Artificial Intelligence in Education:

AI Conversational Agent for Online Collaborative Learning

Jara Martens

Bachelor Thesis – Module 12

University of Twente

BMS Faculty

Department of Psychology

Supervisor: Dr. P. Papadopoulos

2nd Supervisor: Dr. J. Steinrücke

June 27th, 2023

Abstract

The goal of this study was to gain a better understanding of the changes that a conversational agent may make in terms of user behavioural engagement. A conversational agent is an artificial intelligence (AI) tool that may be employed in a variety of scenarios, including education. In the educational domain, it can assist teachers and help the students in their learning development.

A dyad discussion with the participation of a conversational agent was intended to increase the users' behavioural engagement. In this study, 18 participants were asked to take place in an online experiment in which they held a dyad discussion in two Phases, one without Clair, our collaborative AI agent, and the other with Clair. According to the analysis, there is not a significant difference between the two Phases. This study is a good starting point for further research on the changes that a conversational agent can bring about.

Keywords: conversational agent, artificial intelligence, collaborative learning, scripted collaboration, online collaboration, behavioural engagement

Artificial Intelligence in Education: AI Conversational Agent for Online Collaborative Learning

Introduction

Over the past several years, online education has become increasingly utilised and a more relevant topic. As the global COVID-19 pandemic started, in-person instruction had to be minimised and lessons were only offered online. This resulted in the evolution and expansion of online education all over the world, in which the lessons were held on online platforms, such as Zoom or Microsoft Teams (Kansal et al., 2021). Having classes in an online environment can take away the control the teachers have over their classes and challenges the collaborative learning process within a class online (Silalahi & Hutauruk, 2020). With the new focus on online learning communities, the qualities of chatbot technology appear to be even more significant, especially given the limited help that instructors and teaching staff can normally provide (Roll & Wylie, 2016).

Online environments had a significant impact on collaborative learning. Collaborative learning is a widely used approach wherein participants are divided into groups to work as a team to arrive at a shared learning goal (Dillenbourg, 1999; Teasley et al., 2008). Moreover, it is built around the idea that learning requires knowing how to solve problems and resolve conflicts while interacting with one another (Dillenbourg, 1999). The collaborative learning approach can be used in different forms and settings. A form of this approach can, for example, be group discussions (Weinberger et al., 2007).

Every group discussion involves learning challenges. When it comes to specific learning challenges, students seem to be comparable to their classmates (Laal & Ghodsi, 2012). A common foundation for collaborative learning is the assumption that students have unique learning resources and unshared prior knowledge (Schnaubert & Bodemer, 2019). Hereby, there is evidence that learners benefit most from more top-down types of information sharing in collaborative learning. For instance, when students create counterarguments after being presented with information that is different from their own (Jeong et al., 2019).

As previously stated, collaborative learning can be used in a variety of forms. One of the more recently developed forms is computer-supported collaborative learning (CSCL). CSCL can be defined as a form of collaborative learning technique that uses technology to look over, assist and help learners during their learning process and in achieving their learning goals (Fischer et al., 2013; Cress et al., 2021). CSCL gained prominence as technology systems and computers became more and more common in daily life, particularly during and after the global COVID-19 pandemic (Nolan, 2022). Hereby, AI can provide chances to assist computer-supported collaborative learning and the types of ambitious learning practises associated with CSCL (Hmelo-Silver, 2022). Because of that, studies in the field are extremely relevant at the moment, and attempts are being made to discover its potential and overcome its obstacles (Simon, 2022).

Artificial intelligence (AI) is frequently viewed as a “game-changer” in terms of offering individualised educational experiences along with unique creations of innovative opportunities for comprehending the true purpose of the learners (Tegos et al., 2020). The AI systems are focused on creating intelligent machines that are capable of activities that typically require human intelligence, including perception, learning, and argumentation (Wang, 2019). These goals are accomplished by implementing the latest technologies, deep learning and machine learning (Shah et al., 2022). AI systems are applied in a wide range of settings. They can be used in normal day-to-day life, for example, in the form of learning apps. One of the most prominent AI systems in education are conversational agents.

Conversational agents are commonly referred to as conversational AI or chatbots (Tegos et al., 2020). Conversational agents have become an established use of AI in the educational domain (Demetriadis et al., 2018). It has been suggested that conversational agents hold significant promise for learning establishments and companies (Wollny et al., 2021). These agents can be thought of as computer programs that can communicate with students in natural language through auditory or textual means in order to achieve one or more educational outcomes (Murad et al., 2019). A chatbot may use many AI approaches to imitate student-to-teacher or peer-to-peer dialogues, to help the student feel more at ease when interacting with a virtual agent (Demetriadis et al., 2018). Texting has become widely recognized as being among the most intriguing forms of computer-human communication and chatbots are therefore starting to affect a variety of industries including the educational sphere (Tegos et al., 2020). Due to chatbots, natural language processing (NLP) technologies have become more available, allowing developers to create interfaces that simulate human-to-human conversation (Wollny et al., 2021).

Teachers have very limited resources to help every individual in a class with their learning process (Mertens, 2019). In those situations, conversational agents can help the teacher and students in the way that they offer automatic support, as they can constantly be present with the students (Tegos et al., 2016). Conversational agents have emerged as one of the most important educational innovations to direct and assist student interaction using natural language in both collaborative and individual contexts (Caballé & Conesa, 2018). According to research, conversational agents can increase students' motivation and engagement in a task which improves their pedagogy (Tegos et al., 2020). This type of agent can be a useful tool in real educational contexts, as the agent is able to make up for a learner's lack of assistance and therefore improve their learning process (Tegos et al., 2016). In order to give the students a feeling of safety while using conversational agents, the agents must have a narrative and common spoken language, as this produces an informal, confidential, and comfortable conversational mood (Li et al., 2022). One of the most effective tools teachers have to plan and carry out instructional activities to support students in meeting their learning objectives is language (Li & Graesser, 2021). Thus, the success of student learning and the growth of students` language are both influenced by the teacher's use of dialogues, language, and conversational patterns (Li & Graesser, 2021).

Conversational agents have emerged as one of the most important educational innovations to direct and assist student interaction using natural language in both collaborative and individual contexts (Caballé & Conesa, 2018).

Current study

The current study will focus on the newly developed conversational agent, the collaborative learning agent for interactive reasoning (Clair). This system was developed to help both students and teachers. When students are working online, they might be divided, for example via Microsoft Teams or Zoom into breakout rooms, the teacher is hereby not able to follow every conversation at the same time, as they can only join one of the student's breakout rooms at a time (de Araujo et al., 2023). This problem arises not only in online situations but also when teaching in schools. Because one teacher must supervise the work of many students at once, it can be difficult to keep track of each individual's performance on tasks and development (de Araujo et al., 2023). Clair focuses on an online discussion between two students, who are communicating via chat about a given topic. Therefore, Clair's task is to interact with the students in order to help them with the conversational flow and ensure that the students stay focused on the task and collaborate effectively with one another (de Araujo et al., 2023). Therefore, Clair's core instructional approach is to create an intellectually successful conversation. To achieve that goal, Clair uses talk moves. An example of a talk move is "Recapping" where Clair would ask "Can somebody give the partner a summary of what we've covered so far?" (de Araujo et al., 2023).

As mentioned before AI is becoming of great value and can especially be used in the educational domain for CSCL. Because of that, this study will focus on comparing the behavioural engagement of the users between an online discussion without Clair and an online discussion with Clair. Furthermore, testing how the students perceive the quality of the conversation is essential because it can show how well conversational agents operate as mediators for facilitating constructive discourse, knowledge exchange, and meaningful relationships among students. Lastly, another focus of this study will concentrate on the impact Clair has on the responsiveness of the participants, which will represent the active involvement, attention, and willingness to participate in collaborative learning activities with a conversational agent. Examining Clair's effects on participants' responsiveness can provide insights into the effectiveness of the conversational agent.

Therefore, these research questions will be tested:

Research Question 1 (RQ1): How does Clair affect behavioural engagement?

Research Question 2 (RQ2): How do students perceive the quality of their talk?

Research Question 3 (RQ3): Does Clair have an impact on participants' responsiveness?

Method

Participants

The data collection of this study took place at the University of Twente, Enschede from the 4th of April to the 4th of May 2023. The study was conducted in an online environment via Microsoft Teams Meetings and the Graasp system (https://graasp.eu/). The participants were randomly assigned to a username and paired up with another participant to do the tasks. This research had a total of 24 participants, but due to an error that occurred in the Clair, six participants had to be excluded from the data.

Therefore, the final data set consisted of 18 participants. 17 participants are from Germany, and one participant is from the Netherlands. 13 of the participants are females and five are males. The age range of the participants is from 19 - 23 (M_{age} = 21.6, SD_{age} = 1.10). Simple random sampling was used to select the participants through an online system.

Additionally, students were asked to take part in the study and flyers were spread out in the buildings of the University. In order to take part in the study, participants had to (1) be Psychology students, (2) have the ability to write and understand English, and (3) have access to a computer.

Materials and Instruments

For this study a Graasp learning environment was created (Appendix A). The Graasp Environment included the consent form (Appendix B), a questionnaire about the participant's demographical background (Appendix C), a familiarization task description (Appendix D), the discussion topic 1 description (Appendix E), the discussion topic 2 description (Appendix F), a questionnaire about the participant's experience during and after the task, a questionnaire about their opinion and engagement with the Clair the conversational agent (Appendix G), a collaboration tool, with which the participants were paired up randomly, the chat and the conversational agent Clair. Furthermore, the participants needed a laptop or computer to join the online Teams Meeting and open the Graasp system. The items the subjects were required to answer following the experiment were measured in the questionnaire using a five-point Likert scale (Appendix G). Finally, R is the program that enabled the data analysis (R Core Team, 2021).

Independent Variables

As stated, the participants are paired up randomly with another participant, with whom they have to discuss different statements in the chat. Hereby, the participants had two conditions, the first was having a discussion without the interaction of the conversational agent. In the second condition, the participants held the discussion with the interaction of the conversational agent. Therefore, the presence of Clair was the independent variable.

Dependent Variables

Participants’ Behavioural Engagement

In order to measure the outcome of how the conversational agent impacts the participants’ behavioural engagement in the discussion, the engagement of the participants in the first condition without the interaction of the conversational agent was compared with the results of the behavioural engagement during the second condition, where the conversational agent interacted in the discussion. Therefore, the level of behavioural engagement in the online dyad discussion is the dependent variable for RQ1. In order to test the behavioural engagement of the participants the discussions were coded on the items that are shown and defined in Table 1.

Table 1
Definition of the Behavioural Engagement Items.

Item(s)	Definition
Number of messages (scale)	Count the number of messages that were used.
Words per message	The words that were (on average) used per message.
Number of turns	The number of turns during the discussion.
Words per turn	The average number of words used for each turn.
Informative (IN)	Probability of being basic information.
Argumentative (AR)	Probability of being an argumentation.
Asking for Information (AI)	Probability of being a question.
Frequency of high TSIM	Topic keywords have a high frequency of semantic similarity.
Topic accumulation	The ratio of the speaker's total TSIM.
Messages per minute	The number of messages that are posted per minute.

Perceived Talk Quality

For RQ2 the perceived quality of the talk by the participant is the dependent variable. A questionnaire with nine questions separated into three scales was provided to test this (Appendix G). The first subscale is Accurate Knowledge (AK), Learning Community (LC) is the second subscale, and the third subscale is Rigorous Thinking (RT). The respective questions that were used to assess each scale are shown in Table 2. The statements were scored on a five-point Likert scale, with answers ranging from 'Strongly disagree' (1) to 'Disagree' (2) to 'Neutral' (3) to 'Agree' (4) to 'Strongly agree' (5). These questionnaire items were created by altering Chen et al. (2020), students-perceived discursive engagement questionnaire to more specifically test the accountability elements described by Michaels et al. (2013). To test the internal consistency reliability, Cronbach's Alpha was calculated after determining the final dataset. Cronbach's alpha for AK (α = .63) indicated a moderate level of reliability. Further, the scale LC (α = .75) indicates a good level of reliability. Lastly, the RT scale has a low level of reliability, with Cronbach's alpha (α = .56).

Table 2
Scale(s) and Respective Questions of the Perceived Discursive Engagement.

Scale(s)	Respective Questions
Accurate Knowledge (AK)	“I gathered information to support my ideas”, “I checked whether the information that our group gathered was correct”, “I discussed with my classmate what information was needed to progress on the task”
Learning Community (LC)	"I listened to my classmate without interrupting until it was my turn to speak", "I listened to my classmate’s opinions to get inspiration", "I defended my position respectfully while discussing with my classmate"
Rigorous Thinking (RT)	“I checked whether our arguments were clear and coherent”, “I discussed whether our arguments are sufficiently convincing”, “I discussed with my classmate cases where our arguments may not be correct”

Participant's Responsiveness

The dependent variable for RQ3 is the participant's responsiveness to the intervention. This was measured by qualitative coding of the responses to Clair. The codes are divided into three categories, namely Responded (RES), Acknowledged (ACK) and Ignored (ING). RES was coded when the participants gave a direct answer to the intervention of Clair. Acknowledged was coded when participants responded to Clair but did not provide an answer to the intervention of Clair, for example, ‘Clair I don’t know what else to say’ or ‘Clair I don’t understand what you mean’. Lastly, the interventions from Clair that did not get any response are coded as Ignored.

Procedure

The experiment started with two students signing up on the SONA system of the University of Twente, as each experimental group consisted of two people. On the website of SONA, it stated what would be studied during the experiment, the duration of 60 minutes, where the study would take place and how many credits they would receive for their participation. It also included an email address from the researcher they could contact if they had any questions prior to the study. Once they signed up, they received an email with confirmation. One day before their study took place, they received another email which gave them their username, which they needed to log in to the system and the link to the Microsoft Teams meeting.

When the participants joined the Microsoft Teams meeting, they were provided with the needed information to log in to the Graasp System, in which the tasks, dyad discussion and questionnaires took place. After everyone was logged in with their correct username, they were allowed to start filling out the consent form (Appendix B) and the Questionnaire about their demographic background (Appendix C). After everyone was filling out both Phases, they were told to go to the next Phase which was Orientation (Appendix D). The participants had 8 minutes to finish this task. After 8 minutes, they had to go to the next Phase, which was Discussion Topic 1 (Appendix E). For this task, the participants had 15 minutes. After that, they had to do the task under Discussion Topic 2 (Appendix F). During Discussion Topic 2, the conversational agent Clair was turned off and interacted with in their discussion. When the time was over, the participants had to do the last Phase which was filling out questionnaires about their experience with the task and with Clair (Appendix G).

Data Analysis

After the data-collecting procedure was completed, the acquired data was imported from Graasp and prepared for analysis. The dataset was examined for missing data, and the data of six participants were excluded from the data set because an error occurred in the second Phase with the Clair. After that, the dataset was imported into RStudio for data analysis. The data were analysed using the statistical programming language R (Version 4.1.0) along with the interface RStudio (Version 1.4.1717). The packages foreign, tidyverse, janitor, haven, readr, dplyr, psych, modelr and stats were used. Demographic data were analysed using descriptive statistics. A Wilcoxon signed-rank test was used to examine RQ1. The Wilcoxon signed-rank test findings were initially utilised to evaluate the effect Clair had on the behavioural engagement of the participants during the dyad discussions, as posed by RQ1. In order to test RQ2, how students perceived the quality of their talk, the questionnaire was analysed. This was investigated by calculating and comparing the mean values of the three scales. Lastly, by coding the dyad discussion in regard to the reaction of participants towards the questions Clair provided, it was qualitatively tested whether Clair has an impact on participants' responsiveness. This was coded in three reaction types, namely Responded (RES), Acknowledged (ACK) and Ignored (ING) the intervention of Clair.

Results

Research Question 1 - How does Clair affect behavioural engagement?

An overview of the averages and standard deviations for each variable in each Phase is shown in Table 3 and Figure 1. How the Clair affected behavioural engagement within dyads was tested with the Wilcoxon signed-rank test for each variable. The Wilcoxon signed-rank test showed that the intervention of Clair did elicit a statistically significant change in the item 'high frequency of TSIM' in Phase 2 compared to Phase 1 ($Z = 1.34$, $p = 0.002$).

Table 3

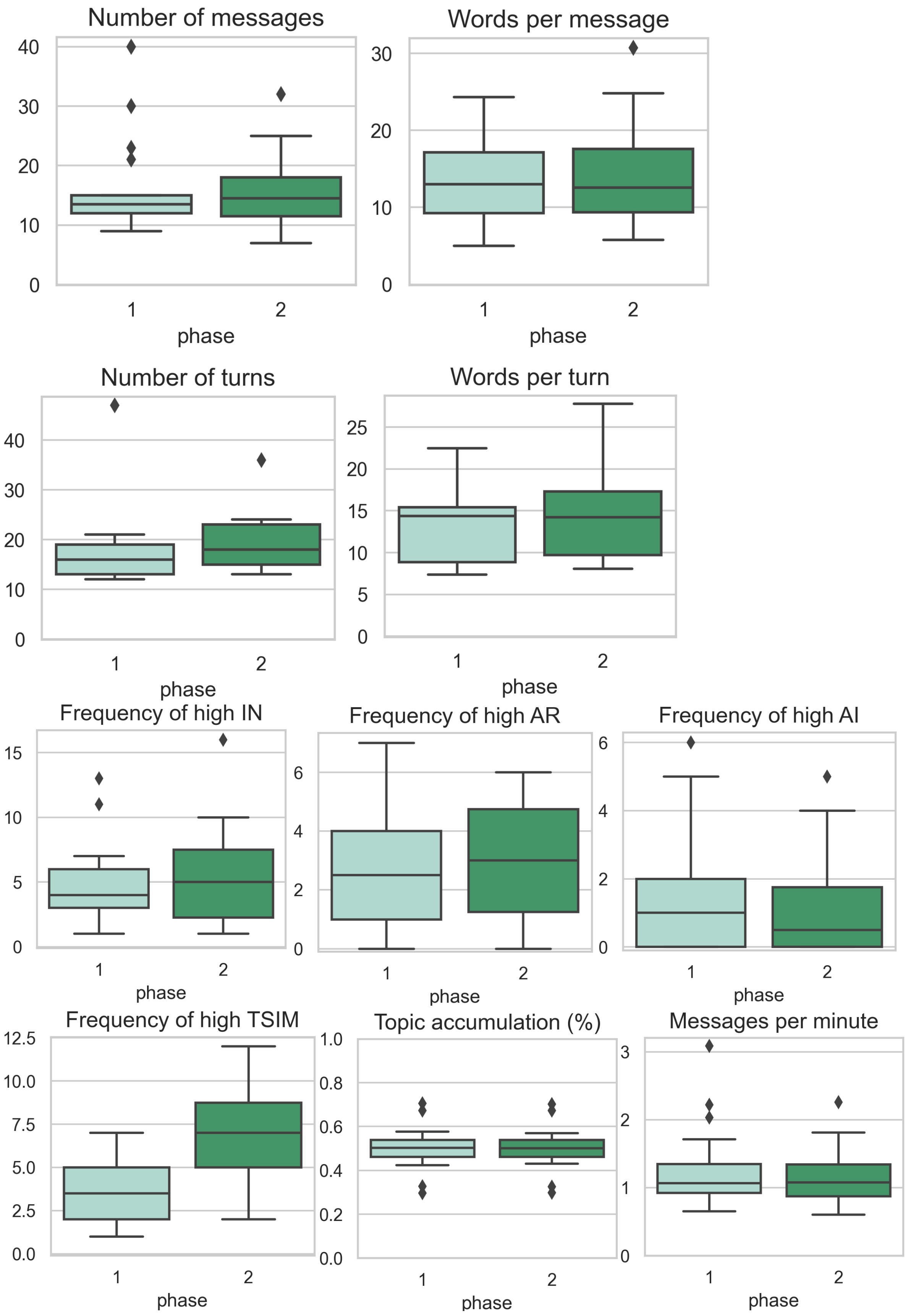
Descriptive Statistics of Metrics from Phase 1 & Phase 2.

	Phase 1 (without Clair)		Phase 2 (with Clair)	
	M	SD	M	SD
Number of messages	16	7.94	15.6	6.19
Words per message	13.3	5.19	14.4	6.62
Number of turns	19.2	10.5	19.7	7.03
Words per turn	13.3	4.86	14.4	6.07
Frequency of high IN	4.83	3.20	5.33	3.82
Frequency of high AR	2.61	1.97	2.94	1.95
Frequency of high AI	1.39	1.85	1.11	1.53
Frequency of high TSIM*	3.67	2.14	6.72	2.52
Topic accumulation (%)	0.50	0.10	0.50	0.10
Messages per minute	1.30	0.61	14.4	0.43

Note. * p <.05; Min. = Minimum Score of the Scale; Max. = Maximum Score of the Scale; Number of messages (Min.: 7, Max.: 40); Words per message (Min.: 5, Max.: 31); Number of turns (Min.: 12, Max.: 47); Words per turn (Min.: 7, Max.: 28); Frequency of high IN (Min.: 1, Max.: 16); Frequency of high AR (Min.: 0, Max.: 7); Frequency of high AI (Min.: 0, Max.: 6); Frequency of high TSIM (Min.: 1, Max.: 12); Topic accumulation (%) (Min.: 0.29, Max.: 0.70); Messages per minute (Min.: 0.6, Max.: 3.1)

Figure 1

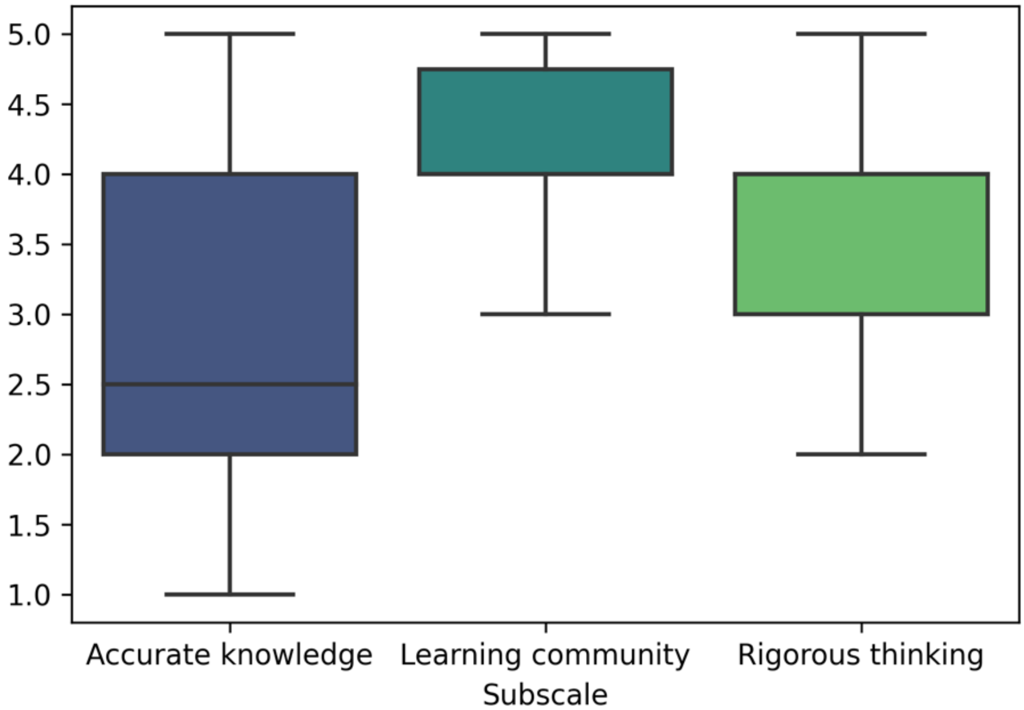
Boxplot of the Descriptive Statistics of Metrics from Phase 1 & Phase 2.



Research Question 2 - How do students perceive the quality of their talk?

The research question "How do students perceive the quality of their talk?" was assessed by comparing the mean values of each of the three scales. According to the mean values from Figure 2, Subscale 2, LC has the greatest mean (*M* = 4.17), suggesting that respondents scored higher on this subscale on average than the other two subscales, implying a sense of perceived belonging, cooperation, and support by the participant with their discussion partner. The mean of Subscale 1 AC (*M* = 2.89) reflects a perceived lack of access to correct information or a lack of comprehension in the examined knowledge domain by the participants. Subscale 3, RT is in the middle, with a mean of 3.5, showing a recognition of the value of critical thinking but possible diversity in their engagement in this area.

Figure 2
Boxplot of the Perceived Discursive Engagement Scale(s).



Research Question 3 - Does Clair have an impact on participants' responsiveness?

In total Clair intervened 33 times during Phase 2 of the research study. A summary of the participants' responsiveness to these interventions is given in Table 4. Hereby, most of the participants responded to the intervention of Clair during their discussion in Phase 2 and actively answered the questions (85%) (Table 4). Most of the time Clair got a response it, also helped the behavioural engagement of both participants and the participants were more drawn to give a more in-depth description of what they meant with their argument. For example, User022: “True and also what the punishment looks like”, Clair: “User022, could you please expand this idea?”, User022: “Punishment could be physical, financial or for example restriction of joyful activities”, User021: “Punishment could be giving something unpleasant or taking away something pleasent”. Another example of a dyad discussion enhanced engagement in the dyad discussion was, User007: “it really decreases your intrinsic motivatoin” Clair: “User007, could you please elaborate more on this?” User007: “well, when you only do something because others want you do and therefore punish/reward you, you don't do it out of your own interest. Furthermore, when keeping up the reward/punishment system for a while, the person might get used to this and always expect others to guide their behaviour in this way”, User007: “this is especially prevalent in upbringing of kids I think”. In this example Clair posted another intervention with which the other User was getting reinvolved in the discussion, Clair: “User, to what your partner just mentioned, would you like to add anything?”, User008: “exactly, I think right now there is kind of a huge debate on social media because many parents say that one should stop rewarding or praising their children after they have done a simple task... so children learn to do things because they like to do it or because the situation expects them to.. and not because the parents reward them”, User008: “like parents should stop saying: yes good job to everything you know”, User007: “I also think that this is especially troubling when parents use severe punishment for educating their children (like ignoring them when they don't behave like wanted by the parents)”.

Only two times the intervention of Clair was acknowledged (Table 4). The participants acknowledged the intervention but did not understand correctly who has to answer, for example, Clair: “User, would you like to add something to what your partner just said?” User023: “@clair, whom of us do you mean?” or did not know what to answer to the question, for example, Clair: “User007, could you please provide more details?” User007: “I don't think I have anything to add”.

Lastly, just three interventions from Clair were ignored by the participants (Table 4). One time the participants ignored Clair because they were currently wrapping up the discussion as the time was over. For example, User016 “FINAL ANSWER Punishment can be helpful in some cases, if it is ethically okay, relevant and if there is no harm or problem in the way” Clair “User015, how does this add to what User already said?”. During another dyad discussion, it was not clear whom Clair meant, for example, Clair: “User, do you agree or disagree with your partner?” and therefore none of the two participants answered.

Table 4
Frequency Table.

	Frequency	Percentage
Responded	28	85%
Acknowledged	2	6%
Ignored	3	9%
Total	33	100%

Discussion

Context of this research

The current study was conducted with the aim to get more insight into and collect additional data on Clair and the effect the system can have on the behavioural engagement of the users. As a result, the purpose of this study was to investigate if the students' behavioural engagement altered according to whether or not Clair intervened in the debate. Similarly, the difference between dyad discussions without and with Clair interaction was explored and tested using 10 items analysing behavioural engagement, a questionnaire implementing the perceived feelings of behavioural engagement by the participants, and codes on how the participants reacted to the Clair intervention.

Main findings

The prior results section covers all of the research questions in this study. The first research question, "How does Clair affect behavioural engagement?" asked if Clair had a substantial influence on behavioural involvement in dyad discussions. According to the findings, the behavioural engagement level of participants in Phase 1 (without Clair) did not change significantly from that of participants in Phase 2 (with Clair). Based on that, just one of the items, topic similarity, which measures the participants' behavioural involvement, revealed a significant difference. The results can be explained in part by the limited sample size and tiny combined impact size of Phase 1 and Phase 2 behavioural engagement. Due to a number of reasons, including the difficulty of quantifying behavioural engagement and the limitations of the research design, there were no significant differences in behavioural engagement between Phases 1 and 2. Behavioural engagement is a multidimensional concept with cognitive, emotional, and behavioural characteristics. The use of a limited selection of ten items to measure behavioural engagement may have underestimated the construct's complexity (Fredricks et al., 2011). Further, the specific setting and sample used in this study are reflected in the results. The findings might not apply to different groups or environments. To guarantee the external validity of the findings, future research should replicate the study with a more varied and representative sample (Creswell & Creswell, 2017).

RQ2 was examined using a three-scale questionnaire, with the learning community scale having the greatest mean compared to the others. The fact that students perceived a strong sense of community and teamwork during their talk sessions is indicated by the higher mean score on the learning community scale (Tran, 2013). This might be credited to the use of cooperative learning techniques, which emphasise collaboration and student contact (Slavin, 1996). According to Gillies (2014), when students see a learning community positively, they may be motivated, engaged, and encouraged throughout their discussions. The differences in mean scores among the three measures suggest that students' assessments of the value of their talk vary depending on the particular factors evaluated. The learning community scale's higher mean might indicate that students place a high priority on the social and collaborative aspects of their discussion sessions (Mercer, 2010).

The third research question: “Does Clair have an impact on participants' responsiveness?” is to determine whether the interventions of Clair were accepted by the users and whether they made an impact on the discussion. Even though there was no significant difference found between the items in Phase 1 and Phase 2 in the results of RQ1, participants were open to the intervention and willing to engage with it, as evidenced by the high acceptance rate and engagement with the conversational agent. This implies that the intervention changed the way the participants behaved and engaged. When comparing the arguments of the participants in Phase 1 and Phase 2, it is clear that once Clair engaged with the participants, the participant's response was more thoroughly articulated, and their argument was explained in more detail. By examining some of the participant dyad discussion samples, it can be shown that the agent most likely contributed to improved communication and increased participation. Most of the time when Clair intervened, the participants reflected more on their arguments and could give a deeper and more detailed explanation of what they wanted to indicate (Grimes et al., 2021). This may not show in the form of a higher behavioural engagement in Phase 2 compared to Phase 1, but while analysing RQ3 it was found that the participants accepted the intervention from Clair and implemented it in their discussions.

Limitations

The performed study's low generalizability is one of its weaknesses. The demographic and setting of the study may have an impact on the study's results. As the study took place online it could not be controlled under which conditions the users participated. The generalizability of the findings would be strengthened by replication in various contexts and with a variety of participant populations (Creswell & Creswell, 2017). Further, without a control group, it might be difficult to tell whether the intervention itself is to blame for the reported results or whether there are other confounding variables. Moreover, as the topics were really similar being able to compare them with each other might have also resulted in the participants having the feeling of repeating themselves as they said something similar in Phase 1, which was also directly stated by some of the participants during the discussions. Additionally, the condition with the interventions of Clair was the last Phase of the study. Participants' comprehension and motivation levels were possibly lower than in Phase 1 and the familiarization task. Lastly, the study was conducted at one point in time, having the same participants work with the conversational agent for a longer period of time would provide deeper insights into how the agent affects the dyad discussions.

On the other hand, there were also strengths of this study. First, the study was performed in a realistic and interactive way. This provides a dynamic and engaging experience for the participant. Replicating natural conversation and potentially increasing participant involvement. On top of that, adding a familiarization task in the beginning, helped the participants to feel more comfortable and value the learning community. Finally, because this study is being conducted online, internet experiments frequently offer real-time data gathering, allowing researchers to obtain fast replies and eliminate recollection biases.

Conclusion and Future Research

There is a lot of potential for additional study in the field of employing AI for education. More study is needed, particularly in the evaluation and research of using conversational agents in the educational area. In order to prevent or limit the weaknesses of this research, it would be a possibility to replicate this study in a real classroom and for a longer period of time. Having real students from a class could ensure the conditions under which the participants take place in the study and could limit the possibility of interruptions. Besides, given the study is not restricted to a single moment in time, greater knowledge or conclusions may be generated by spreading the research over a longer period of time and performing more sessions. Adding a control group to this replication would also give more insights into the actual difference between the dyad discussions with the interaction of Clair and without. Furthermore, including an additional experimental group that begins with Clair and finishes without may reveal whether the study had a sequence effect.

To summarize, additional research in the field of conversational agents in educational settings is necessary and required, with an emphasis on replicating the study in real classrooms over a long period of time to overcome potential constraints and acquire more thorough data.

References

Caballé, S., & Conesa, J. (2018). Conversational Agents in Support for Collaborative Learning in MOOCs: An Analytical Review. In *Lecture notes on data engineering and communications technologies*. Springer International Publishing. https://doi.org/10.1007/978-3-319-98557-2_35

Chen, G., Jiahong, Z., Chan, C. K. K., Michaels, S., Resnick, L. B., & Huang, X. (2020). The link between student-perceived teacher talk and student enjoyment, anxiety and discursive engagement in the classroom. *British Educational Research Journal*, *46*(3), 631–652. https://doi.org/10.1002/berj.3600

Cress, U., Oshima, J., Rosé, C. P., & Wise, A. F. (2021). Foundations, Processes, Technologies, and Methods: An Overview of CSCL Through Its Handbook. In *Springer eBooks* (pp. 3–22). https://doi.org/10.1007/978-3-030-65291-3_1

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

de Araujo, A., Papadopoulos, P. M., McKenney, S., & de Jong, T. (2023). Automated coding of student chats, a trans-topic and language approach. *Computers and Education: Artificial Intelligence*, *4*, 100123.

de Araujo, A., Papadopoulos, P. M., McKenney, S., & de Jong, T. (2023). Transferable and Configurable Conversational Agent.

Demetriadis, S., Tegos, S., Psathas, G., Tsiatsos, T., Weinberger, A., Caballé, S., Dimitriadis, Y., Sanchez, E., Papadopoulos, P. M., & Karakostas, A. (2018). *Conversational Agents as Group-Teacher Interaction Mediators in MOOCs*. https://doi.org/10.1109/lwmoocs.2018.8534686

Dillenbourg, P. (1999). What do you mean by collaborative learning? In P. Dillenbourg (Ed.), Collaborative-learning: Cognitive and computational approaches (pp. 1-19). Oxford, UK: Elsevier.

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a Script Theory of Guidance in Computer-Supported Collaborative Learning. *Educational Psychologist*, *48*(1), 56–66. https://doi.org/10.1080/00461520.2012.748005

Fredricks, J., McColskey, W., Meli, J., Mordica, J., Montrosse, B., & Mooney, K. (2011). Measuring Student Engagement in Upper Elementary through High School: A Description of 21 Instruments. Issues & Answers. REL 2011-No. 098. *Regional Educational Laboratory Southeast*.

Gillies, R. M. (2014). Cooperative learning: Developments in research. International Journal of Educational Psychology, 3(2), 125-140.

Grimes, G. M., Schuetzler, R. M., & Giboney, J. S. (2021). Mental models and expectation violations in conversational AI interactions. *Decision Support Systems*, *144*, 113515. https://doi.org/10.1016/j.dss.2021.113515

Hmelo-Silver, C. E., & Jeong, H. (2022). Synergies among the pillars: designing for computer-supported collaborative learning. In *Handbook of Open, Distance and Digital Education* (pp. 1-16). Singapore: Springer Singapore.

Jeong, H., Hmelo-Silver, C. E., & Jo, K. (2019). Ten years of Computer-Supported Collaborative Learning: A meta-analysis of CSCL in STEM education during 2005–2014. *Educational Research Review*, *28*, 100284. https://doi.org/10.1016/j.edurev.2019.100284

Kansal, A. K., Gautam, J., Chintalapudi, N., Jain, S., & Mittal, M. (2021). Google Trend Analysis and Paradigm Shift of Online Education Platforms during the COVID-19

Pandemic. *Infectious Disease Reports*, *13*(2), 418–428. https://doi.org/10.3390/idr13020040

Laal, M., & Ghodsi, S. M. (2012). Benefits of collaborative learning. *Procedia - Social and Behavioral Sciences*, *31*, 486–490. https://doi.org/10.1016/j.sbspro.2011.12.091

Li, H., Cheng, F., Wang, G. J., & Graeser, A. (2022). The Impact of Conversational Agents' Language on Self-efficacy and Summary Writing. In *Springer eBooks* (pp. 553–559). https://doi.org/10.1007/978-3-031-11644-5_48

Li, H., & Graesser, A. C. (2021). The impact of conversational agents' language on summary writing. *Journal of Research on Technology in Education*, *53*(1), 44–66. https://doi.org/10.1080/15391523.2020.1826022

Mercer, N. (2010). The analysis of classroom talk: Methods and methodologies. *British journal of educational psychology*, *80*(1), 1-14.

Mertens, D. M. (2019). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods*. Sage publications.

Michaels, S., Hall, M. W., & Resnick, L. B. (2013). *Accountable talk sourcebook: For classroom conversation that works*. Pittsburgh, PA: University of Pittsburgh.

Murad, D. F., Iskandar, A. T. P., Fernando, E., Octavia, T. S., & Maured, D. E. (2019). *Towards Smart LMS to Improve Learning Outcomes Students Using LenoBot with Natural Language Processing*. https://doi.org/10.1109/icitacee.2019.8904311

Nolan, E. (2022). Transcending Lockdown: Fostering Student Imagination through Computer-Supported Collaborative Learning and Creativity in Engineering Design Courses. *University of Toronto Quarterly*, *91*(1), 67–87. https://doi.org/10.3138/utq.91.1.01

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Roll, I., & Wylie, R. (2016). Evolution and Revolution in Artificial Intelligence in Education. *International Journal of Artificial Intelligence in Education*, *26*(2), 582–599. https://doi.org/10.1007/s40593-016-0110-3

Schnaubert, L., & Bodemer, D. (2019). Providing different types of group awareness information to guide collaborative learning. *International Journal of Computer-supported Collaborative Learning*, *14*(1), 7–51. https://doi.org/10.1007/s11412-018-9293-y

Shah, M., Kshirsagar, A., & Panchal, J. (2022). *Applications of Artificial Intelligence (AI) and Machine Learning (ML) in the Petroleum Industry*. CRC Press.

Silalahi, T. F., & Hutauruk, A. F. (2020). The application of cooperative learning model during online learning in the pandemic period. *Budapest International Research and Critics Institute-Journal (BIRCI-Journal)*, *3*(3), 1683-1691.

Simon, S. (2022, September 14). *Towards A Comprehensive Framework for Situated Collaborative Learning Tools*. https://hal.archives-ouvertes.fr/hal-03782838

Slavin, R. E. (1996). Research on Cooperative Learning and Achievement: What We Know, What We Need to Know. *Contemporary Educational Psychology*, *21*(1), 43–69. https://doi.org/10.1006/ceps.1996.0004

Teasley, S. D., Fischer, F., Weinberger, A., Stegmann, K., Dillenbourg, P., Kapur, M., & Chi, M. T. H. (2008). Cognitive convergence in collaborative learning. *International Conference of Learning Sciences*, 360-367. http://www.gerrystahl.net/proceedings/icls2008/papers/paper192.pdf

Tegos, S., Demetriadis, S., Papadopoulos, P. M., & Weinberger, A. (2016). Conversational agents for academically productive talk: a comparison of directed and undirected agent interventions. *International Journal of Computer-supported Collaborative Learning*, *11*(4), 417–440. https://doi.org/10.1007/s11412-016-9246-2

Tegos, S., Demetriadis, S., Psathas, G., & Tsiatsos, T. (2020). A Configurable Agent to Advance Peers' Productive Dialogue in MOOCs. In *Lecture Notes in Computer Science* (pp. 245–259). Springer Science+Business Media. https://doi.org/10.1007/978-3-030-39540-7_17

Tran, V. H. (2013). Theoretical Perspectives Underlying the Application of Cooperative Learning in Classrooms. *International Journal of Higher Education*, *2*(4). https://doi.org/10.5430/ijhe.v2n4p101

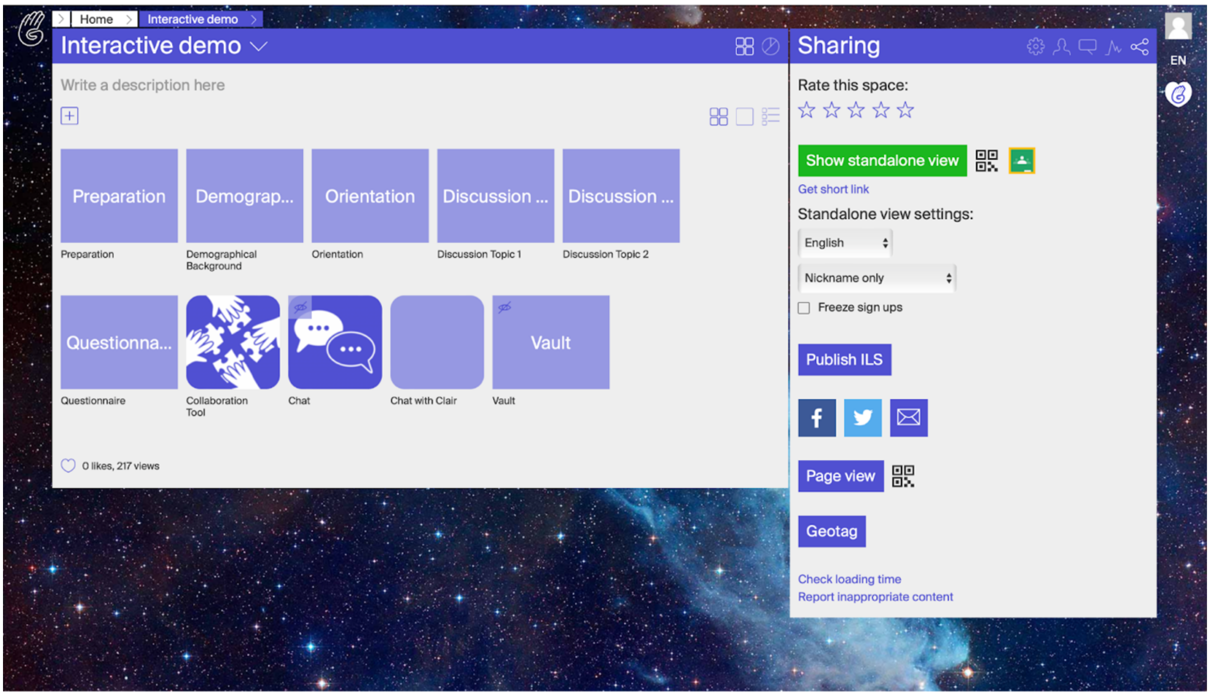
Wang, P. (2019). On Defining Artificial Intelligence. *Journal of Artificial General Intelligence*, *10*(2), 1–37. https://doi.org/10.2478/jagi-2019-0002

Weinberger, A., Stegmann, K., & Fischer, F. (2007). Knowledge convergence in collaborative learning: Concepts and assessment. *Learning and Instruction*, *17*(4), 416–426. https://doi.org/10.1016/j.learninstruc.2007.03.007

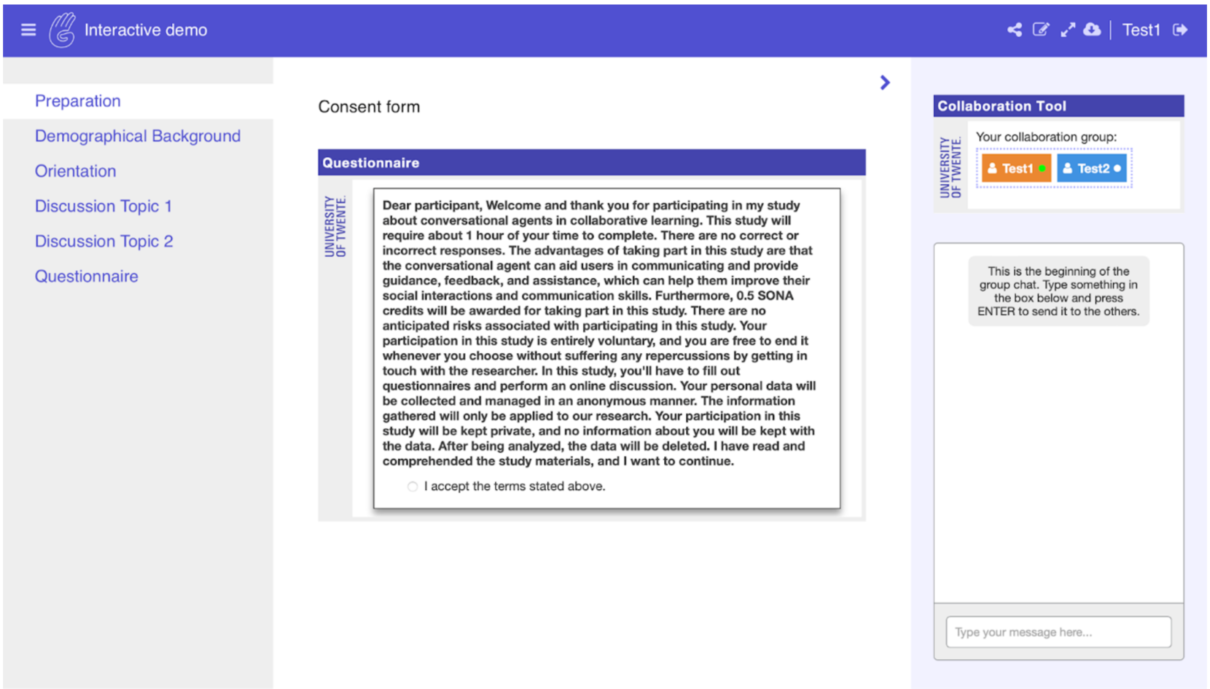
Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M., & Drachsler, H. (2021). Are We There Yet? - A Systematic Literature Review on Chatbots in Education. *Frontiers in Artificial Intelligence*, *4*. https://doi.org/10.3389/frai.2021.654924

Appendix

Appendix A - Graasp Environment



Appendix B – Consent Form



Consent form

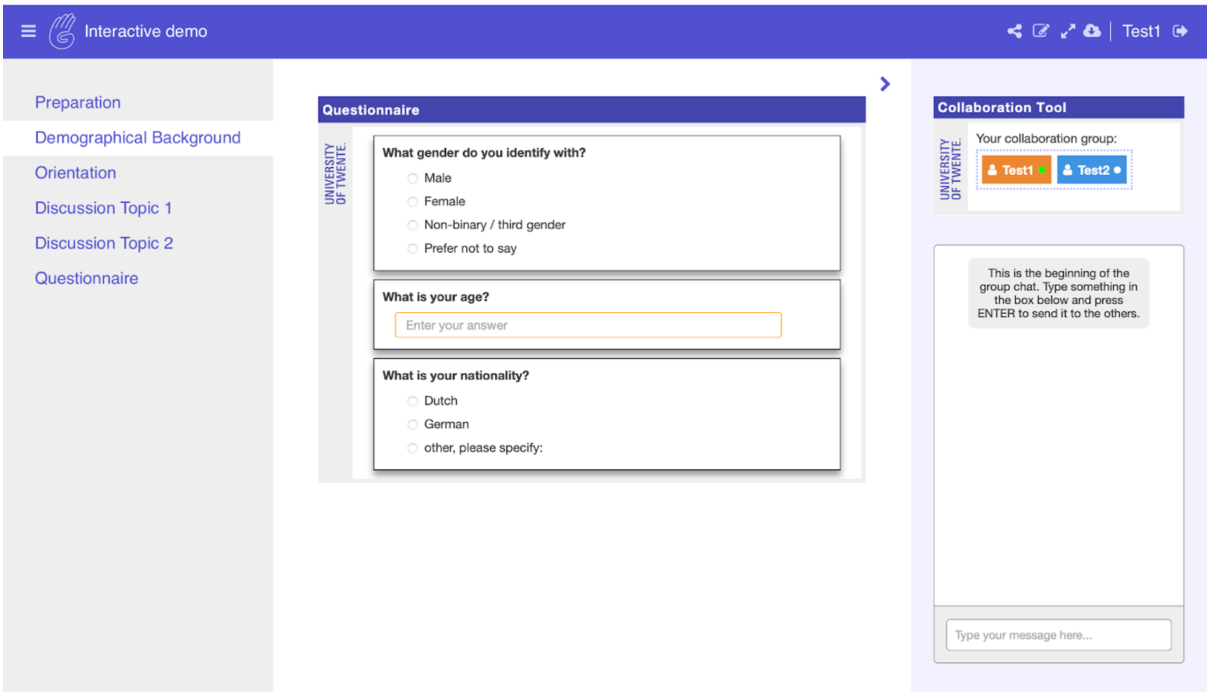
Questionnaire

Dear participant, Welcome and thank you for participating in my study about conversational agents in collaborative learning. This study will require about 1 hour of your time to complete. There are no correct or incorrect responses. The advantages of taking part in this study are that the conversational agent can aid users in communicating and provide guidance, feedback, and assistance, which can help them improve their social interactions and communication skills. Furthermore, 0.5 SONA credits will be awarded for taking part in this study. There are no anticipated risks associated with participating in this study. Your participation in this study is entirely voluntary, and you are free to end it whenever you choose without suffering any repercussions by getting in touch with the researcher. In this study, you'll have to fill out questionnaires and perform an online discussion. Your personal data will be collected and managed in an anonymous manner. The information gathered will only be applied to our research. Your participation in this study will be kept private, and no information about you will be kept with the data. After being analyzed, the data will be deleted. I have read and comprehended the study materials, and I want to continue.

○ I accept the terms stated above.

This is the beginning of the group chat. Type something in the box below and press ENTER to send it to the others.

Appendix C - Demographical Background



Questionnaire

What gender do you identify with?

○ Male
○ Female
○ Non-binary / third gender
○ Prefer not to say

What is your age?

Enter your answer

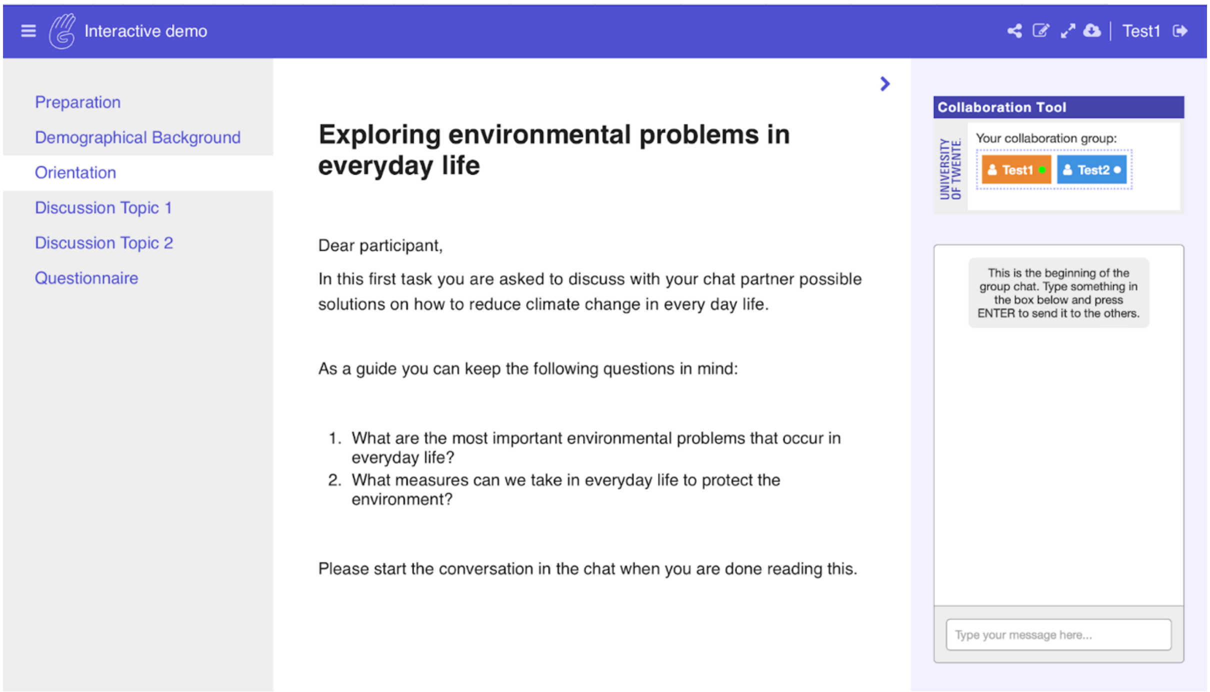
What is your nationality?

○ Dutch
○ German
○ other, please specify:

This is the beginning of the group chat. Type something in the box below and press ENTER to send it to the others.

Type your message here...

Appendix D – Orientation Task



Exploring environmental problems in everyday life

Dear participant,

In this first task you are asked to discuss with your chat partner possible solutions on how to reduce climate change in every day life.

As a guide you can keep the following questions in mind:

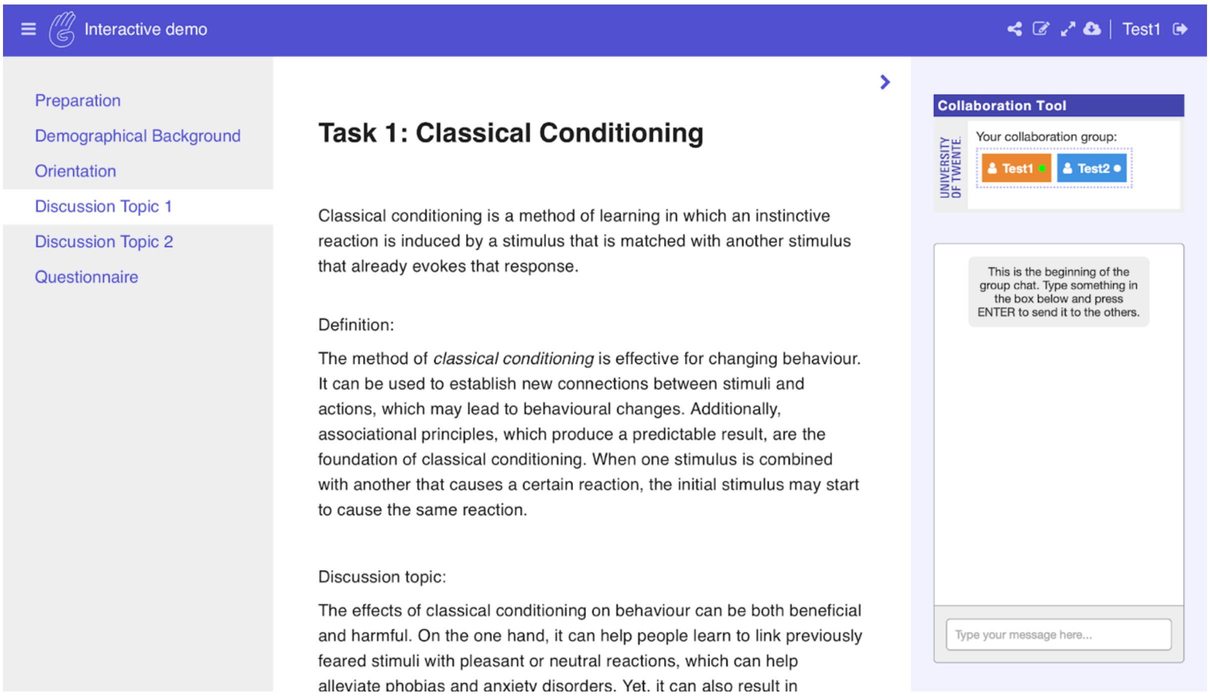
1. What are the most important environmental problems that occur in everyday life?
2. What measures can we take in everyday life to protect the environment?

Please start the conversation in the chat when you are done reading this.

This is the beginning of the group chat. Type something in the box below and press ENTER to send it to the others.

Type your message here...

Appendix E – Discussion Topic 1



This is the beginning of the group chat. Type something in the box below and press ENTER to send it to the others.

Task 1: Classical Conditioning

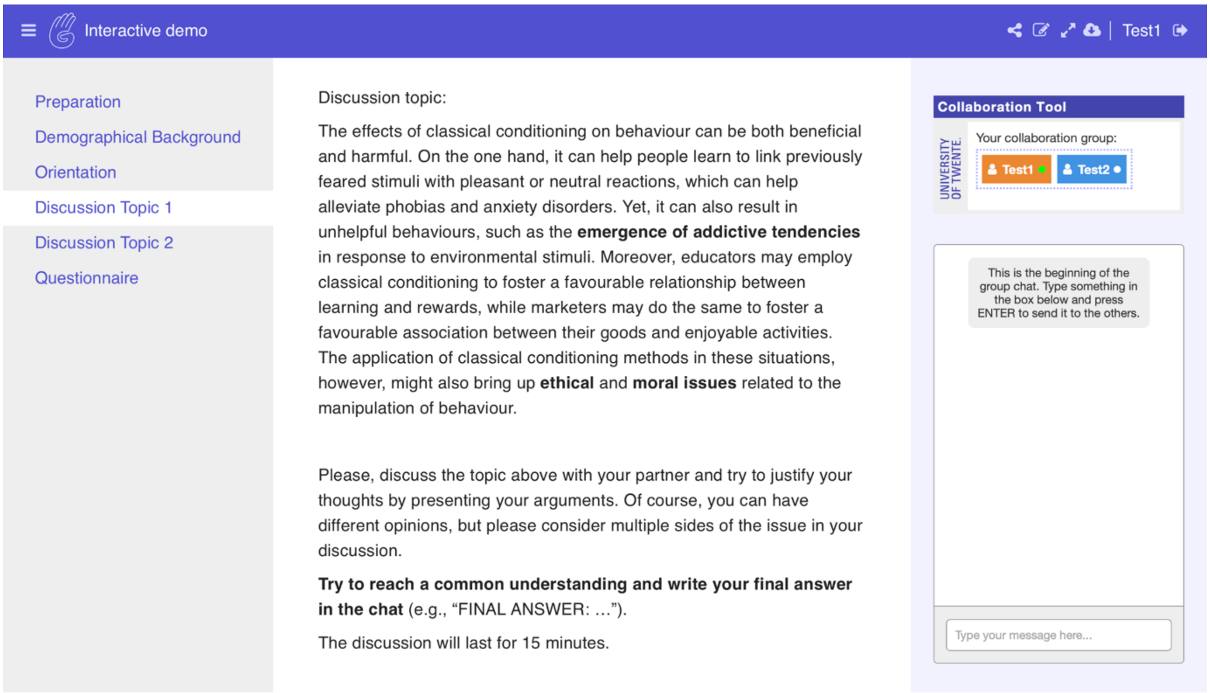
Classical conditioning is a method of learning in which an instinctive reaction is induced by a stimulus that is matched with another stimulus that already evokes that response.

Definition:

The method of *classical conditioning* is effective for changing behaviour. It can be used to establish new connections between stimuli and actions, which may lead to behavioural changes. Additionally, associational principles, which produce a predictable result, are the foundation of classical conditioning. When one stimulus is combined with another that causes a certain reaction, the initial stimulus may start to cause the same reaction.

Discussion topic:

The effects of classical conditioning on behaviour can be both beneficial and harmful. On the one hand, it can help people learn to link previously feared stimuli with pleasant or neutral reactions, which can help alleviate phobias and anxiety disorders. Yet, it can also result in



This is the beginning of the group chat. Type something in the box below and press ENTER to send it to the others.

Discussion topic:

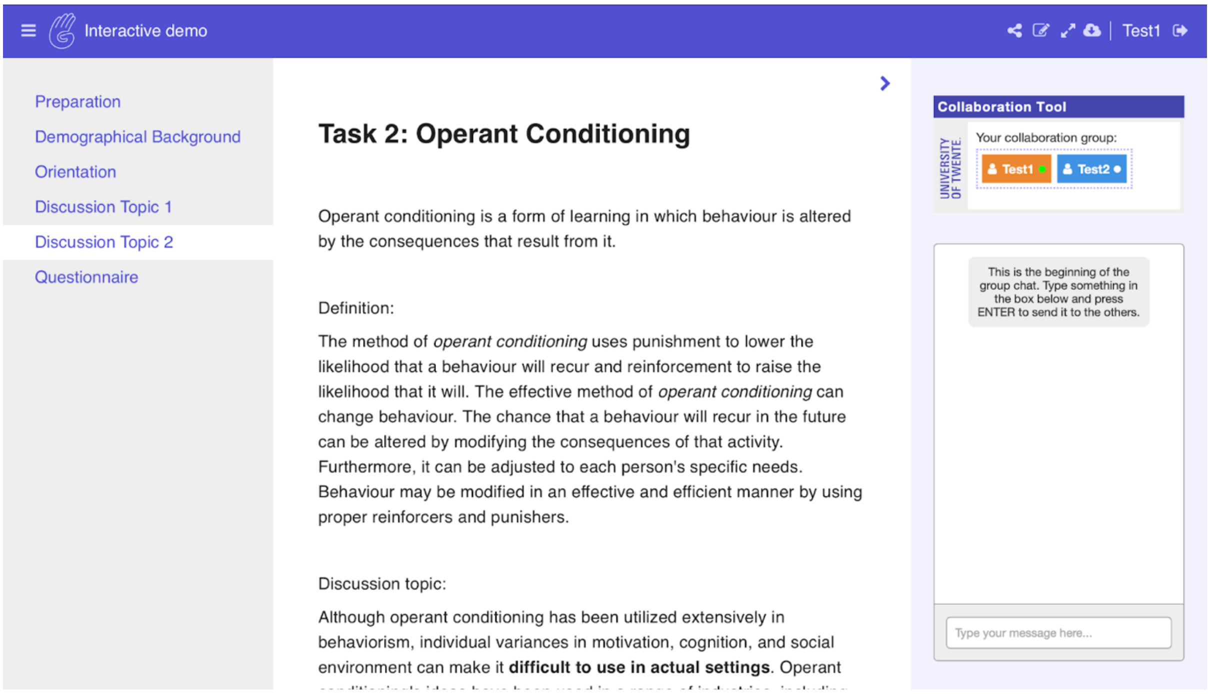
The effects of classical conditioning on behaviour can be both beneficial and harmful. On the one hand, it can help people learn to link previously feared stimuli with pleasant or neutral reactions, which can help alleviate phobias and anxiety disorders. Yet, it can also result in unhelpful behaviours, such as the **emergence of addictive tendencies** in response to environmental stimuli. Moreover, educators may employ classical conditioning to foster a favourable relationship between learning and rewards, while marketers may do the same to foster a favourable association between their goods and enjoyable activities. The application of classical conditioning methods in these situations, however, might also bring up **ethical** and **moral issues** related to the manipulation of behaviour.

Please, discuss the topic above with your partner and try to justify your thoughts by presenting your arguments. Of course, you can have different opinions, but please consider multiple sides of the issue in your discussion.

Try to reach a common understanding and write your final answer in the chat (e.g., "FINAL ANSWER: …").

The discussion will last for 15 minutes.

Appendix F – Discussion Topic 2



Task 2: Operant Conditioning

Operant conditioning is a form of learning in which behaviour is altered by the consequences that result from it.

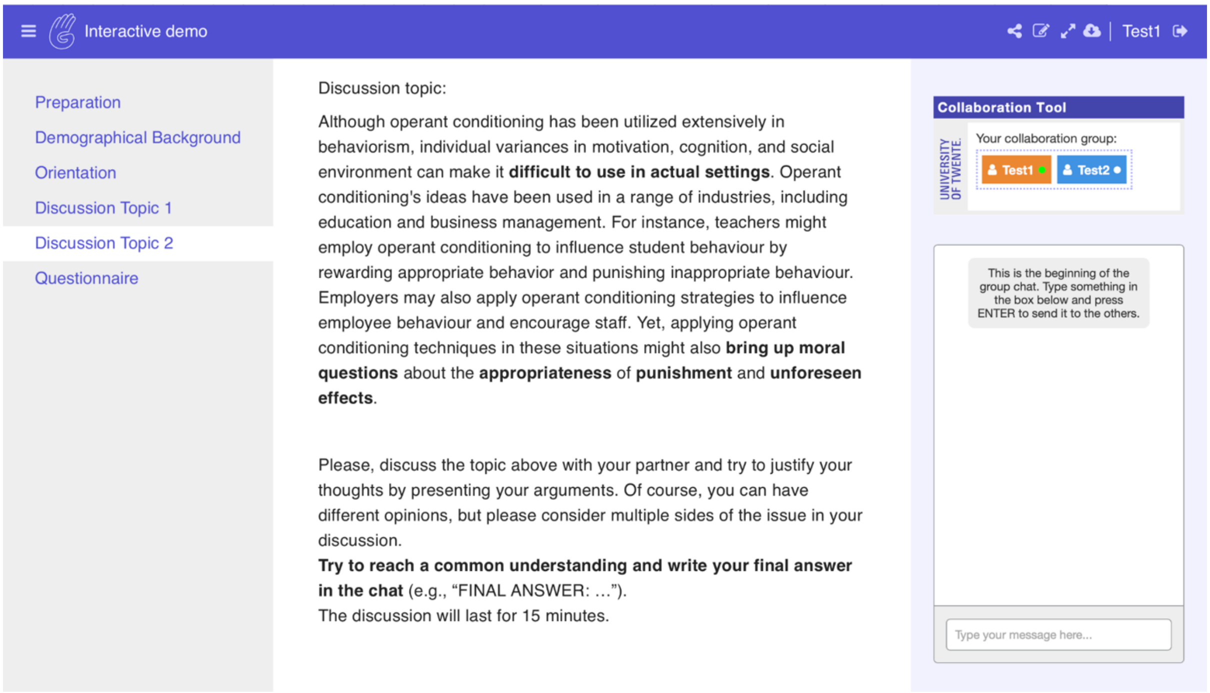
Definition:

The method of *operant conditioning* uses punishment to lower the likelihood that a behaviour will recur and reinforcement to raise the likelihood that it will. The effective method of *operant conditioning* can change behaviour. The chance that a behaviour will recur in the future can be altered by modifying the consequences of that activity. Furthermore, it can be adjusted to each person's specific needs. Behaviour may be modified in an effective and efficient manner by using proper reinforcers and punishers.

Discussion topic:

Although operant conditioning has been utilized extensively in behaviorism, individual variances in motivation, cognition, and social environment can make it **difficult to use in actual settings**. Operant

This is the beginning of the group chat. Type something in the box below and press ENTER to send it to the others.



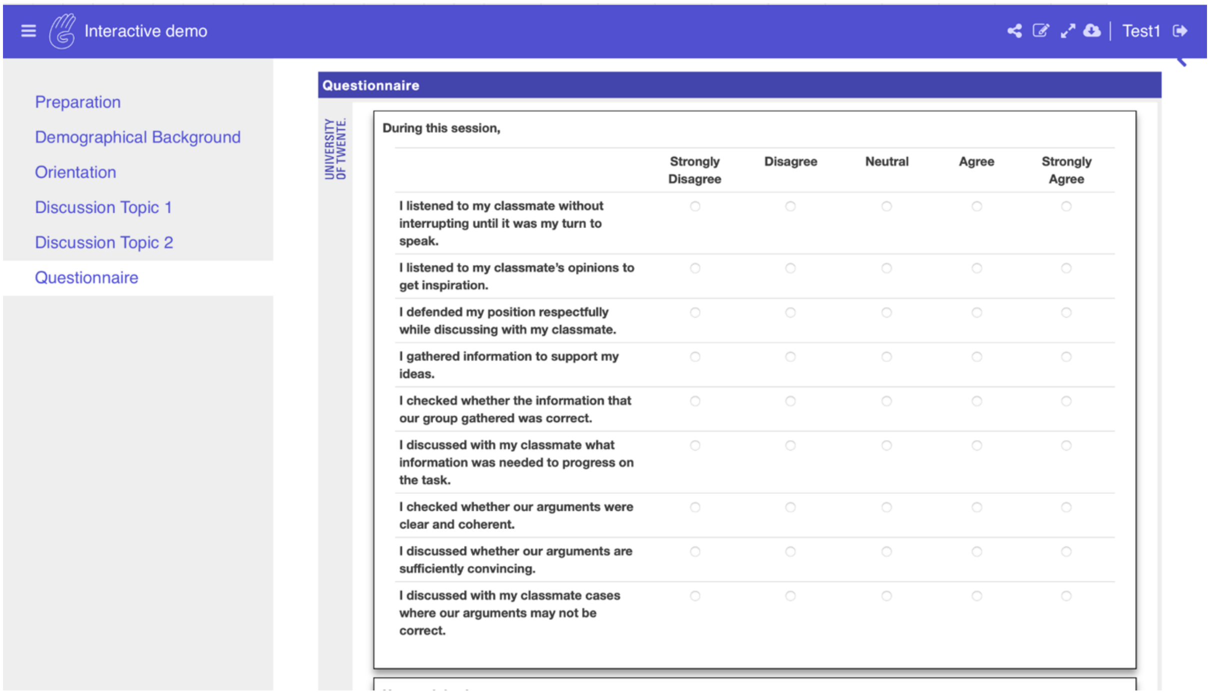
Discussion topic:

Although operant conditioning has been utilized extensively in behaviorism, individual variances in motivation, cognition, and social environment can make it **difficult to use in actual settings**. Operant conditioning's ideas have been used in a range of industries, including education and business management. For instance, teachers might employ operant conditioning to influence student behaviour by rewarding appropriate behavior and punishing inappropriate behaviour. Employers may also apply operant conditioning strategies to influence employee behaviour and encourage staff. Yet, applying operant conditioning techniques in these situations might also **bring up moral questions** about the **appropriateness** of **punishment** and **unforeseen effects**.

Please, discuss the topic above with your partner and try to justify your thoughts by presenting your arguments. Of course, you can have different opinions, but please consider multiple sides of the issue in your discussion.
Try to reach a common understanding and write your final answer in the chat (e.g., “FINAL ANSWER: …”).
The discussion will last for 15 minutes.

This is the beginning of the group chat. Type something in the box below and press ENTER to send it to the others.

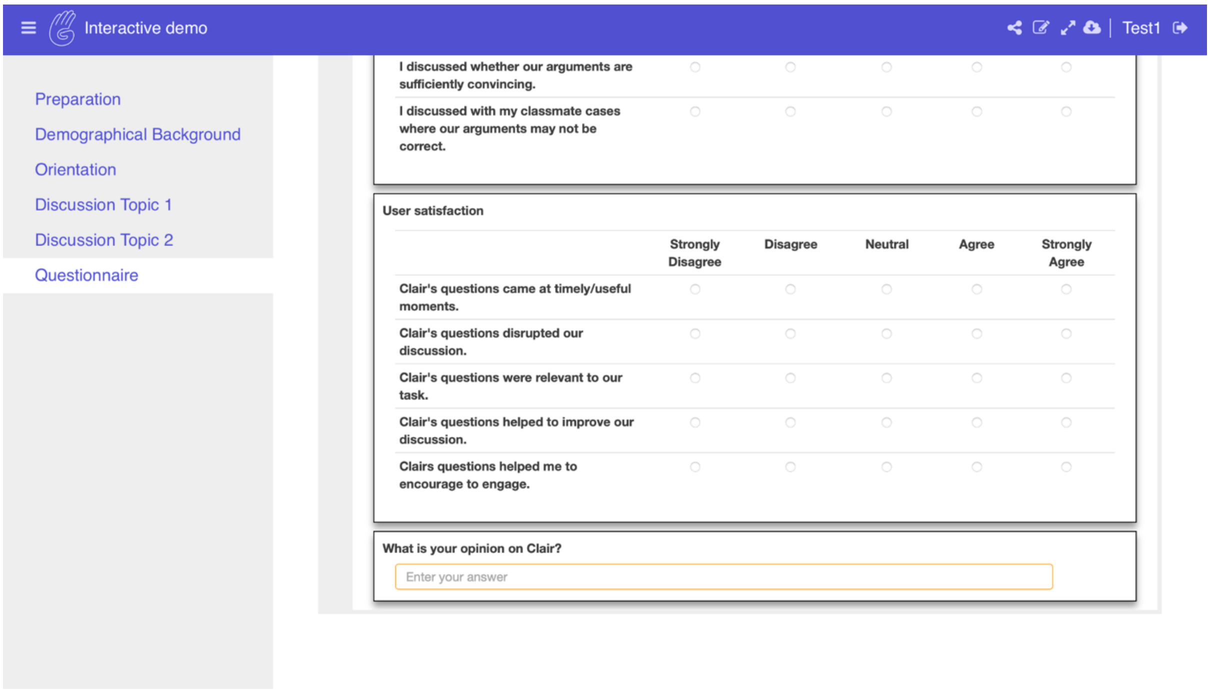
Appendix G – Questionnaire



Questionnaire

During this session,

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I listened to my classmate without interrupting until it was my turn to speak.	○	○	○	○	○
I listened to my classmate's opinions to get inspiration.	○	○	○	○	○
I defended my position respectfully while discussing with my classmate.	○	○	○	○	○
I gathered information to support my ideas.	○	○	○	○	○
I checked whether the information that our group gathered was correct.	○	○	○	○	○
I discussed with my classmate what information was needed to progress on the task.	○	○	○	○	○
I checked whether our arguments were clear and coherent.	○	○	○	○	○
I discussed whether our arguments are sufficiently convincing.	○	○	○	○	○
I discussed with my classmate cases where our arguments may not be correct.	○	○	○	○	○



I discussed whether our arguments are sufficiently convincing.	○	○	○	○	○
I discussed with my classmate cases where our arguments may not be correct.	○	○	○	○	○

User satisfaction

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Clair's questions came at timely/useful moments.	○	○	○	○	○
Clair's questions disrupted our discussion.	○	○	○	○	○
Clair's questions were relevant to our task.	○	○	○	○	○
Clair's questions helped to improve our discussion.	○	○	○	○	○
Clairs questions helped me to encourage to engage.	○	○	○	○	○

What is your opinion on Clair?

Enter your answer