

Evaluation of interventions in blended learning using a communication skills serious game

Raja Lala¹, Gemma Corbalan¹, and Johan Jeuring^{1,2}

¹Utrecht University, Netherlands

²Faculty of Management, Science and Technology, Open University Netherlands

Abstract. Serious games often employ a scripted dialogue for player interaction with a virtual character. In our serious game Communicate, a domain expert develops a structured, scripted scenario as a sequence of potential interactions in an authoring tool. A player is often a student learning communication skills and a virtual character represents a person that a student talks to. In the original version of Communicate, a player ‘converses’ with a virtual character by clicking on one of the multiple statement options. Since 2018, we perform blended learning sessions for final year computer science students using Communicate. Our goal is to improve these sessions and in this paper, we apply the action research method over three semesters to iteratively improve these blended learning sessions. In the first semester, our baseline, we conduct sessions where students play a scenario in multiple choice format. In the second semester, we enhance Communicate by enabling a student to enter open text input in an improved scenario. In the third semester, we enhance a session by incorporating peer teaching. Students fill in an evaluation survey after a session and we compare the evaluation of students from the three semesters. Results show that student ratings are significantly higher in sessions incorporating peer teaching compared to the baseline.

1 Introduction

Most professions require communication skills, for example a doctor needs to communicate with a patient [17], while an IT expert discusses system requirements with a client. Generic communication skills such as conflict management, or being assertive are useful for many professionals. Communication skills are best learned through practice, in role play or with a simulated patient [1].

Communicate [4] is a serious game for practising communication skills. A communications expert/teacher scripts a scenario in an authoring tool that provides expressive constructs for variability in a dialogue [6]; see the screenshot at the left hand side in Fig. 1. In the original version of the game, a player navigates through a dialogue by selecting a statement option from one of the scripted player statements, see the screenshot at the right hand side in Fig. 1. This is similar to a multiple choice question answer format.



Fig. 1. Communication skills scenario simulation

In this paper, we apply the action research method [8] to teach collaboration skills in a blended learning session to final year bachelor computer science students at Utrecht University. We aim to improve these sessions. Lewin [8] describes action research as: ‘a spiral of steps, each of which is composed of a circle of planning, action and fact-finding about the result of the action’. In our case, a step is a set of blended learning sessions that we teach per semester. We collect student evaluation in these sessions. At the end of the semester, we critically reflect on the sessions and also analyse the student evaluations to identify potential improvements. We then introduce improvement(s) to the blended learning sessions in the next semester, again collect student evaluations, and compare these to the previous semester.

Final year computer science students at Utrecht University need to work in a project team (of 10 to 12 students each) to develop a software product for a real client. Prior to 2018, we gave a teamwork lecture in a single classroom session to all students. Since the semester spring-summer 2018, we provide a blended learning session: a face-to-face workshop for collaboration skills using Communicate per student project team. We teach a student to handle a collaboration situation where another student in a team (represented as a virtual character in Communicate) does not follow agreed team quality measures (integration tests). In consultation with communication skills experts, we developed a scenario called Collaborate for our students, which uses a protocol consisting of the subject-phases: Approach, Express, Discuss and Agree next steps. We present our study from blended learning sessions with successive batches of final year bachelor computer science students from three semesters: spring-summer 2018, fall-winter 2018 and spring-summer 2019.

We start with sessions where a student plays the Collaborate scenario in a multiple choice manner (in spring-summer 2018). These sessions form our baseline case. In a session an instructor introduces Communicate, students play a scenario, the instructor explains the communication protocol that forms the basis of the scenario, and students play the scenario again. After playing, students and the instructor have a plenary discussion. Students fill in an evaluation form after each session. This form gathers student perceptions about some of the didactic aspects of Communicate. Dhaqane et al [2] find a correlation of student satisfaction with student performance in higher education. The evaluation form includes five statements on student satisfaction and self-efficacy, for example: I know better how to give relevant feedback, measured on a 5-point Likert scale, rating from 1 (completely disagree) to 5 (completely agree). Fig. 2 shows all

the questions. Additionally, the form includes two open questions: ‘What do you think about the game?’ and ‘Suggestions to improve the scenario?’

In addition to the evaluation of students, we also look at the application of activity theory to human computer interaction design [5] for potential improvements to our blended learning sessions. Multiple choice versus open text (test) questions are two common forms encountered in learning and education. Choosing from multiple options is similar to a multiple choice test, which often evokes low level cognitive processing, whereas open text response often requires complex thinking [9]. Ozuru et al [14] conduct a study to compare text comprehension from multiple choice and open text questions. Ozuru et al. find a positive correlation of the performance on open text questions with the quality of an explanation, and a positive correlation of the performance on multiple choice questions with the level of prior knowledge related to the text. In *Communicate*, a student reads and responds to a text from a virtual character, and gets a score on her performance in a scenario. This is a similar activity as in the study by Ozuru et al [14], where a student reads a short text while explaining preselected sentences. An open text response should lead a better quality of a response and possibly higher cognitive skills. In fall-winter 2018, we change the activity mechanism for a student to interact with the virtual character by typing open text input.

After fall-winter 2018, we evaluate the blended learning sessions again. A common feedback from students is the wish to play more scenarios. For the activity mechanism we change the interaction between students within a blended learning session to encourage insight in peer behaviour. Goldschmid et al [3] review peer teaching (when a student teaches another student) in higher education. The authors find that peer teaching among students can enhance active participation, develop skills in cooperation and interaction when used in conjunction with other teaching and learning methods. In spring-summer 2019 we incorporate peer teaching in our blended learning sessions by asking students to play *Communicate* in pairs and explain their motivation of a statement choice at a step of a scenario to each other. The interaction with a virtual character is reverted to the baseline case of spring-summer 2018, i.e. multiple-choice.

In this paper, we compare student evaluations from the three semesters. Our research question is: ‘How does student evaluation vary with different interventions (multiple choice, open text and peer teaching) in blended learning sessions using *Communicate*?’

This paper is organised as follows. Section 2 discusses related work. Section 3 presents the interventions in the blended learning sessions in more detail and compares the student evaluations from the three semesters, Section 4 discusses results and Section 5 presents conclusions and future work.

2 Related Work

Provoost et al [15] perform a scoping review of embodied conversational agents (ECAs) in clinical psychology. The authors find that an ECA has a positive effect

on user engagement and effectiveness of the interventions. However, Provoost et al. find only a limited number of evaluation and implementation studies of ECAs, in particular with larger sample sizes and control groups. The authors advocate a 'low tech' (simple) implementation to improve psychological interventions in the field.

Mazza et al [10] present a learning environment aimed at a student to learn how to interview a patient. A simulation in this environment includes a set of patient videos of interconnected doctor visits. A professional actor plays a virtual patient and simulates moods, attitudes and responses. The authors present a component that matches an open text speech input to a set of available choices. Eight students tested the free speech simulation and were in general positive about this component. However, the students noted problems with input for which no, or an erroneous, match was found. In the former case the software asks a student to rephrase her input, and in the latter case the software matches an input to an incorrect event. To address the problems, the authors propose to extend the set of available choices for a student. This seems to involve hiring the same actor to act responses to the extended set of choices, which might be cumbersome.

Van der Lubbe et al [18] develop a virtual training environment to teach a player (especially older adults) about situations where a potential swindler tries to gain trust. The author present a prototype consisting of six scenarios. In a scenario, a player interacts with a virtual character by either clicking on an option from multiple choice options or by speaking aloud an option. A speech module detects the level of assertiveness in case of a spoken response. The authors conduct an evaluation with a focus group of five security advisors. The focus group was in general positive about the prototype and expressed desire for more scenarios, being able to go back within a scenario, alter an answer and navigate to the tips/feedback menu.

Ochs et al [12] develop a virtual reality (VR) based simulation to train a doctor to break bad news. An ECA 'acts' as a virtual patient and follows a scenario. A player's (doctor's) speech input is interpreted real time by a human operator who selects a semantic match to one of the available 136 prototypical sentences. These prototypical sentences are based on a previously transcribed corpus of doctor interactions with trained actors playing a patient. The matched sentence is sent to a dialogue system that generates a verbal and non-verbal response of the ECA. Evaluation of the VR experience shows an positive impact of the environment display on the sense of presence, sense of co-presence, and believability of the virtual patient [13].

The work presented in this paper differs from the research described above: we incorporate multiple interventions (based on feedback from students) using the same learning environment over three semesters with successive student groups. We also compare the evaluation from the students over the three semesters.

3 Method and results

Action research [8] consists of cycles of taking action, and investigating the effects of the action. A step is a cycle of planning, action and fact finding. Our final year computer science students work in a software project as part of their curriculum towards the end of their bachelor program. Within this course, we provide a workshop session per team to address collaboration skills. Our goal is to improve these sessions and an action step-cycle is described in the Introduction (Section 1). Students from each semester fill in the same survey after a session, covering the following questions: five communication learning questions (see Fig. 2) with options on a 5-point Likert scale, and two open evaluation questions (*1. What did you think of the game? 2. Suggestions to improve the scenario?*). We analyse this evaluation and introduce improvements to the blended learning sessions. In this section, we describe the blended learning sessions in each semester in more detail in cycles of action and results.

3.1 Baseline multiple choice sessions of spring-summer 2018

In our first version of Communicate (**first intervention**), a player navigates through a simulation and converses with a virtual character by clicking a statement option from one of the prescribed player statements. This is similar to a multiple choice question answer format; see the right hand side of the screenshot of our learning environment in Fig. 1.

In spring-summer 2018, we organised blended learning sessions where students played the Collaborate scenario. A student interacted with a virtual character by choosing an option from the multiple choice statements at a step of the Collaborate scenario. There were a total of 82 students assigned in eight project teams of 10 to 12 students each. After a project team session, students filled in the evaluation form. A total of 75 students filled in the form. Fig. 2 and Table 1 show the student Likert scale ratings on the five learning questions (label SS18 for spring-summer 2018 intervention).

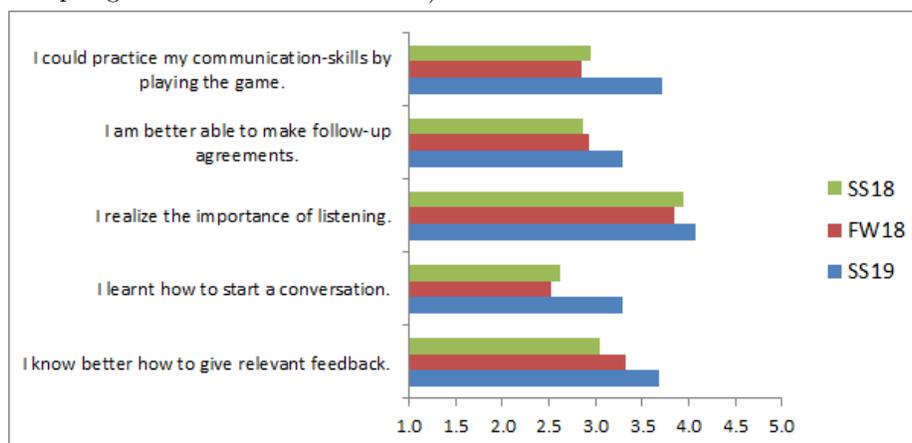


Fig. 2. Chart representation of student rating on the five learning questions

In the feedback to the two open evaluation questions, students expressed the wish to input open text and had reservations about limited statement choices. To address this feedback, we conducted sessions in the same semester to gather open text input from students [16]. Gathering open text input also enabled us to create a dataset to develop and test several NLP methods [16], and to improve the Collaborate scenario. An improved scenario in the game is a teaching material improvement [11]. We applied a clustering algorithm to the list containing students’ open text and the predefined statement options for each node in our scenario [7]. Two experts analysed the clustered input and improved the Collaborate scenario in multiple ways. First, if two predefined options at a step of a scenario were contained in the same cluster, then one of them was removed, since they were too similar when compared to the variety of student open text responses. Second, if a number of student open text responses formed a cluster that did not include any predefined scripted statement options, we added a new statement option (similar to the open text responses in the cluster), at that step of the scenario. Third, we modified a response from the virtual character to better frame a context and provide step feedback at some nodes of the scenario. We tested various NLP methods on the dataset created from open text input in spring-summer 2018, but the results were not entirely satisfying [16]. We choose the best matching NLP match method for use in the session of fall-winter 2018.

Another common feedback from the two open evaluation questions was that the VC was seemingly dumb and some students playing the scenario did not want to help / collaborate with the VC. Students also commented about the texture and gestures of the virtual character. We developed a new virtual character with better texture, gestures and animation. We simplified the scenario dialogue to give the VC a ‘happy go lucky’ character: an engaged and social team member, who works somewhat irregularly.

3.2 Intervention open text input, fall-winter 2018

In the blended learning session of fall-winter 2018 (**second intervention** abbreviated to FW18), we incorporate open text in Communicate. A student enters an open text at a step of the scenario Collaborate and the NLP method attempts to match this input to one of the scripted statement choices at that step of Collaborate. Since the NLP match method is not entirely accurate, we scaffold a match result. Communicate highlights the best match when an open text matches to at least one scripted statement and gives a sequence of hints when an open text does not match with any scripted statements at a step. In total 52 students were assigned to five project teams. After playing in open text sessions, 40 students filled in the evaluation form. The evaluation of the students in the five process questions is shown in Fig. 2 and Table 1 (label FW18 for fall-winter 2018 intervention).

The baseline is the evaluation from spring-summer 2018 (abbreviated to SS18) when the students played the Collaborate scenario (Collaborate) in multiple choice format. We compare the (Likert) student ratings on the five learning questions in Table 1 to the student ratings baseline spring-summer 2018. The

first column in Table 1 displays the feedback Question that students rated on a 5-point Likert Scale rating from 1 (completely disagree) to 5 (completely agree). The second column denotes the semester and the third column shows the number of students who filled in the evaluation form. The fourth column shows the average rating of the students of the particular question and the fifth column shows the standard deviation to the mean score. In the fifth column, the p-value indicates the likelihood that an increase (or decrease) in rating is due to chance. A p -value <0.05 is significant, and a smaller p -value indicates a high likelihood that an increase (or decrease) in rating is due to an intervention.

Question	Intervention	N	Mean	Std. Dev.	p -value
I could practice my communication skills by playing the game.	SS18	75	2.95	1.089	
	FW18	40	2.85	1.027	0.639
	SS19	78	3.72	0.952	<0.000
I am better able to make follow up agreements.	SS18	75	2.87	1.082	
	FW18	40	2.93	0.971	0.769
	SS19	78	3.29	0.899	0.009
I realize the importance of listening.	SS18	75	3.95	1.138	
	FW18	40	3.85	1.075	0.654
	SS19	78	4.08	0.879	0.431
I learnt how to start a conversation.	SS18	75	2.61	1.051	
	FW18	40	2.53	0.960	0.651
	SS19	78	3.29	0.955	<0.000
I know better how to give relevant feedback.	SS18	75	3.05	0.985	
	FW18	40	3.33	0.917	0.144
	SS19	78	3.68	0.830	<0.000

Table 1. Student ratings on the five communication process propositions

Despite the extra scaffolding steps there was no significant deterioration (nor improvement) in the five learning questions ratings compared to spring-summer 2018. In the feedback to the two open questions, students were positive about the sessions but had reservations about the matching and expressed a wish to play more scenarios.

3.3 Intervention students peer teach and play multiple scenarios, spring-summer 2019

In spring-summer 2019 in our **third intervention** (abbreviated to SS19), we incorporate peer teaching and multiple scenarios in the blended learning sessions. To increase the number of scenarios, we requested communication skills teachers from other faculties to share scenarios and received scenarios on breaking bad news, self reflection, giving feedback etc. In these SS19 sessions, an instructor introduces Communicate, and students play multiple scenarios; at least *Collaborate* and one other scenario (e.g. breaking bad news, giving feedback etc). Students play the scenarios in multiple choice mode. The instructor explains the communication protocol of *Collaborate*: Approach, Express, Discuss and Agree. Thereafter, students play a scenario in pairs of two, where a student explains

her statement choice to respond to a virtual character to the other student. We incorporate this action with the goal to provide insight into peer-behaviour, to teach each other and improve interaction in a session [3].

In total 81 students were assigned to eight project teams of which 78 students filled in the evaluation form after the sessions. In SS19 there is a significant improvement in four of the five learning questions ratings compared to the baseline, see Table 1. The fifth question: *‘I realize the importance of listening.’* receives high ratings in all three semesters. The scenario seems to elicit the importance of listening in general. In the feedback to the two open questions, some students were ‘pleasantly surprised’ by the quality of the game and virtual character, some found that the game needed no further improvements, some wished for more feedback in *Collaborate*. We argue that the improved student evaluation is directly related to the third intervention.

4 Discussion

In the student evaluation, student ratings on the Likert scale questions show no significant difference in fall-winter 2018 sessions incorporating open text input versus the baseline of multiple choice input of SS18. In the response to the evaluation form open questions, students made remarks about the quality of an NLP match and that could be a reason that student ratings are not significantly higher. Students also complained about typing multiple times due to the extra scaffolding steps (highlighting and hints) and that could be another reason for no significant rating increase.

Results show that the ratings of the spring-summer 2019 sessions incorporating peer teaching and multiple scenarios are significantly higher for four of the five questions compared to the baseline of spring-summer 2018, see Section 3.3. Application of activity theory to human computer interaction design [5] often focusses on interaction. Our results suggest that while interaction (e.g. multiple choice versus open text input) is important, it is crucial to investigate blended learning sessions with respect to pedagogical aspects such as, in our case, peer teaching.

5 Conclusions and future work

In this paper we plan, implement and evaluate interventions in our blended learning sessions using Communicate. We start with sessions using multiple choices in Communicate, enhance Communicate by enabling a student to enter open text input in an improved scenario with an improved virtual character and finally incorporate peer teaching and playing multiple scenarios in a session. Results show student ratings of open text input sessions do not significantly differ from multiple choice and that ratings of sessions incorporating peer teaching and multiple scenarios are significantly higher compared to multiple choice sessions.

For future work, we plan to have blended learning sessions again with multiple scenarios and peer teaching. As a difference to spring-summer 2019, we plan to

have scenarios in both open text (in Collaborate scenario) and multiple choice (breaking bad news, giving feedback etc.) input modes. In these sessions, for open text input we plan to match with minimal scaffolding, where with matched input a simulation continues as if a virtual character has understood the input (i.e. no extra highlight step) and with unmatched input, we present the available statement options (i.e. no hint step). After playing, a debrief step will be introduced to ask the students in a session to reflect on a recent experience and have a plenary discussion on what went well and what could be improved. After guided sessions with a new batch of students, we can collect student feedback from the sessions.

6 Acknowledgments

This activity has partially received funding from the European Institute of Innovation and Technology (EIT). This body of the European Union receives support from the European Unions Horizon 2020 research and innovation programme. This research was also partially supported by the 644187 EC H2020 RAGE project.

This work has been developed in collaboration with the Software and Game-project of the bachelor Computer Science in the faculty of Science, Utrecht University. The authors also acknowledge Majanne Wolters, Marjan van den Akker and Michiel Hulsbergen for their help in teaching and scenario improvement; and to Jordy van Dortmund and Marcell van Geest for implementing the desired changes to Communicate.

References

1. Marianne Berkhof, H. Jolanda van Rijssen, Antonius J.M. Schellart, Johannes R. Anema, and Allard J. van der Beek. Effective training strategies for teaching communication skills to physicians: An overview of systematic reviews. *Patient Education and Counseling*, 84(2):152–162, 2011.
2. Mahad Khalif Dhaqane and Nor Abdulle Afrah. Satisfaction of students and academic performance in benadir university. *Journal of Education and Practice*, 7(24):59–63, 2016.
3. Barbara Goldschmid and Marcel L Goldschmid. Peer teaching in higher education: A review. *Higher Education*, 5(1):9–33, 1976.
4. Johan Jeuring, Frans Grosfeld, Bastiaan Heeren, Michiel Hulsbergen, Richta IJntema, Vincent Jonker, Nicole Mastenbroek, Maarten van der Smagt, Frank Wijmans, Majanne Wolters, and Henk van Zeijts. Communicate! — a serious game for communication skills —. In *Proceedings EC-TEL 2015: Design for Teaching and Learning in a Networked World: 10th European Conference on Technology Enhanced Learning*, volume 9307 of *LNCS*, pages 513–517. Springer, 2015.
5. Victor Kaptelinin and Bonnie A Nardi. *Acting with technology: Activity theory and interaction design*. MIT press, 2006.
6. Raja Lala, Johan Jeuring, Jordy van Dortmund, and Marcell van Geest. Scenarios in virtual learning environments for one-to-one communication skills training. *International Journal of Educational Technology in Higher Education*, 14(1):17, May 2017.

7. Raja Lala, Marcell van Geest, Stefan Ruseti, Johan Jeuring, Mihai Dascalu, Jordy van Dortmund, Gabriel Gutu-Robu, and Michiel Hulsbergen. Enhancing free-text interactions in a communication skills learning environment. In *Proceedings of the 13th international conference on Computer supported collaborative learning*, June, 2019.
8. Kurt Lewin. Action research and minority problems. *Journal of social issues*, 2(4):34–46, 1946.
9. Michael E Martinez. Cognition and the question of test item format. *Educational Psychologist*, 34(4):207–218, 1999.
10. R. Mazza, L. Ambrosini, N. Catenazzi, S. Vanini, D. Tuggener, and G. Tavarnesi. Behavioural simulator for professional training based on natural language interaction. In *EDULEARN18 Proceedings*, 10th International Conference on Education and New Learning Technologies, pages 3204–3214. IATED, 2-4 July, 2018 2018.
11. Nienke Nieveen and Elvira Folmer. Formative evaluation in educational design research. *Design Research*, 153:152–169, 2013.
12. Magalie Ochs, Philippe Blache, and Grégoire De Montcheuil. What common ground between a human and a virtual agent? the case of task-oriented dialogues for breaking bad news. In *22nd Workshop on the Semantics and Pragmatics of Dialogue (SemDial 2018)*, 2018.
13. Magalie Ochs, Daniel Mestre, Grégoire De Montcheuil, Jean-Marie Pergandi, Jorane Saubesty, Evelyne Lombardo, Daniel Francon, and Philippe Blache. Training doctors social skills to break bad news: evaluation of the impact of virtual environment displays on the sense of presence. *Journal on Multimodal User Interfaces*, 13(1):41–51, 2019.
14. Yasuhiro Ozuru, Stephen Briner, Christopher A Kurby, and Danielle S McNamara. Comparing comprehension measured by multiple-choice and open-ended questions. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 67(3):215, 2013.
15. Simon Provoost, Ho Ming Lau, Jeroen Ruwaard, and Heleen Riper. Embodied conversational agents in clinical psychology: a scoping review. *Journal of medical Internet research*, 19(5):e151, 2017.
16. Stefan Ruseti, Raja Lala, Gabriel Gutu-Robu, Mihai Dascalu, Johan Jeuring, and Marcell van Geest. Semantic matching of open texts to pre-scripted answers in dialogue-based learning. In *Proceedings of the 20th International Conference on Artificial Intelligence in Education*, June, 2019.
17. Jonathan Silverman, Suzanne Kurtz, and Juliet Draper. *Skills For Communicating With Patients, Third Edition*. Radcliffe Publishing Limited, 2013.
18. Laura M van der Lubbe, Charlotte Gerritsen, Daniel Formolo, Marco Otte, and Tibor Bosse. A serious game for training verbal resilience to doorstep scams. In *International Conference on Games and Learning Alliance*, pages 110–120. Springer, 2018.
19. Wim Westera, Baltasar Fernandez-Manjon, Rui Prada, Kam Star, Andrea Molinari, Dominic Heutelbeck, Paul Hollins, Rubén Riestra, Krassen Stefanov, and Eric Kluijfhout. The rage software portal: Toward a serious game technologies marketplace. In *International Conference on Games and Learning Alliance*, pages 277–286. Springer, 2018.