOC). *British Journal of Educational Technology, 49*(1), 88–100. https://doi.org/10.1111/bjet.12528

Bower, M. (2019). Technology-mediated learning theory. *British Journal of Educational Technology, 50*(3), 1035–1048. https://doi.org/10.1111/bjet.12771

Boyle, T., & Ravenscroft, A. (2012). Context and deep learning design. *Computers & Education, 59*(4), 1224–1233. https://doi.org/10.1016/j.compedu.2012.06.007

Burgoon, J. K., & Walther, J. B. (1990). Nonverbal expectancies and the evaluative consequences of violations. *Human Communication Research, 17*(2), 232–265. https://doi.org/10.1111/j.1468-2958.1990.tb00232.x

Cappella, J. N., & Greene, J. O. (1982). A discrepancy-arousal explanation of mutual influence in expressive behavior for adult and infant-adult interaction. *Communication Monographs, 49*(2), 89–114. https://doi.org/10.1080/03637758209376074

Chatterjee, A., Gupta, M., & Agrawal, P. (2020, January). FAQaugmenter: Suggesting questions for enterprise FAQ pages. In *Proceedings of the 13th international conference on web search and data mining* (pp. 829–832). https://doi.org/10.1145/3336191.3371862

Coughlan, T., Lister, K., Seale, J., Scanlon, E., & Weller, M. (2019). Accessible inclusive learning: Foundations. In R. Ferguson, A. Jones, & E. Scanlon (Eds.), *Educational visions: Lessons from 40 years of innovation* (pp. 51–73). London: Ubiquity Press. https://doi.org/10.5334/bcg.d.

De Saint Laurent, C. (2018). Defence of machine learning: Debunking the myths of artificial intelligence. *Europe's Journal of Psychology, 14*(4), 734–747. https://doi.org/10.5964/ejop.v14i4.1823

De Souza, N. S., & Perry, G. T. (2021). Women's participation in MOOCs in the IT area. *Computers & Education, 173*, Article 104270. https://doi.org/10.1016/j.compedu.2021.104270

Di Gaetano, S., & Diliberto, P. (2018). Chatbots and conversational interfaces: Three domains of use. In *Fifth international workshop on cultures of participation in the digital age* (pp. 62–70). http://ceur-ws.org/Vol-2101/paper8.pdf.

Ferguson, R. (2019). Teaching and learning at scale: Futures. In R. Ferguson, A. Jones, & E. Scanlon (Eds.), *Educational visions: Lessons from 40 years of innovation* (pp. 33–50). London: Ubiquity Press. https://doi.org/10.5334/bcg.c.

Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior, 97*, 304–316. https://doi.org/10.1016/j.chb.2019.01.020

Howden, L. M., & Meyer, J. A. (2011). *Age and sex composition, 2010*. US Department of Commerce, Economics and Statistics Administration, US Census Bureau.

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), 6. https://doi.org/10.18178/ijmerr.8.5.813-818

Iniesto, F., McAndrew, P., Minocha, S., & Coughlan, T. (2019, September). Auditing the accessibility of MOOCs: A four-component approach. In *European conference on technology enhanced learning* (pp. 650–654). Cham: Springer. https://doi.org/10.1007/978-3-030-29736-7_59.

Jenkins, M. C., Churchill, R., Cox, S., & Smith, D. (2007, July). Analysis of user interaction with service oriented chatbot systems. In *International conference on human-computer interaction* (pp. 76–83). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-73110-8_9.

Julia, K., & Marco, K. (2021). Educational scalability in MOOCs: Analysing instructional designs to find best practices. *Computers & Education, 161*, Article 104054. https://doi.org/10.1016/j.compedu.2020.104054

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*(2), 136–153. https://doi.org/10.1037/0033-295X.93.2.136

Kasilingam, D. L. (2020). Understanding the attitude and intention to use smartphone chatbots for shopping. *Technology in Society, 62*. https://doi.org/10.1016/j.techsoc.2020.101280

Kaveri, A., Gunasekar, S., Gupta, D., & Pratap, M. (2015, October). Decoding the Indian MOOC learner. In *2015 IEEE 3RD international conference on MOOCS, innovation and technology in education (MITE)* (pp. 182–187). IEEE. https://doi.org/10.1109/MITE.2015.7375311.

Kent, M. L., & Taylor, M. (2002). Toward a dialogic theory of public relations. *Public Relations Review, 28*(1), 21–37. https://doi.org/10.1016/S0363-8111(02)00108-X

Kim, S., Lee, J., & Gweon, G. (2019, May). Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–12). https://doi.org/10.1145/3290605.3300316

Kloos, C. D., Catálan, C., Muñoz-Merino, P. J., & Alario-Hoyos, C. (2018, September). Design of a conversational agent as an educational tool. In *2018 learning with MOOCS (LWMOOCS)* (pp. 27–30). IEEE. https://doi.org/10.1109/LWMOOCS.2018.8534591.

Lambert, S. R. (2020). Do MOOCs contribute to student equity and social inclusion? A systematic review 2014–18. *Computers & Education, 145*, Article 103693. https://doi.org/10.1016/j.compedu.2019.103693

Littlejohn, A., & Hood, N. (2018). *Reconceptualising learning in the digital age: The (Un)democratising of learning*. Singapore: Springer. https://doi.org/10.1007/978-981-10-8893-3

Liu, M., Zou, W., Shi, Y., Pan, Z., & Li, C. (2019). What do participants think of today's MOOCs: An updated look at the benefits and challenges of MOOCs designed for working professionals. *Journal of Computing in Higher Education, 32*, 1–23. https://doi.org/10.1007/s12528-019-09234-x

(n.d.) National Geographic. Continents https://www.nationalgeographic.org/encyclopedia/Continent/.

Poquet, O., Jovanovic, J., & Dawson, S. (2020). Differences in forum communication of residents and visitors in MOOCS. *Computers & Education, 156*, Article 103937. https://doi.org/10.1016/j.compedu.2020.103937

Rhee, C. E., & Choi, J. (2020). Effects of personalization and social role in voice shopping: An experimental study on product recommendation by a conversational voice agent. *Computers in Human Behavior, 109*, Article 106359. https://doi.org/10.1016/j.chb.2020.106359

Robino, G. (2018, January 24). *Is 2018 the year of voice interfaces?* ConvComp.it. https://convcomp.it/il-2018-%C3%A8-lanno-delle-voice-user-interfaces-3134ca939497.

Rubin, V. L., Chen, Y., & Thorimbert, L. M. (2010). *Artificially intelligent conversational agents in libraries*. Library Hi Tech. https://doi.org/10.1108/07378831011096196

Sandoval, Z. V. (2018). Design and implementation of a chatbot in online higher education settings. *Issues in Information Systems, 19*(4). https://doi.org/10.48009/4_iis_2018_44–52

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(3), 417–424. https://doi.org/10.1017/S0140525X00005756

Seltzer, T., & Mitrook, M. A. (2007). The dialogic potential of weblogs in relationship building. *Public Relations Review, 33*(2), 227–229. https://doi.org/10.1016/j.pubrev.2007.02.011

Selwyn, N. (2007). e-Learning or she-learning? Exploring students' gendered perceptions of education technology. *British Journal of Educational Technology, 38*(4), 744–746. https://doi.org/10.1111/j.1467-8535.2006.00659.x

Shaw, A. (2012). Using chatbots to easily create interactive and intelligent FAQ webpages. *Journal of Applied Global Research, 5*(15), 10–15.

Shawar, B. A., & Atwell, E. (2007). Chatbots: Are they really useful? *LDV Forum, 22*(1), 29–49. http://www.jlcl.org/2007_Heft1/Bayan_Abu-Shawar_and_Eric_Atwell.pdf.

Shumanov, M., & Johnson, L. (2021). Making conversations with chatbots more personalized. *Computers in Human Behavior, 117*, Article 106627. https://doi.org/10.1016/j.chb.2020.106627

Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the facebook messenger. *Computers & Education, 151*, Article 103862. https://doi.org/10.1016/j.compedu.2020.103862

Song, D., Rice, M., & Oh, E. Y. (2019). Participation in online courses and interaction with a virtual agent. *International Review of Research in Open and Distance Learning, 20*(1). https://doi.org/10.19173/irrodl.v20i1.3998

Spector, J. M. (2020). Remarks on progress in educational technology. *Educational Technology Research & Development, 68*, 833–836. https://doi.org/10.1007/s11423-020-09736-x

Suchman, L., Blomberg, J., Orr, J. E., & Trigg, R. (1999). Reconstructing technologies as social practice. *American Behavioral Scientist, 43*(3), 392–408.

Turing, A. M. (2009). Computing machinery and intelligence. In R. Epstein, G. Roberts, & G. Beber (Eds.), *Parsing the turing test* (pp. 23–65). Dordrecht: Springer. https://doi.org/10.1007/978-1-4020-6710-5_3.

Valtolina, S., Barricelli, B. R., & Di Gaetano, S. (2020). Communicability of traditional interfaces VS chatbots in healthcare and smart home domains. *Behaviour & Information Technology, 39*(1), 108–132. https://doi.org/10.1080/0144929X.2019.1637025

Wambsganss, T., Winkler, R., Söllner, M., & Leimeister, J. M. (2020, April). A conversational agent to improve response quality in course evaluations. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems* (pp. 1–9). https://doi.org/10.1145/3334480.3382805

Winkler, R., Hobert, S., Salovaara, A., Söllner, M., & Leimeister, J. M. (2020, April). Sara, the lecturer: Improving learning in online education with a scaffolding-based conversational agent. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–14). https://doi.org/10.1145/3313831.3376781