Social Robots in Collaborative Learning: Consequences of the Design on Students' Perception

Alix Gonnot, Christine Michel, Jean-Charles Marty, Amélie Cordier

Social Robots in Collaborative Learning: Consequences of the Design on Students' Perception

Alix Gonnot[1], Christine Michel[1], Jean-Charles Marty[2] and Amélie Cordier [3]
[1]*INSA de Lyon, Univ Lyon, CNRS, LIRIS, UMR 5205, F-69621 Villeurbanne, France*
[2]*Université de Savoie Mont-Blanc, CNRS, LIRIS, UMR 5205, F-69621 Villeurbanne, France*
[3]*Hoomano, Lyon, France*

Keywords: Collaborative Learning, Social Robots, Human-Robot Interaction

Abstract: The interest in using social robots in education is growing as it appears that they could add a social dimension that enhances learning. However, there is little use of robotics in collaborative learning contexts. This shows a lack of knowledge about students' perception of social robots and their use for education purposes. This paper aims to fill this gap by analyzing, with experimental methods: (1) the influence of specific ways of interaction (facial expressions, voice and text) on the students' perception of the robot and, (2) students' acceptability criteria for using robots in a classroom. The target objective is to help the design of future learning situations. The study shows that the ways used to interact produce significant differences in the perception of the animation, the likeability, the attractiveness, the safety and the usability of the robot. The study also shows that major improvements must be made on the design of the hedonic characteristics of the interactions, especially identification and stimulation, to favor the student's acceptance of this kind of learning support tools.

1 INTRODUCTION

Robots are increasingly used in education where they can serve many purposes (e.g. tools to learn how to develop software, telepresence) and more recently, social robots offer a new opportunity for supporting people in various tasks. We believe that social robots can improve collaborative learning since they can bring back the social dimension, essential for learning, into the standard tools used in Computer Supported Collaborative Learning (CSCL).

1.1 Collaborative Learning

Dillenbourg (1999) defines collaborative learning as a type of learning where two or more people try to learn something together. This form of learning is known for being beneficial on various levels such as academic, social and psychological (Laal and Ghodsi, 2012).

Jermann *et al.* (2004) exposed two different approaches to support collaborative learning. The structuring approach supports collaborative learning by carefully designing the activity beforehand. The regulating approach consists in mediating directly the interactions within the group. Mirroring tools, metacognitive tools and guiding systems can be used to accomplish all or part of this process. However, a lot of work remains to be done in order to support the teacher in an effective way. The proposed numeric tools are sometimes inadequate since they do not provide the necessary "human interaction" that is central for group regulation.

1.2 Social Robots

With the development of robotic technologies, the use of robots inside classrooms is spreading. They are currently mainly used as a new teaching tools (Church et al., 2010) or as telepresence tools (Tanaka et al., 2014).

Social robots are humanoid, autonomous robots that provide users with "social" abilities such as communicating, cooperating or learning from people. They usually communicate with humans through channels that are typically dedicated to the communication with other humans (voice, gesture, emotions, etc.). Most of the time, people anthropomorphize them and assume they have a form of social intelligence or social skills (Breazeal, 2003).

The social aspects of social robots can be interesting in educational contexts. Social behaviour is an essential, albeit tacit, component of the student-teacher interaction (Kennedy et al., 2017). Robots are presented as able to add embodiment and a social dimension to the learning activity (Mubin et al., 2013). Saerbeck *et al.* (2010) also claim that *"employing social supportive behavior increases learning efficiency of students"*. Social characteristics of robots can thus constitute a progress when compared with standard tools. This statement leads us to consider robots as potential regulating tools as those presented in Section 1.1.

1.3 Social Robotics in Collaborative Learning

Strohkorb *et al.* (2016) conducted an experiment to find out if a social robot could influence the collaboration between two children that were playing the same game. Mitnik *et al.* (2008) proposed a framework for a mediator robot and experimented on it. Short and Matáric proposed a model formalizing a moderation process (2015) and a moderation algorithm (2017) that are designed to be used with social robots. The human-robot interactions used in the various experiments presented above are only one-way interactions, usually robot to human. There are no communications between the students and the robot. Programming robots to interact with humans is a demanding technical challenge and today, a human cannot interact with a social robot at the same level of complexity they could interact with a human.

Furthermore acceptance problems arise when social robots are used in class. According to Sharkey (2016), social robots may generate privacy issues because robots record and/or use information about their environment and users. Moreover, Sharkey asks if it is moral or safe to let humans think that a machine is clever and to take the risk that they become attached to it. It also seems problematic to put a robot in charge of a classroom without human supervision. However, it seems "*that the attitude towards social robots in schools is cautious, but potentially accepting*" (Kennedy et al., 2016).

We are inclined to think that using specific means of communication on the robot (displaying text, using its voice, etc.) will influence the students' perception of the device. The aim of the paper is to find out which means of communication used by the robot are the most adequate considering several dimensions. This is a prerequisite for setting up collaborative learning scenarios that use social robots. In this study we choose to focus only on the students' perception of the social robot during the interaction (ethical problems related to the use of this technology in the classroom are not considered).

In order to address those questions, we designed an experiment that is described in the Method section of this paper. We then expose the results and the implications for our project in the Results and Discussion sections.

2 METHOD

The goal of this experiment is to better understand the students' perception of a social robot guiding them through a learning task and to determine whether this perception is influenced by the use of certain functionalities of social robots, namely text-to-speech and the ability to express emotions.

During the experiment, the participants are divided into groups of three and are guided by the robot through a collaborative learning task. The robot gives instructions to the groups using different interaction modalities. The participants are then asked to fill out a questionnaire about their perception of the robot.

2.1 Task

The collaborative task proposed to the participants is a lesson on the use of decision matrix, a decision-making method designed to evaluate several options by comparing them on a finite number of criteria.

The robot instructs the participants to read a document explaining what a decision matrix is and how to use it. It then guides them through an exercise were they are supposed to advise a fictional company on the selection of new smartphones for its employees. The robot gives step by step instructions to the participants to make them fill and use the decision matrix to take the decision. Once an instruction given by the robot is completed by the group, the participants touch the screen of the robot to get the next instruction.

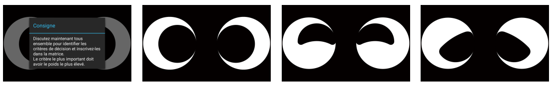
Twice during the process, the robot gives feedback to the participants before giving the next instruction. The first piece of feedback is given when the participants have chosen the criteria used to evaluate the smartphones and the second piece of feedback is given after the selection of the smartphone accordingly to the data presented in the decision matrix. Each time, if the participants provide the right answer, the robot gives a positive feedback and if the participants provide the wrong answer, the robot gives a negative feedback and provides them with an indication to help pinpoint the error. The participants are then given a chance to correct their answer and if they are still wrong, the correct answer is offered to them.

2.2 Robot

The robot we used for this experiment is a prototype running on Android. This robot is a non-humanoid robot equipped with motorized wheels. A tablet is placed on its head and is usually used to display its eyes. The robot is able to deliver information in various ways such as displaying text on its screen or speaking. It is also able to display several facial expressions representing various emotions.

We combined those functionalities to build three different behaviors. In the first behavior, the robot shows a neutral expression all the time and instructions are given to participants in text form through a dialog box displayed on the robot screen on top of its eyes. The instructions are displayed for 30 seconds on the screen. In the second behavior, the robot shows a neutral expression all the time and instructions are pronounced out loud for the participants. In the third behavior, instructions are also pronounced out loud and a neutral expression is shown most of the time. When the robot gives feedback to participants however, a joyful facial expression is shown when the feedback is positive and a sad facial expression is shown when the feedback is negative. The voice used in the second and third behavior is the default Android text-to-speech voice.

Those three behaviors constitute our three conditions for the experiment: Instructions given in text form (C1), Instructions given with a neutral voice (C2) and Instructions given with a neutral voice and facial expressions (C3).



(a) Text (b) Neutral (c) Joyful (d) Sad

Figure 1: Elements displayed on the robot's screen during the experiment.

We believe that the participants' perception of the robot will vary if the robot uses its "social" functionalities such as talking or showing facial expressions. Our hypothesis can be formulated as: *Participants will have a better perception of the robot when it gives them instructions vocally with facial expressions rather than with voice or text only (H1).*

For the robot to deliver instructions and feedback in a way that adapts to the participants progression and errors, we chose to use the Wizard of Oz technique (Kelley, 1984) to control the robot. A human operator is physically present in the room and makes the robot react appropriately to the unfolding events. For example, if the participants completed an instruction that had not been given yet, the operator would make the robot skip this instruction. The operator also decides if the answers provided by the participants are correct and chooses the error message delivered by the robot if needed. The participants were not informed that the robot was entirely controlled by the operator.

2.3 Questionnaire

At the end of the task, the participants are asked to fill out a questionnaire meant to understand how the participants of the study perceive the robot. The questionnaire is divided into two different parts.

In the first part, we chose to use the Godspeed questionnaire (Bartneck et al., 2009) that is a standard questionnaire designed to measure user's perception of a robot. As the Godspeed questionnaire does not question the user about the usability of the robot and the user experience, we decided to also use the AttrakDiff questionnaire (Hassenzahl et al., 2003). Since our participants are all students in France or working in France, we chose to use French translations of the Godspeed and the AttrakDiff questionnaires to increase participants' understanding of the questions. For the Godspeed questionnaire, we chose to offer our own translation, based on the one already proposed on Bartneck's website. For the AttrakDiff questionnaire, we used the official translation (Lallemand et al., 2015).

The second part of the questionnaire contains three open-ended questions: "What could we do to improve the robot?"(Q1), "Would you be ready to use such a robot in class (or in a more general learning situation)? Why?" (Q2) and "Do you have other remarks?" (Q3).

2.4 Analysis

The first part of the questionnaire consists of two semantic differential scale questionnaires: Godspeed and AttrakDiff. As recommended by the authors (Bartneck et al., 2009) mean scores were computed for each scale of the Godspeed questionnaire. Five dimension are then analyzed: Anthropomorphism, Animation, Likeability, Perceived Intelligence and Perceived Safety. The same processing was applied to the AttrakDiff questionnaire (Hassenzahl et al., 2003) for each of its scale. Four dimensions are then analyzed: Pragmatic Quality, Hedonic Quality-Stimulation, Hedonic Quality-Identification, Attractiveness. The ANOVA method was then used to determine the influence of the conditions C1, C2, C3 on these nine dimensions.

The second part of the questionnaire contains open-ended questions. We carried out a thematic analysis (Braun and Clarke, 2006) on the participants' answers for questions Q1 Q2 and Q3. This analysis highlighted two main types of themes: participants' expectations about the robot (feedback, emotions, dynamism) and robot's means of interaction (voice, movement, content). Another layer of analysis was added for question Q2 in order to determine the valence of the answers. We identified three kinds of answers: "Yes", "No", "Yes, if improvements".

2.5 Participants

21 persons participated to the experiment in total. The participants were divided into groups of three and each group was associated to one of the three conditions described in Section 2.2. One group (3 participants) was removed from the study because its members did not all fill out the final questionnaire.

The remaining participants were equally distributed between the three conditions and the proportions of male and female participants were equivalent. The vast majority of participants are engineering school students. Two participants are PhD students and one is a young design engineer. Their ages range from 20 to 26 years old.

3 RESULTS

3.1 Semantic Differential Scale Questionnaires

In order to detect variations between the three conditions in the Godspeed and AttrakDiff questionnaires results, we used the ANOVA method. We performed the tests on each scale of the two questionnaires to determine whether the mean scores for each condition were significantly different.

3.1.1 Godspeed Questionnaire

As shown on Table 1, significant differences exists between the three conditions for the Animation, Likeability and Perceived Safety indicators of the Godspeed questionnaire. No significant differences were found for the Anthropomorphism and Perceived Intelligence scales. It means that the interactions involving speaking or expressing emotions we used during the experiment influenced the perception of the animation, likeability and perceived safety but did not influence the way the robot's is anthropomorphized by the participants or how intelligent it seems to be. It then seems that the anthropomorphic characteristics attributed to our robot are mainly due to its physical form, that stay the same in all conditions, and it could explain why we did not detect any variations for the Anthropomorphism scale. In the same way, we can guess that the perceived intelligence is directly tied to the material delivered by the robot in the experiment and no variations were detected on the Perceived Intelligence because that material remains identical in the three conditions.

Table 1: p-value and significance of the ANOVA test ($p < 0.05$) for the Godspeed questionnaire.

Variable	p	Sig
Anthropomorphism	0.308	No
Animation	0.010	Yes
Likeability	0.024	Yes
Perceived Intelligence	0.075	No
Perceived Safety	0.040	Yes

The mean scores (see Figure 2) for the condition C3 are higher than those for condition C2 on the Animation, Likeability and Perceived Safety variables, meaning that the participants like the robot better and have the perception of a better animation and safety when the robot uses facial expressions when it speaks. Similar results are observed when the robot speaks (conditions C2 and C3) rather than when it only displays text on its screen (condition C1). Our hypothesis is thus partially validated for the animation, likeability and perceived safety criteria. Participants have a better perception of these characteristics when the robot gives them instructions vocally with facial expressions rather than with voice or text only.

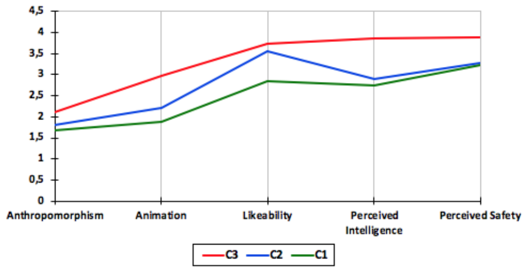


Figure 2: Mean scores of the Godspeed questionnaire in the three conditions.

3.1.2 AttrakDiff Questionnaire

As shown on Table 2, significant differences exists between the three conditions for the Pragmatic Quality and Attractiveness scales of the AttrakDiff questionnaire. No significant differences were found for the hedonic variables, meaning that using the robot was perceived as equally pleasant in all three conditions.

Table 2: p-value and significance of the ANOVA test ($p < 0.05$) for the AttrakDiff questionnaire.

Variable	p	Sig
Pragmatic Quality	0.020	Yes
Hedonic Quality Stimulation	0.243	No
Hedonic Quality Identification	0.137	No
Attractiveness	0.038	Yes

The data presented on Figure 3 show that the mean scores for the condition C3 are higher than those for condition C2 for the Pragmatic Quality and Attractiveness variables. The mean scores are also higher for condition C2 than for condition C1 for the two variables.

This means that the participants find the robot to be significantly more usable and more attractive when it gives them instructions vocally with facial expressions rather than with voice or text only.

We can confirm that our hypothesis is partially validated for the Pragmatic Quality and Attractiveness variables.

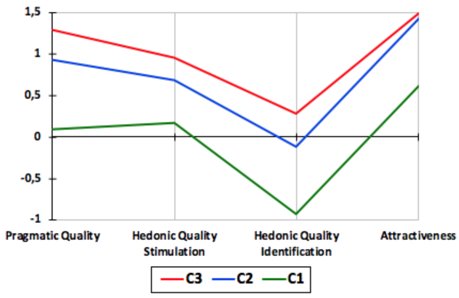


Figure 3: Mean scores of the AttrakDiff questionnaire in the three conditions.

3.2 Open-ended Questions

3.2.1 Statements About the Experiment

Table 3 presents the results of the thematic analysis that was performed on the participant's answers to the open-ended questions. 76 different items were identified and distributed into 11 themes. The different items are equally distributed between conditions C1, C2 and C3 (respectively 26, 22, 28). The most represented themes in those 76 items are the feedback (20), the voice (20), the movement (11) and the interactions (10). Items are equally distributed in our three conditions for each of those themes.

Items related to *feedback* are similar throughout our three conditions. Participants want the cognitive abilities of the robot to be more developed so that it could be able to guide them, answer questions or provide custom explanations.

Table 3: Themes identified in the open-ended questions answers.

Theme	C1	C2	C3	All
Feedback	7	5	8	20
Voice	9	4	7	20
Movement	4	3	4	11
Interaction	2	3	5	10
User emotion	3	1	1	5
Material/content	0	2	0	2
Class dynamism	0	2	0	2
Robot emotion	0	0	2	2
Naturalness	0	2	0	2
Ease of use	1	0	0	1
Presence	0	0	1	0
All	26	22	28	76

Expectations formulated regarding the *voice* are different when they come from participants of condition C1 or from participants of conditions C2 and C3. The C1 participants express their regret that the robot does not speak and that it is not possible to interact vocally with it. The C2 and C3 participants would also like to interact vocally with the robot but they wish for a voice more "realistic, natural, pleasant" and less "robotic" as well.

When considering the *movement* theme, we can note that the items express the same ideas in all conditions. Participants would like for the robot to produce head movements or moves in order to make the class more dynamic but also to intervene in the discussion, to address a student in particular and to be more "present".

Items associated to the *interaction* theme were mainly proposed by participants of condition C3. They think specific interactions such as the blinking of the robot's eyes are interesting and regret that there is not more of those.

Items related to the *user emotion* theme are mostly positive. Participants of all conditions indicated that they appreciated this type of learning activity that is fun and new. Participants of conditions C2 and C3 even pointed out that the class was more dynamic when conducted this way, however they think that more content should be provided.

Finally, only C2 and C3 participants expressed some expectations regarding the *anthropomorphization of the robot*, asking for more emotion, presence and naturalness from the robot.

3.2.2 Are Students Ready to Welcome Social Robots in Class?

As stated in Section 2.4, the answers given by the participants to the question Q2 were classified in three categories: ”Yes”, ”No” and ”Yes, if improvements”.

The answers qualified as ”Yes” or ”No” were clearly positive or negative such as *”Yes, absolutely (saving time rather than calling the teacher)”* or *”No, because it may not be able to answer to my questions (...)”*.

In the third category, ”Yes, if improvements” we find answers that were potentially accepting, under the condition that improvements were made on the robot. An example of answers classified in this category is *”No, having to read on a screen makes it as interesting a chat bot on a computer. However, if it was more ”interactive”, why not!”*.

Table 4: Distribution of the answers into the three categories for each condition.

Category	C1	C2	C3	All
Yes	1	2	0	3
Yes (improv.)	2	2	1	5
No	3	2	5	10
All	6	6	6	18

The Table 4 presents the distribution of the answers into the three categories. We can count 8 positive answers (”Yes” and ”Yes, if improvements”) out of 18 and 10 negative answers (”No”). With those numbers, it is difficult to discern a tendency from the participants to accept or reject the use of robots in class, even if the scale tips a little on the negative side.

Arguments in favor of the use of robots in class were that it could be interesting to use it to help the teacher. A participant said that using the robot could save time rather than calling the teacher. It also could make the class more dynamic, be motivating and finally, a participant even mentioned that adding the robot was like adding a nice person to the group. Overall, it seems that the robot could make the learning more enjoyable.

When examining the arguments against the use of the robot in class, it appears that the reason for rejecting the use of this technology in class could be that the robot is not sophisticated enough. Indeed, several participants point out that the robot is not interactive enough to be used in class, that its voice is unpleasant or that the robot is useless if it is only used to display text. Others express doubts about the robot’s abilities to answer questions, explain things and illustrate its explanations with examples taken from personal experience. The robot was also perceived as difficult to use or as a potential waste of time.

4 DISCUSSION

4.1 Threats to Validity

Although our study seems to yield interesting results and partially validates our hypothesis, it is important to note that our sample was rather small and constituted mostly of engineering school students. The lack of participants and diversity among them means that the results obtained are deeply bound to the context of the experiment and may not be generalizable as such. In order to confirm the insights yielded by our experiment, it is essential to conduct future studies with a greater number and variety of students.

4.2 Acceptability of the Robot in the Classroom

When counting the negative and positive answers given by the participants in question Q2, we noted that we got slightly more ”no” than ”yes” or ”yes, if improvement”. In the same way, the results of the questionnaires showed that anthropomorphism and animation (with respective average values of 1.86/5 and 2.36/5) were considered insufficient by the participants. In the Attrakdiff questionnaire, the low values of Hedonic Qualities of stimulation (HQ-S) show that the robot should be better used to support stimulation. The negative values of the Hedonic Qualities of identification (HQ-I) indicate that the participants develop no identification with the robot or the situation. Moreover, in the majority of the negative answers, the participants state that they would not use the robot in class because the robot is not interactive or intelligent enough to be useful for learning purpose. The conditions expressed in the ”Yes, if improvement” answers referred to the same arguments. This could led us to believe that most participants do not wish to use social robots in the classroom.

However, we can also note that the values of the likeability, perceived intelligence and perceived safety indicators are quite good (with respectively 3.38, 3.17 and 3.46/5). They show that the robot is well perceived. Similarly, the attractiveness is greater than 1 (1.43 for C2 and 1.50 for C3) as soon as the robot speaks. Pragmatic qualities are also greater than 1 (1.29 for C3) as soon as the robot expresses emotions. This means that it is considered pleasant, but also clear, controllable, effective and practical. The low values of hedonic qualities show that the design is not refined enough for learning purpose. Comparative analysis between conditions C1, C2 and C3 showed that the more the robot’s own features were developed, the more positive the experience was.

These results let us think that the behavior of the robot used for the experiment was too basic: the robot was only pronouncing or displaying instructions for a sequential task according to participants' progress. Experimentation did not involve robot mobility or customization of answers. Furthermore, the experiment did not include voice control.

Our belief is that if we were able to make the robot more intelligent, more interactive or if it was perceived as such by the students, even in the long run, most of them will agree to use it in class.

The following section provides the improvements that appear to be the most critical.

4.3 Improving the Robot

The most critical aspects to improve are features that promote hedonic qualities of identification and stimulation. These are the lowest, and for some of them negative, values of AttrakDiff. In addition, they will enable users to achieve the be-goals more satisfactorily, that is to say to find reasons why they will continue to find the robot interesting and stimulating for their own development. This aspect is fundamental in education. To implement these qualities, it will be necessary to maintain a high level of animation and interactivity to meet the expectations of the participants. Intelligence expectations are also very high.

4.3.1 Working on Hedonic Qualities of Stimulation

The HQ-S can be strengthened by features that make the robot original, creative and captivating. The anthropomorphic characteristics can be used to design interactions using the voice, the movements or the emotions that serve this objective.

Participants suggested improvements such as making the robot able to nod, move or address a specific person in the group. They suggested also to make the robot more natural, expressive, dynamic and enthusiastic. Very concrete suggestions on the matter were also provided, such as slow down the blinking animation to make it look more natural and to make the robot follow the users with its eyes.

4.3.2 Working on Hedonic Qualities of Identification

The HQ-I are stimulated by personalized interactions that reinforce the professional/realistic aspects of the situation or the link with others.

Individual feedback functionalities can serve this purpose by stimulating access to appropriate information or content, assessment of activities or advice for task completion. Some participants proposed to make the robot more intelligent, for example by providing more instructions, talking more, making the students able to understand the mistakes they made and help them. It was also suggested that the robot ask them questions about their progress in the task. The robot could also be able to take initiatives and intervene in the users' discussion, directly by arguing on the content or by calling a specific student by his first name.

In a more general way, it could also be interesting to use the robot as a mediator in collaborative learning activities. For example, the robot could analyze students' work time or participation in order to regulate the collective work. It could play a specific role in a project or game learning situation and follow an adaptive scenario to improve learners' immersion and motivation.

This experiment is a first step towards the introduction of new tools, the social robots, in learning situations and the emergence of new ways of teaching with technology.

5 CONCLUSION

This paper analyses the students' point of view on introducing social robots in collaborative learning environments.

The use of social robots in education is growing as it appears that they could add a social dimension that enhances learning. Some experiments have been engaged but the interactions used, especially the communication between the human and the robot, are not considered as natural and interactive enough. Moreover, the use of social robots may cause privacy and acceptance problems.

The aim of this paper is, on one hand, to test the influence of specific ways of interaction on the students' perception of a social robot. Our hypothesis is that the participants do have a better perception of the robot when it gives them instructions vocally with facial expressions rather than with voice or text only. On the other hand, we aim to identify student's acceptability criteria of a social robot in class in order to help the design of future learning situations.

We presented the results of a comparative study conducted with potential users interacting with the social robot Ijini in a collaborative problem solving learning situation, using different ways of communication for the robot.

The experimentation shows that significant differences where found on five variables: animation, likeability, attractiveness, safety and usability. No significant differences were found on anthropomorphism, perceived intelligence and hedonic quality.

The analysis of the participants' recommendations shows that they can potentially accept social robots in the classroom if we come up with a better design. The major improvements to be made are to support hedonic qualities of identification and stimulation. The stimulation goal could be achieved by using anthropomorphic characteristics such as voice, movement and expression of emotion in order to make the robot more interactive. The identification goal could be achieved with the intelligence and animation characteristics of the robot. They can be used to provide the students with personalized feedback or to play an adaptive role in collaborative situations.

In the next steps of our work we will implement the discovered critical improvements and conduct a new and larger study to confirm the insights that were exposed in this work. This next study will also provide us with the opportunity to explore the idea to use the social robot as a regulating tool for collaborative learning activities.

REFERENCES

Bartneck, C., Croft, E., and Kulic, D. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1(1):71–81.

Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3:77–101.

Breazeal, C. (2003). Toward sociable robots. *Robotics and Autonomous Systems*, 42(3):167–175.

Church, W. J., Ford, T., Perova, N., and Rogers, C. (2010). Physics with robotics - using lego mindstorms in high school education. In *AAAI Spring Symposium: Educational Robotics and Beyond*.

Dillenbourg, P. (1999). *What do you mean by collaborative learning?* Oxford: Elsevier.

Hassenzahl, M., Burmester, M., and Koller, F. (2003). *AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität*, pages 187–196. Vieweg+Teubner Verlag, Wiesbaden.

Jermann, P., Soller, A., and Lesgold, A. (2004). Computer software support for collaborative learning.

Kelley, J. F. (1984). An iterative design methodology for user-friendly natural language office information applications. *ACM Trans. Inf. Syst.*, 2(1):26–41.

Kennedy, J., Baxter, P., and Belpaeme, T. (2017). Nonverbal immediacy as a characterisation of social behaviour for human–robot interaction. *International Journal of Social Robotics*, 9(1):109–128.

Kennedy, J., Lemaignan, S., and Belpaeme, T. (2016). The cautious attitude of teachers towards social robots in schools. In *Proceedings of the 21th IEEE International Symposium in Robot and Human Interactive Communication, Workshop on Robots for Learning*.

Laal, M. and Ghodsi, S. M. (2012). Benefits of collaborative learning. *Procedia - Social and Behavioral Sciences*, 31:486–490.

Lallemand, C., Koenig, V., Gronier, G., and Martin, R. (2015). Création et validation d'une version française du questionnaire attrakdiff pour l'évaluation de l'expérience utilisateur des systèmes interactifs. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, 65(5):239 – 252.

Mitnik, R., Nussbaum, M., and Soto, A. (2008). An autonomous educational mobile robot mediator. *Autonomous Robots*, 25(4):367–382.

Mubin, O., Stevens, C. J., Shahid, S., Al Mahmud, A., and Dong, J.-J. (2013). A review of the applicability of robots in education. *Journal of Technology in Education and Learning*, 1:209–0015.

Saerbeck, M., Schut, T., Bartneck, C., and Janse, M. D. (2010). Expressive robots in education: varying the degree of social supportive behavior of a robotic tutor. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1613–1622. ACM.

Sharkey, A. J. C. (2016). Should we welcome robot teachers? *Ethics and Information Technology*, 18(4):283–297.

Short, E. and Mataric, M. (2015). Multi-party socially assistive robotics: Defining the moderator role.

Short, E. and Mataric, M. J. (2017). Robot moderation of a collaborative game: Towards socially assistive robotics in group interactions. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*.

Strohkorb, S., Fukuto, E., Warren, N., Taylor, C., Berry, B., and Scassellati, B. (2016). Improving human-human collaboration between children with a social robot. In *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*, pages 551–556. IEEE.

Tanaka, F., Takahashi, T., Matsuzoe, S., Tazawa, N., and Morita, M. (2014). Telepresence robot helps children in communicating with teachers who speak a different language. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*, pages 399–406. ACM Press.