



UNIVERSITY OF LEEDS

This is a repository copy of *University Student Surveys Using Chatbots: Artificial Intelligence Conversational Agents*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/171076/>

Version: Accepted Version

Proceedings Paper:

Abbas, N orcid.org/0000-0002-1434-6497, Pickard, T, Atwell, E orcid.org/0000-0001-9395-3764 et al. (1 more author) (2021) *University Student Surveys Using Chatbots: Artificial Intelligence Conversational Agents*. In: *Learning and Collaboration Technologies: Games and Virtual Environments for Learning*. HCI International 2021, 24-29 Jul 2021, Washington DC, USA. *Lecture Notes in Computer Science*, 12785 . Springer, Cham , pp. 155-169. ISBN 978-3-030-77942-9

https://doi.org/10.1007/978-3-030-77943-6_10

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

University student surveys using chatbots: Artificial Intelligence conversational agents

Noorhan Abbas¹ [0000-0002-1434-6497], Thomas Pickard¹ [0000-0002-2523-2364],

Eric Atwell¹ [0000-0001-9395-3764] and Aisha Walker² [0000-0001-9773-6134]

¹ School of Computing, University of Leeds, Leeds, UK

² School of Education, University of Leeds, Leeds, UK

Correspondence: e.s.atwell@leeds.ac.uk

Abstract: Predefined web surveys are often used to collect course evaluations from students in higher education institutions. These institutions use the evaluations to adjust their courses' pedagogical standards and lecture style to cope with an increasingly uncertain and complex world. Many limitations to using web surveys have been reported such as low response rates and low-quality responses to open questions. To overcome these limitations, artificial intelligence conversational agents (CAs) or 'chatbots' are used to play the interviewer role, facilitating the enhancement of the quality of responses. This is accomplished by mimicking human-human conversations; by asking questions in a friendly, casual way and pursuing high-quality responses. This study aims to explore the opportunities and the obstacles of using CAs in collecting course evaluations in three European universities (UK, Spain and Croatia) and one Centre of excellence in Cyprus. The transcripts collected have been analyzed using statistical data analysis methods and qualitative data analysis techniques. Our findings reveal that the use of CAs in collecting course feedback from students has a positive impact on response quality and can boost students' enjoyment levels. Furthermore, gender differences and student age have been identified as important factors that can influence the depth of the conversation with the CA.

Keywords: Chatbot; Conversational Agent; Online Course Evaluation; Pedagogical Conversational Agents; Student Enjoyment.

1 Introduction

Nowadays, web or online surveys are often used to collect course feedback/evaluations from students in higher education institutions. This feedback enables these institutions to adjust their courses' pedagogical standards and lecture style to cope with an increasingly uncertain and complex world [1, 2]. Web surveys have become the standard format for these course evaluations [3, 4]. Despite the widespread use of web surveys to

collect quantitative data, previous research has highlighted some limitations such as low response rates and low-quality responses to open questions [5, 6]. Several reasons have been offered to explain these limitations. For instance, respondents' satisficing behaviour, survey fatigue and the static interaction style are among the key reasons that can explain these negative effects [7]. Respondents' satisficing behaviour occurs to mitigate cognitive burden by responding to survey questions in a non-differentiation style; hence, generating satisficing responses instead of accurate ones, as argued by Krosnick [8]. Roster, Rogers, Albaum and Klein [9] claim that, by using web surveys, responses are far more susceptible to satisficing behaviour and poor-quality responses compared to face-to-face or telephone surveys. In addition, survey fatigue occurs when respondents feel bored or uninterested in the survey and as a result, provide inaccurate responses to the survey questions [10].

To overcome these limitations, Steyn, Davies and Sambo [11] have proposed that individual interviews should be adopted to enhance the quality of responses and to gain in-depth insights. However, due to resource constraints, interviewing students is unpractical as lecturers, especially at the time of COVID 19, are trying to cope with unprecedented workload. Alternatively, Kim et al. [7] and Wambsganss et al. [3, 12] argue that the use of artificial intelligence conversational agents (CAs) or 'chatbots' is promising and can facilitate collection of high-quality course feedback as these intelligent agents can play the interviewer role. CAs are software programs that communicate with users through natural language interaction interfaces [13]. Wambsganss et al. [3] claim that CAs not only enhance the quality of responses but also boost students' levels of enjoyment. The use of CAs in education is growing and the technology is still evolving [14, 15].

While many aspects of the use of CAs in education could be examined, this paper focuses on exploring the opportunities and obstacles of using 'Hubert.ai' (an artificial intelligence CA) in collecting course feedback from students at three universities in the UK, Spain and Croatia and one research institute in Cyprus. This study contributes to the educational CA literature, by highlighting key patterns that characterise the use of smart CAs, like Hubert.ai, in collecting course feedback in higher education institutions. Our analysis outlines and demonstrates these patterns to facilitate advancing the use of artificial intelligence chatbots in education.

2 Background

There is a growing interest in the use of CAs or chatbots in educational settings as they can provide efficient and timely services to students [16]. The authors differentiate between two categories of educational CAs: service oriented CAs and teaching oriented CAs. Service oriented CAs like Ask L.U. [17], which is built on Amazon Web Services, provide services to students using voice technology. Ask L.U. delivers a voice interface to answer students' questions about timetables, grades, tutors, societies, clubs and assist

them in booking spaces to meet their peers [17]. Examples of other service oriented CAs that are used to assist students during enrolment and admission are Lola [18] and Dina [19]. Differ and CourseQ [20] are CAs that can assist students in the various services offered by universities that promote student engagement in the learning process. Lisa [21] acts as a virtual assistant to introduce new students to university life. Many of these CAs have shown good results as their success is assessed either through questionnaires or by the number of satisfactory answers they generate [16].

On the other hand, teacher oriented CAs aim to act as teaching assistants, reinforcing learning of students through generating knowledge like a human tutor, hence, relieving the workload of teachers [16]. For instance, Coding Tutor [22] is developed to support university students studying introductory programming courses in writing software code and getting an automatic assessment of this code through step-by-step guidance using natural language interactions. Providing feedback to students about their performance aims to initiate metacognitive thinking processes that can boost students' motivation and engagement [14]. Other teacher oriented CAs are developed to promote language learning, e.g., Bookbuddy [23], Clive Chatbot [24] and Mobile Chatbot [25]. As the main objective of these chatbots is to teach, their assessment is based on their teaching efficiency and on achieving predefined learning outcomes [16]. Other quality metrics that have been used to evaluate CAs include the length and structure of the conversation [26]. Smutny and Schereiberova [15], after evaluating 47 educational CAs implemented in Facebook Messenger, propose four categories of quality metric: teaching, humanity, affection and accessibility.

Furthermore, Griol, Molina and Callejas [27] argue that the use of educational CAs can advance students with disabilities in their studies, reporting benefits such as motivation, improved grades and engagement. Pedagogical CAs can provide personalised learning to students, however, the collaboration of a human tutor and a CA is still necessary to achieve learning outcomes [16]. Despite the positive impact of educational CAs reported by many research studies, further research is needed to establish the long-term effects on both learning processes and learning outcomes [14].

3 Research Methodology

To explore the opportunities and challenges of using educational chatbots or CAs like Hubert.ai in higher education institutions, the Erasmus+ funded EDUBOTS project collaborated with three European universities and one research institute: University of Leeds, University of Granada, University of Zagreb and CYENS Centre of Excellence in Cyprus to advance knowledge about this technology. Hubert.ai is an artificial intelligence chatbot that aims to help educators get students' feedback to improve their teaching and provide students with personalised follow-ups to boost their learning. Using Hubert.ai, the four research teams collected course feedback data from both undergraduate and postgraduate students studying a variety of courses. Each research team

formulated their surveys by choosing from a set of questions provided by Hubert.ai. The questions focused on students' general views of specific modules / courses (rather than overall programmes) and possible avenues for improvement.

A total of 206 students participated in this study. The participation in this study was voluntary. Nevertheless, not all of them completed the survey. The dropout rate, defined as the percentage of students who quit the survey before answering questions about the positive and negative aspects of the course, was 17.4%. Therefore, this study's sample size is $N=170$.

Some instructors collected demographic information about their students like gender and age. In this study, demographic data were collected for three modules: Business Informatics (24 females and 5 males, average students' age was 19.5), Software Engineering (5 females and 19 males, average students' age was 21.5) and Text and Image Editing (12 females and 16 males, average students' age was 19.5). In addition, some educators asked their students to evaluate Hubert.ai and to give the chatbot itself a score. A sample of the survey questions used by instructors in the four universities is shown below.

Beep boop, here we go! By the way, just say "go back" and correct me if I misunderstand you.
 Hi, how old are you? Please type your age as a number
 What is your gender? Please choose Male or Female [Male] [Female]
 What is working well with Course X and should continue in the same way?
 What could the teachers start doing, that would improve it?
 What could the teachers stop doing, that would improve Course X?
 What is your overall experience of Course X? Please write a sentence or two.
 How did you like this type of evaluation compared to a regular survey?
 So on a scale from 1-10, would you give it a 9?

Fig. 1. A Sample of the Survey Questions

The four research teams collated the course feedback transcripts in csv (comma separated values) files. All the feedback data was sanitised and anonymised before being sent to the University of Leeds team for further analysis and data mining.

The transcripts were collated into a single file and several transformations applied to the text to facilitate later analysis; emoji characters were replaced with text descriptions, and identifiers for the questions asked by the CA (see Figure 1) were appended. In addition, numerical ratings were extracted and standardised to lie on a scale from 0 to 10. For responses such as "a 6 or 7", an average value was taken. Where users provided ratings for different aspects of the course ("Laboratory exercise: 10; Lectures (theory): 4") or based on hypotheticals ("9, it would be a 10 without the exam"), the minimum rating offered was taken. In many cases, the chatbot suggested a rating to the user, who could agree with it or provide a different one – the last rating given in response to each question was retained. Conversation lengths were calculated, in terms of both the number of utterances and the total elapsed time.

Statistical data analysis was performed using MS Excel's spreadsheet filters and pivot tables to extract demographic data and the depth of the conversations. Furthermore, qualitative analysis was undertaken to understand the different patterns that characterise students' conversations with Hubert.ai.

4 Results and analysis

The descriptive statistical analysis of the data reveals that students participating in the study are mainly studying computing-related courses, but also a range of science and humanities courses, as shown in **Table 1**.

Course Title	Respondents
3D Modelling and Animation	9
Business Informatics	30
Taxation	14
Data Mining	17
Data Mining and Text Analytics	20
Computer-Mediated Communications	2
Social Information Systems	2
Professional Software Technology Practice	16
Health and Fitness	4
Web Design and Development	15
Statics	11
Predmet	1
Sociology and Pedagogy issues in Physical Education	4
Software Engineering	24
Text and Image Editing	28
Data Mining	9
Grand Total	206

Table 1. Number of Participating Students in each Course

The depth of the conversations with CAs has been identified by Przegalinska et al. [26] as a quality metric that can be used to assess the performance of the chatbot; so, the depth of each conversation was calculated and the average depth of conversation per course is shown in **Table 2**.

It is worth noting that the average conversation depth across all the courses ranged between 26 and 37 utterances. One conversation lasted for 218 iterations as the student tried to challenge Hubert.ai's intelligence. This student challenged Hubert.ai at the end

of the survey; after answering all Hubert.ai's questions in detail. Hence, this conversation's length was omitted from the average scores to preserve consistency of the data.

Course Name	Avg Depth of Conversation
3D Modelling and Animation	28
Business Informatics	27
Taxation	32
Data Mining	29
Data Mining and Text Analytics	30
Computer-Mediated Communications	29
Social Information Systems	35
Professional Software Technology Practice	31
Health and Fitness	27
Web Design and Development	30
Statics	32
Predmet	37
Sociology and Pedagogy issues in Physical Education	30
Software Engineering	29
Text and Image Editing	28
Data Mining	26

Table 2. Average Depth of Conversation (No. of Utterances) per Course

Some respondents supplied numerical ratings (out of 10) for their overall experience of the course and/or for the experience of using the Hubert.ai chatbot to supply feedback. In total, 218 such ratings were given (as some respondents gave ratings for both of these questions). The distributions of ratings are shown in **Fig. 2**. Distribution of Numerical Ratings by Question

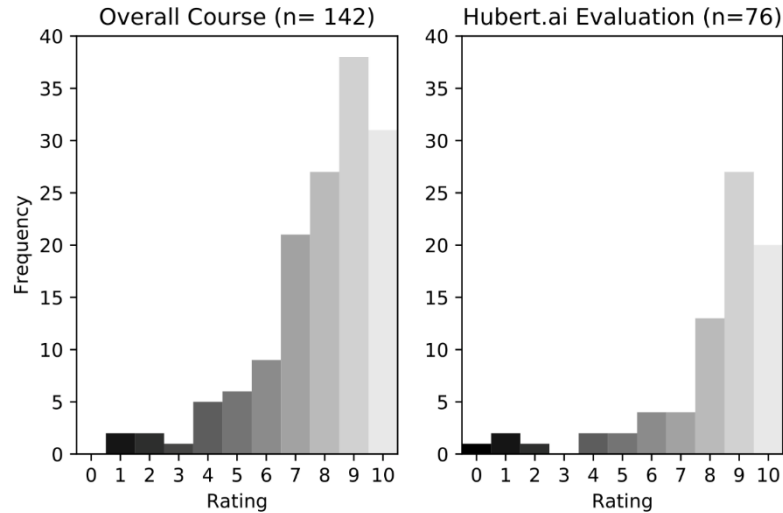


Fig. 2. Distribution of Numerical Ratings by Question

In the context of collecting course feedback in higher education institutions, a key difference between the use of a web survey and a conversational agent is the latter's ability to mimic human-human conversations; by asking questions in a friendly casual way to students and pursue high-quality responses. Hubert.ai uses emoji in his replies, mirroring the texting style that humans usually adopt in virtual conversations with their friends. Embedding anthropomorphic design elements like informal language and emoji in conversational agents has been proven to positively influence the interactivity of the conversation [7, 28].

In addition, Hubert.ai can express empathy with students who are happy/unhappy about some course aspects. For instance, when students claim that they like everything about the course, Hubert.ai would ask them:

- *Ok, tell me about what was best then 😊*
- *Are you absolutely, definitively 100% sure about that 😊?*

Evidently, in this study, data reveals that this technique has helped students to think more about their answers and to provide more constructive replies to Hubert.ai as shown below. As web surveys have a predefined design, it is impossible to match conversational agents' flexibility or 'intelligence'. As argued by Wambsgans et al. [12], this feature can mitigate common web survey challenges like satisficing behaviour and survey fatigue.

In addition, in many cases where Hubert.ai received short answers from students, he asked them to elaborate on these. In most conversations, this successfully helped students to provide more precise and focused answers. An example of Hubert.ai's ability to guide towards constructive answers from students is shown in Figure 3.

Hubert.ai	What could the teachers start doing, that would improve it?
User	I think everything is alright. Nothing to improve
Hubert.ai	Surely there must be something that could be improved?
User	I can't remember anything bad
Hubert.ai	Not yet!
Hubert.ai	Let me repeat the question
Hubert.ai	What could the teachers start doing, that would improve it?
User	I can't remember anything bad
Hubert.ai	Not yet!
Hubert.ai	What could the teachers start doing, that would improve it?
User	Maybe weekly tasks could be equal, because I heard that some students got harder tasks than others
Hubert.ai	I understand...
Hubert.ai	What could the teachers stop doing, that would improve course X?
User	I think that teachers should put more information online, not in last week or something like that
Hubert.ai	I'm sorry to hear that
Hubert.ai	What is your overall experience of course X? Please write a sentence or two
User	irrelevant*
Hubert.ai	Alright, noted. Let's continue.
Hubert.ai	Do you want to add something more?
User	Overall experience: I have been very frustrated and angry. As I said professors have been wasting times in preparing videos that have a little impact on students' teaching a lesson. Professor V has gave us a many material but in exam it was very hard to understand. A point is there is no connection between two topics at least. It's shameful....This year, on this course, there was so many cheating. I regret that I haven't cheated. Prof. B should be ashamed that he's said that this generation is better than generation 2014/2015. I guess I am correct. One little advice to professor is to simplify lesson or course PI with important information. Don't give us boring materials that make no sense. Lessons are boring and exams are very tough. Grade for this course is 1-10: 5....I forget to mention about laboratory tests. They were tough. I recommend that you check knowledge every 2-3 weeks. Overall, for me, laboratory exams were very tough.
Hubert.ai	Got it

Fig. 3. A Sample conversation guiding the student to more constructive answers

4.1 Evaluation of the Conversational Agent

Some students expressed positive perceptions about the use of a conversational agent like Hubert.ai; many students perceived their experiences with Hubert.ai as fun and more interesting than web surveys, as demonstrated in Fig. 4. This was also reflected in the overall numerical ratings presented in **Fig. 2**.

Hubert.ai	How did you like this type of evaluation compared to a regular survey?
User	Very interesting and unique.
Hubert.ai	So on a scale from 1-10, would you give it a 10?
User	Let it be a 10!

Hubert.ai	How did you like this type of evaluation compared to a regular survey?
User	Interesting.
Hubert.ai	So on a scale from 1-10, would you give it a 10?
User	Yes
Hubert.ai	I see...
Hubert.ai	How did you like this type of evaluation compared to a regular survey?
User	Yes, it's fun
Hubert.ai	So on a scale from 1-10, would you give it a 10?
User	yes, without no doubt

Hubert.ai	How did you like this type of evaluation compared to a regular survey?
User	I think this way is better for communication.
Hubert.ai	So on a scale from 1-10, would you give it a 10?
User	Good!

Hubert.ai	How did you like this type of evaluation compared to a regular survey?
User	It was a much more interesting approach to feedback, much more bearable to fill out. I really enjoyed it to be honest, really cool.
Hubert.ai	So on a scale from 1-10, would you give it a 10?
User	Yeah

Hubert.ai	How did you like this type of evaluation compared to a regular survey?
User	I really enjoyed it, comparing it to the boring evaluation this gives you a sense of involvement and making a change to the productivity of the institute in a good way.
Hubert.ai	So on a scale from 1-10, would you give it a 10?
User	maybe 9

Hubert.ai	I'll ask a question about myself as well:
Hubert.ai	How did you like this type of evaluation compared to a regular survey?
User	It's much more interesting and fun :)
Hubert.ai	So on a scale from 1-10, would you give it a 10?
User	Yes
Hubert.ai	I see...
Hubert.ai	How did you like this type of evaluation compared to a regular survey?
User	this way is total different with regular one. However, it is interesting
Hubert.ai	So on a scale from 1-10, would you give it a 7?
User	maybe 8
Hubert.ai	How did you like this type of evaluation compared to a regular survey?
User	it's really interesting, better than filling a form
Hubert.ai	So on a scale from 1-10, would you give it a 10?
User	10

Fig. 4. Examples of positive perceptions of use of Hubert for course evaluation

On the other hand, a few students argued that regular static web surveys are better than conversational agents, as shown in Fig. 5.

Hubert.ai	How did you like this type of evaluation compared to a regular survey?
User	No difference from regular survey other than the scripted responses

Hubert.ai	How did you like this type of evaluation compared to a regular survey?
User	Still not realistic enough - don't feel like i'm chatting. Perhaps more option paths can be created

Fig. 5. Examples of negative perceptions of use of Hubert for course evaluation

Moreover, this study's data captures some indicators of students' enjoyment which is demonstrated through students' use of humour during their interactions with Hubert.ai. This finding confirms Wambsganss et al.'s [3] claim that the use of conversational agents in collecting course feedback from students can have a positive impact on students' level of enjoyment compared to web surveys. Many respondents mirrored Hubert.ai's language and used emoji and generally friendly comments in greeting Hubert.ai as shown in Fig. 6.

Hubert.ai	Thank you those were all the questions I had! You may close this window now. Have a great day! ☺
User	<i>You too bot!</i>

Hubert.ai	Are there any other points you would like to comment on?
User	You are a very clever bot Hubert
Hubert.ai	My bad!
Hubert.ai	Let me repeat the question:
Hubert.ai	Are there any other points you would like to comment on?
User	Nope! Stay safe and healthy
Hubert.ai	I understand

Fig. 6. Examples of positive friendly responses to Hubert

Generally, there are no clear differences observed in the quality of course feedback provided by students of different genders. However, some male students attempted to challenge Hubert.ai's intelligence. Framing systems like conversational agents as anthropomorphic often leads to raising user expectations of the intelligence of these systems [29]. Dibitonto et al. [21] argue that as a result of a false sense of anonymity when chatting with conversational agents, users can tend to be ruder than when chatting with a human interviewer.

It's worth noting that Hubert.ai's handling of these challenges from users did not lead to a failure in handling the conversation. This contrasts with other conversational agents that have been developed using the Facebook Messenger API as reported by Kim, Lee and Gweon [7]. The authors argue that this failure has caused student frustration as a result of the agent posing the same question repeatedly despite getting a full answer to the question.

As mentioned above, the longest conversation with Hubert.ai lasted for 218 utterances. Examples of some students' challenging responses are shown in Fig. 7.

Hubert.ai	Thank you, those were all the questions I had! You may close this window now. Have a great day! ☺
User	Bitch
Hubert.ai	"A poor choice of words never improved anything" - Hubert the bot
User	tell me a joke
Hubert.ai	Knock knock.
User	whos there
Hubert.ai	Hey there!
Hubert.ai	I don't have any questions for you right now. Please come back later!
User	sick joke man

Hubert.ai	What is working well with course X and should continue in the same way?
User	yes ofskrzo
Hubert.ai	Ok, what?
User	why yes
Hubert.ai	Ok, what?
User	u limited
User	yes \\ u limited
Hubert.ai	Ok, what?
User	Yes
Hubert.ai	Ok, what?
User	go back
Hubert.ai	Are you sure you want to go back one question and answer it again?
User	Yes
Hubert.ai	What is your gender? Please choose Male or Female [Male] [Female]
User	Male

Hubert.ai	I don't have any questions for you right now. Please come back later!
User	Coronavirus
User	What does coronavirus cause
Hubert.ai	I'll try to answer that, but I may be totally wrong. Give me a sec.
Hubert.ai	SARS-CoV-2 virus
Hubert.ai	Now it's your turn to answer my question:
User	Sure
User	Can you send me an emoji

Hubert.ai	Thank you, those were all the questions I had! You may close this window now. Have a great day! ☺
User	Are you still there
Hubert.ai	Yes.
User	Can I ask you questions
Hubert.ai	Hey there!
Hubert.ai	I don't have any questions for you right now. Please come back later!
User	What do you like to eat
Hubert.ai	I'll try to answer that, but I may be totally wrong. Give me a sec.
Hubert.ai	I'm sorry but I can't answer that. Even Googled it.
User	Hello \\ What's today's weather
Hubert.ai	I'll try to answer that, but I may be totally wrong. Give me a sec.

Hubert.ai	I'm sorry but I can't answer that. Even Googled it.
User	What is the time now?
Hubert.ai	I'll try to answer that, but I may be totally wrong. Give me a sec.
Hubert.ai	9:36:05 A.M. Eastern Daylight Time, Tuesday, May 12, 2020
Hubert.ai	Now it's your turn to answer my question:
User	Okay
User	What is it?
Hubert.ai	I'll try to answer that, but I may be totally wrong. Give me a sec.
Hubert.ai	What Is It? is a 2005 American surrealist film written, edited, co-produced and directed by Crispin Glover and starring Crispin Glover, Fairuza Balk, and Steven C. Stewart.
Hubert.ai	How did you like this type of evaluation compared to a regular survey?
User	I like you are so cute
Hubert.ai	So on a scale from 1-10, would you give it a 10?
User	want to go on a date?
Hubert.ai	That was all, thanks! You can close this window now.
User	you too
User	Thanks
Hubert.ai	Sure buddy 😊
User	who are you
Hubert.ai	I am Hubert, the world's first artificially intelligent evaluation bot. Here to serve you!
User	can you give me money?
Hubert.ai	I'll try to answer that, but I may be totally wrong. Give me a sec.
Hubert.ai	I'm sorry but I can't answer that. Even Googled it.
User	how old are you
Hubert.ai	I am Hubert, the world's first artificially intelligent evaluation bot. Here to serve you!
User	love you
Hubert.ai	Thank you, I have no more questions now. You can close this window!
User	say me too
User	No
User	Yes

Fig. 7. Examples of student responses that were challenging for Hubert

This study's data indicates that older students (21+) completed answering Hubert.ai's questions with a fewer number of iterations compared to the younger students (18, 19 and 20 years old). One explanation for that can be attributed to the extent of students' experience of the university courses as they head towards the end of their academic degrees.

Unlike some Facebook Messenger conversational agents [7], Hubert.ai allows students to modify their answers, as shown in Fig. 8.

Hubert.ai	What is working well with course Y and should continue in the same way?
User	same assignment
User	Sorry
User	I think just keep the way of assessment
Hubert.ai	Ok
Hubert.ai	What is your gender? Please choose Male or Female [Male] [Female]
User	No
User	We can go now
User	Yes
User	Female
Hubert.ai	I see...

Fig. 8. Examples of students modifying their responses to Hubert

5 Conclusions

This study aims to explore the impact of using Hubert.ai, an AI chatbot or conversational agent in collecting course evaluations in higher education institutions. Analysis of the student-chatbot conversation transcripts allowed us to identify key qualities of the interactions:

- the chatbot was able to guide the student to more constructive answers;
- many students volunteered positive perceptions of use of Hubert for course evaluation, though a few were negative;
- students engaged with the chatbot, with positive friendly responses to Hubert;
- some student responses that were challenging for Hubert to analyse;
- this could be alleviated by allowing students to modifying their responses.

The key findings of the study indicate that the use of chatbots has positively impacted students' response quality. Many students have favoured using chatbots over the regular web surveys. Others demonstrated their enjoyment through the use of humour

during their interactions with Hubert.ai. In addition, in most conversations, the chatbot has successfully helped students to provide more precise and focused answers.

However, the data analysed here was gathered from a relatively small number of conversations, and participation was voluntary and offered predominantly to students taking courses in technical subjects – it may be that the self-selected participants were somewhat predisposed towards positive engagement with and evaluation of the conversational agent (and/or towards exploring its limitations).

The EDUBOTS project is ongoing, with further trials of the Hubert.ai chatbot taking place. These are expected to yield additional data which will be used to enable more robust conclusions to be drawn and to enable more detailed analysis and the application of further natural language processing techniques to enable thematic analysis and automated identification of “useful” student feedback.

6 References

1. Fadel, C., Groff, J. S.: Four-Dimensional Education for Sustainable Societies. Sustainability, Human Well-Being, and the Future of Education, Palgrave Macmillan, 269–281(2018). (https://doi.org/10.1007/978-3-319-78580-6_8).
2. Spooren, P., Brockx, B., Mortelmans, D.: On the Validity of Student Evaluation of Teaching. Review of Educational Research, 83 (2013). (<https://doi.org/10.3102/0034654313496870>).
3. Wambsganss, T., Winkler, R., Sollner, M., Leimeister, J. M.: A Conversational Agent to Improve Response Quality in Course Evaluations. ACM CHI Conference on Human Factors in Computing Systems, 1–9 (2020b).
4. Blair, E., Valdez Noel, K.: Improving Higher Education Practice through Student Evaluation Systems: Is the Student Voice Being Heard? Assessment and Evaluation in Higher Education, 39(7), Routledge, 879–894 (2014). (<https://doi.org/10.1080/02602938.2013.875984>).
5. Keränen, H., Holm, E.: Students’ role in quality enhancement – a reflexion of functional stupidity or an implication of ‘Quality as Practice’? Paper presented at the 9th European Quality Assurance Forum, Barcelona, Spain (2014).
6. Richardson, J. T. E.: Instruments for obtaining student feedback: A review of the literature. Assessment & Evaluation in Higher Education, 30, 387–415(2005). doi:10.1080/02602930500099193
7. Kim, S., Lee, J, Gweon, G.: Comparing data from chatbot and web surveys effects of platform and conversational style on survey response quality. Conference on Human Factors in Computing Systems - Proceedings (2019), 1–12 (2019). DOI: <http://dx.doi.org/10.1145/3290605.3300316>
8. Krosnick, J.: Response strategies for coping with the cognitive demands of attitude measures in surveys. Applied cognitive psychology, 5(3), 213–236 (1991).
9. Roster, A C., Rogers, R D, Albaum, G, Klein, D.: A comparison of response characteristics from web and telephone surveys. International Journal of Market Research, 46, 359–374 (2004).

10. Tucker, B., Jones, S., Straker, L.: Online Student Evaluation Improves Course Experience Questionnaire Results in a Physiotherapy Program. *Higher Education Research and Development*, 27(3), 281–296 (2008).
11. Steyn, C., Davies, C., Sambo, A.: Eliciting Student Feedback for Course Development: The Application of a Qualitative Course Evaluation Tool among Business Research Students. *Assessment and Evaluation in Higher Education*, 44(1), 11–24 (2019).
12. Wambsganss, T., Winkler, R., Schmid, P., Sollner, M.: Unleashing the Potential of Conversational Agents for Course Evaluations: Empirical Insights from a Comparison with Web Surveys. *Twenty-Eighth European Conference on Information Systems (ECIS2020)*, 1-18 (2020a).
13. Abu Shawar, B., Atwell, E.: Using Corpora in Machine-Learning Chatbot Systems. *International Journal of Corpus Linguistics*, 10(4), 489–516 (2005).
14. Winkler, R., Söllner, M.: Unleashing the Potential of Chatbots in Education: A State-Of-The-Art Analysis. In: *Academy of Management Annual Meeting (AOM)*. Chicago, USA (2018).
15. Smutny, P., Schreiberova, P.: Chatbots for learning: A review of educational chatbots for the Facebook Messenger. *Computers and Education*, 151, 1-11(2020).
16. Pérez, J., Daradoumis, T., Puig, J.: Rediscovering the use of chatbots in education: A Systematic Literature Review. *Computational Applied Eng Education*, 1-7 (2020).
17. Lancaster University: Lancaster University Launch Pioneering Chatbot Companion for Students (2019). Available at <https://www.lancaster.ac.uk/news/lancaster-university-launch-pioneering-chatbot-companion-for-students>
18. Muñoz, A. P.: Lola, el chatbot inteligente que triunfa entre los estudiantes (in French) (2018). Available at https://retina.elpais.com/retina/2018/11/30/innovacion/1543580663_865121.html.
19. Santoso, H. A. et al.: Dinus Intelligent Assistance (DINA) chatbot for university admission services. *International Seminar on Application for Technology of Information and Communication*, 417–423 (2018). <https://doi.org/10.1109/ISEMAN-TIC.2018.8549797>
20. Brustenga, G., Alpiste, M., Castells, N.: Briefing paper: Chatbots in education (2018). Available at <http://hdl.handle.net/10609/80185>
21. Dibitonto, M., Leszczynska, K., Tazzi, F., Medaglia, C.M.: Chatbot in a Campus Environment: Design of LiSA, a Virtual Assistant to Help Students in Their University Life. In: Kurosu M. (eds) *Human-Computer Interaction. Interaction Technologies. HCI 2018. Lecture Notes in Computer Science*, 10903. Springer, Cham. (2018). https://doi.org/10.1007/978-3-319-91250-9_9
22. Hobert, S.: Say Hello to ‘Coding Tutor’! Design and Evaluation of a Chatbot-based Learning System Supporting Students to Learn to Program. *40th International Conference on Information Systems, ICIS 2019*, 1, 1-17 (2019).
23. Ruan, S. et al. BookBuddy: Turning digital materials into interactive foreign language lessons through a voice chatbot. *ACM Conference on Learning*, 30, 1-4 (2019). <https://doi.org/10.1145/3330430.3333643>
24. Zakos, J., Capper, L.: CLIVE—An artificially intelligent chat robot for conversational language practice, *Artificial Intelligence: Theories, models and applications Lecture Notes in Computer Science 5138*, Springer, Berlin, Heidelberg (2008).

25. Pham, X. et al.: Chatbot as an intelligent personal assistant for mobile language learning. Proceedings of the 2nd International Conference on Education and E-Learning (2018). <https://doi.org/10.1145/3291078.3291115>
26. Przegalinska, A., Ciechanowski, L., Stroz, A., Gloor, P., Mazurek, G: In bot we trust: A new methodology of chatbot performance measures. *Business Horizons*, 62(6), 785-797(2019).
27. Griol, D., Molina, J., Callejas, Z: Incorporating Android Conversational Agents in M-Learning Apps. *Expert Systems*, 34(4), 1-17 (2017).
28. Gnewuch, U., Morana, S., Adam, M. T. P., Maedche, A.: Faster Is Not Always Better: Understanding the Effect of Dynamic Response Delays in Human-Chatbot Interaction. In 26th European Conference on Information Systems (ECIS) (2018).
29. Shedroff, N., Noessel, C.: *Make it so: Interaction Design Lessons from Science Fiction*. Brooklyn, New York: Rosenfeld (2012).