



Designing a Learning Robot to Encourage Collaboration Between Children

Wouter Kaag^(✉), Mariët Theune, and Theo Huibers

University of Twente, Drienerlolaan 5, 7522 NB Enschede, The Netherlands

k.w.kaag@alumnus.utwente.nl,

{m.theune,t.w.c.huibers}@utwente.nl

Abstract. Collaboration is an important skill for children to learn. In this paper we present a small-scale study exploring how technology can be used to elicit collaboration between children. We developed a prototype of a tablet-based robot (‘surfacebot’) that tried to perform a specific task, while children acted as tutors by giving feedback on the surfacebot’s actions. The surfacebot used the feedback to improve its actions by means of reinforcement learning. A pilot study with an early prototype showed that children were engaged and provided consistent feedback to the surfacebot, but showed little collaboration. Instead they made individual decisions and took turns in providing feedback. Based on these observations we made several changes to the prototype that were meant to stimulate collaboration between the children. In our main study with the revised prototype, we measured collaboration using an annotation scheme based on a collaborative problem solving framework. The results suggest that the revisions of the prototype indeed led to more extensive collaborative behavior, with 4 of the 9 participating pairs of children establishing a division of roles that necessitated perspective taking and mutual exchange of information.

Keywords: Collaboration · Primary school children · Learning robot

1 Introduction

Collaboration, referred to as a 21st century skill [1, 5], is an important skill for children to learn since it relates to critical thinking, meta-cognition and motivation [14]. Improving a skill requires practice, and there might be a long-term education gain when children discover collaboration for themselves [2]. It has been shown that beneficial effects regarding learning and development, particularly in the early years or primary education, can occur when children work in pairs or small groups [25]. Furthermore, self-esteem and attitudes towards others are mentioned as beneficial outcomes of collaborative learning in the classroom [4, 21]. But collaboration can be challenging for young primary school children, as children below the age of 7 have not yet developed all of the cognitive skills

required for effective peer collaboration, such as recursive perspective taking [26]. Therefore, this study focused on using technology to foster collaboration between children.

To encourage collaboration between children, we devised an interactive technology concept with a robot as a teachable agent. Computer agents taught by children are known as teachable agents [3]. As our teachable agent we used a surfacebot, see Fig. 1. Surfacebots were originally developed as affordable, mobile and flexible robots to be used in collaborative storytelling activities on a non-digital tabletop [6]. A surfacebot has two parts: a tablet and a base with wheels, making it capable of movement, sound and visual representations. The tablet can be used as a character display [24] and as an interactive interface [7].



Fig. 1. The surfacebot as used in our study.

Our objective was to explore how an activity with a surfacebot can be designed to encourage collaboration between pairs of primary school children. We expected that primary school children could benefit from settings that encourage social interactions and collaboration, in particular children in the age of 5–7, because their collaboration skills start developing [2]. Our main research question was: *How can the capabilities of the surfacebot be utilized to create an engaging activity that effectively encourages collaboration between children?* In order to answer this question, we iteratively designed prototypes and tested them with children. We also developed an annotation scheme based on the framework for collaborative problem solving skills by Hesse et al. [11] for evaluating the level of collaboration between children during an activity.

Firstly, in Sect. 2 we describe several applications designed for collaborative activities. Section 3 describes our first concept to stimulate collaboration between children using the surfacebot. In the pilot study (Sect. 4), the degree of collaboration between children using this prototype is explored. The findings from the pilot study lead to a second version of the prototype (Sect. 5). In the main study (Sect. 6), we analyzed the level of collaboration using this new prototype. We end with a discussion of the results (Sect. 7) and our conclusions and recommendations (Sect. 8).

2 Related Work

Much work on collaborative technologies for children has been carried out in the context of creative applications such as collaborative storytelling for children [10]. Early examples are KidPad and the Klump: storytelling technologies that allowed children to work independently, but encouraged collaboration by providing added benefits of collaborative actions in terms of efficiency or fun [2]. A more extreme approach to collaboration was taken in the Story Table system, which forced children to work together by requiring multiple-user touch actions (performed by children simultaneously) for certain crucial operations [28]. In a non-storytelling context, Woodward et al. tried to encourage collaboration through role division in a digital tabletop game for children [27].

Some work has been done on collaborative storytelling for children with robots moving on a tabletop. An example is RoboTale, of which the main character is a robot similar to the surfacebot. RoboTale successfully stimulated collaboration among small groups of school children in the form of passing tangibles to each other (prompted by the size of the table and scarcity of resources) and discussing the plot of the story they created together [15]. Earlier work using the surfacebot as a teachable agent had it as the main character in a playful story-based activity, traveling around ‘in France’ (on the tabletop) and being taught French words by pairs of children [24]. In this activity, collaboration was enforced through a fixed role division, where one child was teaching the words to the surfacebot, and the other child was moving the surfacebot.

Other research on robots as teachable agents explored how children perceive and correct the handwriting of a robot. A study with 24 children (aged 7–8) acting individually as handwriting tutors of a Nao robot showed that these children paid attention to the learning of a robot and were capable of providing corrections using a slider or by demonstration [8]. For a survey of studies with robots as teachable agents for children, see [13].

A major inspiration for our work was Sophie’s Kitchen [22], an application featuring a virtual agent (Sophie) that learned from human feedback how to bake a cake, using a reinforcement learning algorithm called Q-learning. Experiments were carried out with different versions of Sophie’s Kitchen, to investigate how adult users wanted to teach the agent. The first experiment allowed users to provide feedback using a slider. The results showed that being able to guide the agent’s attention through feedback resulted in a faster learning interaction, compared to only providing feedback after the fact. In another experiment Sophie used gazing behavior to indicate which action she was about to take (‘transparency behavior’). This led to people providing guidance more often when it was required and less often when not. Another version of Sophie had so-called ‘undo’ behavior: retracting an action (if possible) after receiving negative feedback. This was shown to further improve the learning behavior of the agent.

Our work combines various aspects of these prior works, while also showing some important differences. Our teachable agent is like Sophie in that it uses Q-learning to learn from human feedback, but our users are pairs of children instead of individual adults. Like [8, 24] we use a robot as our teachable agent,

but our primary aim is not learning-by-teaching but stimulating collaboration between children through the teaching activity. Finally, like [2, 15] we try to nudge children into collaboration through the design of the activity, instead of forcing it like [28].

3 First Prototype

Collaboration involves a “mutual engagement of participants in a coordinated effort to solve a problem together” [20, p. 70] and is affected by the structure and design of a task [14]. We kept this in mind as we developed a first prototype of a collaborative activity with the surfacebot as a teachable agent. To encourage children to collaborate while teaching the surfacebot, we created a story that portrays the surfacebot as a bear called Ted (shown on the surfacebot’s tablet screen) who wants to get dressed to go outside, but needs the help of the children to find the right clothes as he does not know which clothes fit the weather.

The clothing items are spread across four locations on the tabletop, each representing a different room in the bear’s house and showing several associated items of clothing (printed on cards) associated to that location, see Fig. 2. The surfacebot moves around these locations and selects clothing items to wear. Children can provide feedback on the surfacebot’s actions using a slider. The idea is that the surfacebot learns from the feedback and adjusts its decision making based on it. At some point, it decides to ‘go outside’ indicating the end of a round. If the selected clothes do not match the weather, another round starts. These actions and learning ability of the surfacebot were simulated in the first prototype.

Collaboration is characterized by a symmetrical structure: a symmetry of goals, actions, knowledge and status [9]. Therefore the activity was designed to have a symmetrical structure to encourage collaboration. This was done by giving children a shared goal of assisting the surfacebot in completing its task, as well as equal knowledge and opportunities to interact with the surfacebot.

The prototype consisted of three parts, see Fig. 3. The first part is an application for the surfacebot’s tablet, further referred to as the ‘character display’. It displays a bear, the clothes it is wearing and the thoughts it has; see Figs. 2 and 3 (middle). Preceding an action, the bear would think about a piece of clothing, shown in a thought cloud, see Fig. 4(a). After 3 s, the thought disappeared and the bear could be seen wearing the item, see Fig. 4(b). The thought cloud was included as a ‘transparency behavior’ [22] of the surfacebot about its upcoming actions (clothes choices). It was used to give children the time to act and to pave the way for eventually implementing a form of undo behavior, which could lead to faster learning of the surfacebot [22].

The second part of the prototype was the ‘reward interface’, through which children could communicate feedback to the surfacebot; see Fig. 3 (left). It included two interactive elements: a slider and a ‘send’ button (see Fig. 6(a) for a more close-up view). The slider enabled children to communicate a degree of right or wrong. The idea was that the slider could stimulate negotiation, and



Fig. 2. Impression of the first prototype. The surfacebot with the character display is positioned next to one of the four locations with its associated clothes cards.

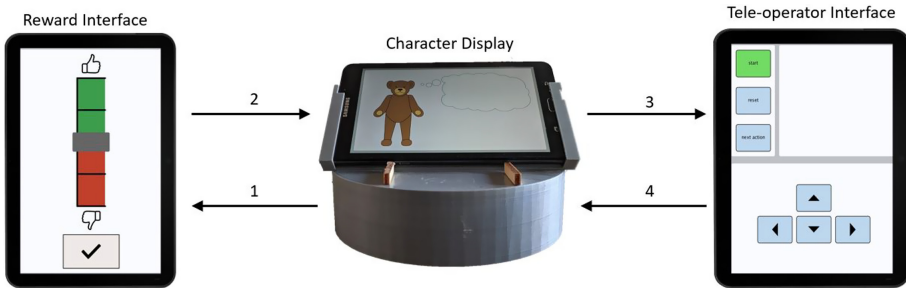


Fig. 3. The three main components of the prototype. The character display is the server, and communicates with the clients: the reward (left) and tele-operator (right) interfaces. The communication involves (1) status of the activity, (2) value and timing of feedback, (3) the name of next action according to the script and (4) controlling the activity (e.g. starting the next action) and controlling the surfacebot's movement.

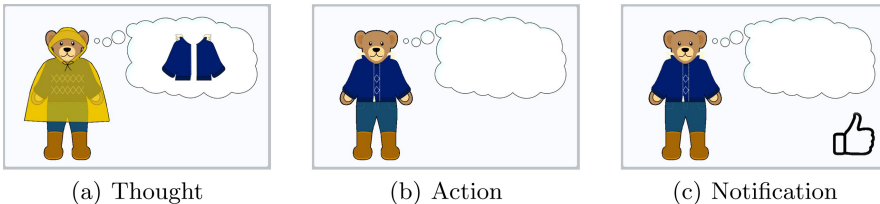


Fig. 4. Character display showing (a) what clothing item the bear is thinking of, (b) the bear wearing the item, and (c) a feedback notification (thumbs up or down).

could be used to reach a consensus when opinions differed, for example by going for an intermediate value. The send button was used to confirm the feedback. When children communicated feedback, a notification appeared on the character display, see Fig. 4(c).

The third part is the tele-operator interface; see Fig. 3 (right). In the first prototype, the surfacebot had no learning ability yet and could not move autonomously. In our tests with this prototype, a Wizard-of-Oz approach was used to make it seem like the robot acted and moved autonomously and that their feedback had an effect on the robot's choices. The tele-operator interface was used to control the activity and fake the autonomous behavior of the surfacebot, following a script with sequences of actions to simulate learning. Over the course of three rounds, the surfacebot's actions became increasingly more accurate and ultimately led to a set of clothes appropriate to the weather scenario.

4 Pilot Study

The first prototype was tested in a pilot study with 12 children at a local daycare facility. The study was approved by the Ethical Committee of our faculty and the parents of the children had given consent for their participation. The age of the children varied between 4 and 8 years old, with an average age of 5.75 years. With the study, we aimed to validate the concept and identify areas of improvement by getting a first impression of how children engaged, collaborated and provided feedback in the designed activity.

4.1 Method

In the pilot study, 6 pairs of children participated in sessions of ± 10 min. The tests were conducted in a separate room at the daycare facility. Figure 5 gives a top view impression of the setup. The procedure was as follows. First, the children received a short introduction in which the setup, the task and how the tablet worked was explained and demonstrated. Then they began their first round, giving feedback to the surfacebot as it was putting on clothes, until the surfacebot decided to 'go outside'. During the activity, the facilitator was always present to answer the children's questions. The children were not guided or motivated by the facilitator to provide feedback and could stop the activity at any time they liked. After three rounds, the children were thanked for their participation and helpfulness. They were made aware of the Wizard-of-Oz method (the facilitator controlling the surfacebot) used during the activity.

How the pairs of children acted and collaborated in the activity was observed during the test, and notes were taken afterwards. The sessions were recorded if consent for this was given by children's guardians. Tablet interactions were logged on the tablet to get an insight into the frequency and values of the children's feedback.

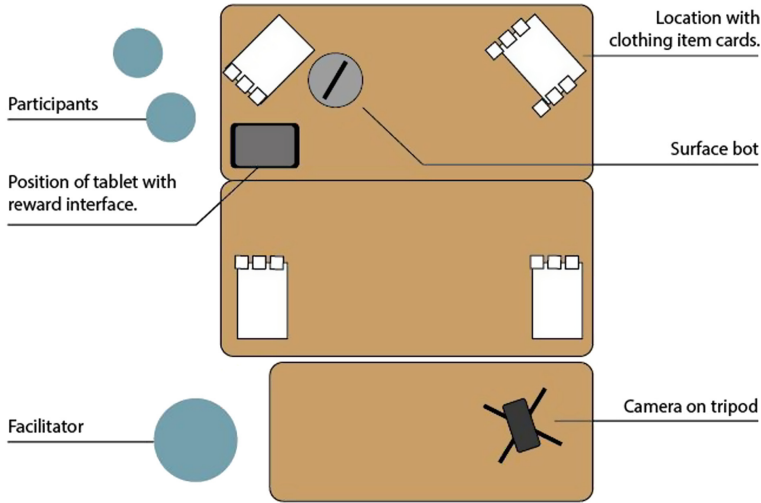


Fig. 5. Impression of the setup of the pilot study. The participants started the activity near the reward interface. The surfacebot moved between the locations at the corners of the table. The camera, on a tripod, was located on a table next to the activity. The facilitator also remained nearby for controlling the surfacebot.

4.2 Results

Limited collaboration between children was observed. Collaboration mostly took the form of providing feedback to the robot in turns. There was little conflict or negotiation about the task, but rather about who operated the tablet. Most children decided to walk around with the tablet, which created a situation where the activity was easily doable for one child. It allowed one child to track the robot's actions and communicate feedback at the same time. This may have provided little incentive for collaboration. This is confirmed by one pair of children who were observed to divide roles. During their session, the tablet remained on the table. It resulted in a situation where the children relied on each other to share information and opinions about the robot's actions. One followed and observed the robot while the other stayed close to the tablet and communicated feedback.

Children were observed to be motivated and enthusiastic during the activity. The logged data showed that the majority of the children continued to be engaged by consistently providing feedback, except for one pair who were both very young and shy (age 4 and 5). They had little communication and did not make use of the tablet. The logged data also showed that children used the reward interface in a binary way, with mainly extreme feedback values being communicated to the robot. It might be due to a certain unanimity that children had about which actions of the surfacebot were right and wrong. This made the slider somewhat superfluous.

The results correspond to the characteristics of the developmental stage of children aged 3 to 7 years, who are self-centered and prefer parallel play [16].

The majority of the children were engaged in the activity and provided feedback throughout the activity. This suggests that children could handle their role as tutor, and remained motivated during the activity. We concluded that ensuring a symmetrical structure [9] allows collaboration, but is not enough to elicit it.

5 Second Prototype

We modified the first prototype based on the insights obtained from the pilot in order to encourage collaboration more effectively.

Fixed Tablet. Firstly, an important change to the setup was made by fixing the tablet to the table. This would make observing the robot and giving feedback at the same time more difficult, so we expected this to stimulate collaboration in the form of a role division [27].

Improved Reward Interface. A new reward interface was created, see Fig. 6(b). The green-red gradient represents the transition from completely right to completely wrong. The aim was to clarify the use of the slider to encourage more diverse feedback.

Ambiguity. Ambiguous tasks tend to foster collaboration, as disagreements and misunderstandings can cause communication in the form of explanations and reasons [9]. Therefore, more ambiguity was added to the activity by providing more choices between similar items of clothing. Also, the names of the items shown on the cards were made more general to ensure they did not hint to a certain outfit or type of weather. For example, one item used in the first prototype had the description: “winter shoes”. This was changed to: “shoes”. It was expected that the less specific names could lead to more discussion among the children about the objects, which would require collaboration to provide unanimous feedback.

Reinforcement Learning. The prototype was extended with actual autonomous behavior of the robot, making the robot capable of choosing its own actions. The role of the tele-operator was limited to navigating the surfacebot to its next location (picked by the surfacebot). The script used for the robot’s actions was replaced by a Q-learning framework, inspired by the work of [22]. It enabled the surfacebot to take actions and to learn from feedback. Q-learning is a form of reinforcement learning where an agent derives the optimal policy on how to act in the current state of its environment, given a transition model that describes the state transitions –the state resulting from an action in a given state– and a reward function which contains the reward received based on a state transition. In our prototype the surfacebot’s state was based on the clothes it was wearing. The children’s feedback determined the rewards on the robot’s actions to maximize the cumulative reward, meaning: which clothes lead to most positive feedback. The surfacebot decided to either explore new states or exploit

known states. Exploration means taking random actions in order to get information about the reward of being in a certain state. Exploitation means going for the action with the highest reward. As action selection strategy, the surfacebot used the epsilon-greedy [23] approach. This strategy enabled an increasing focus on exploiting what was learned, which led to the surfacebot picking the right clothes faster in the later rounds of the activity given that children provided feedback regularly. In other words: at first the robot would try on clothes at random, collecting feedback on its choices, and gradually it would focus more on picking clothes for which it had received positive feedback earlier.

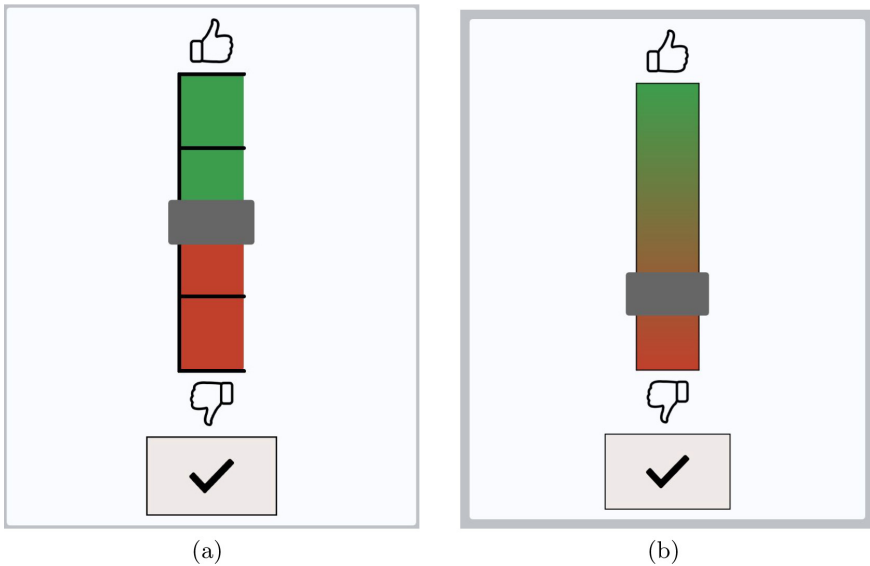


Fig. 6. The reward interface used in the first prototype (a), and the improved interface (b) used in the second prototype. The slider can be used to determine the feedback value. The send button below the slider sends the feedback to the surfacebot.

Undo Behavior. Also inspired by [22], an action cancelling behavior was implemented to provide an immediate response by the surfacebot. If negative feedback was received while the surfacebot was ‘thinking’ about a certain action, the action would be cancelled and the thought cloud would show a different action.

6 Main Study

In the pilot children were engaged and provided frequent feedback, but showed limited collaboration in the activity with the first prototype. In the main study, we explored to which extent the improvements made to the prototype encouraged collaboration. To analyse the collaborative behavior, an annotation scheme was developed for assessing the level of collaboration between children.

6.1 Participants

The study was conducted with 9 pairs of children at a primary school. The age of the children was between 6 and 10 years (mean = 8.00). Two different classes were involved, but it was ensured that the pairs consisted of children from the same class. The parents or guardians of the participating children received a brochure informing them of the nature of the study and the data being collected, and gave consent for their child's participation and for video recording the sessions.

6.2 Setup

The tests were conducted on two different days. The first day, pairs 1–6 participated in the study. One week later, pairs 7–9 participated. The tests were held in two different classrooms, but both rooms were considered to be a familiar setting to the children. The setup was the same as that of the pilot study.

6.3 Procedure

The procedure used in this study was largely similar to the pilot study. The main difference was that before the start of the activity, we tried to get insight into the level of agreement in the preferences of the children regarding the clothing items. The facilitator told the story introducing the surfacebot, and explained the setup and activity. Then the children were shown a list of all clothing items and had to indicate individually which of the items they preferred. At this point, they had not had the opportunity yet to share or discuss their preferences. We expected that if the children's preferences differed, they would need to negotiate to provide consistent feedback to the robot, therefore displaying a higher degree of collaboration. If children were completely in agreement on the items of clothing the robot should get, the task would become unambiguous and there would be little need for negotiation.

6.4 Annotation Scheme for Assessing Collaboration

We developed an annotation scheme to assess the level of collaboration between children during the activity with the surfacebot. The scheme was based on the collaborative problem solving framework of Hesse et al. [11]. Their framework consists of indicators of social skills that form the 'collaborative' part and indicators of cognitive skills that constitute the 'problem solving' part. Our focus was on evaluating the level of collaboration between children, not their skill in solving the task, therefore we used only the indicators of the collaborative part.

Hesse et al. distinguish three classes of indicators that subsume the social skills: participation, perspective taking and social regulation [11]. We took their definitions of the indicators as a starting point and adapted them to assess the level of collaboration between two children in the context of our task, see Table 1. Below we explain how we used the indicators, supported by invented dialogue examples inspired by the observed behaviour of pairs in the pilot study.

Table 1. The annotation scheme used in the main study. Each indicator is listed followed by the definition of Hesse et al. [11], and our interpretation of it in this study.

Indicator	Definition	Description
Participation		
Action	Acts within environment	A child operates the tablet e.g., sends a reward
Interaction	Interacts with, prompts and responds to the contribution of others	A child interacts with the other child
Perspective taking		
Adaptive responsiveness	Ignores, accepts or adapts contributions of others	A child responds to the contribution of another child e.g., disagreeing with a reward sent
Audience awareness	Awareness of how to adapt behaviour to increase suitability for others	A child shares information that is not available to the other e.g., what action the robot thinks of
Social regulation		
Negotiation	Achieves a resolution or reaches compromise	A child attempts to reach an understanding with the other child e.g., about the value of the reward
Self evaluation	Recognises own strengths and weaknesses	A child vocally reflects on an (inter) action regarding the activity
Trans-active memory	Recognises strengths and weaknesses of others	A child vocally reflects on an (inter) action of the other child
Responsibility initiative	Assumes responsibility for ensuring parts of task are completed by the group	A child makes an effort to involve the other child in the activity

Participation. This class consists of the indicators *action* and *interaction*. *Action* was described as “participation of an individual, irrespective of whether this action is in any way coordinated with the efforts of other group members” [11, p.42]. In the context of our activity, we defined actions as the possible interactions with the robot, i.e. operating the reward interface. *Interaction* takes place when children respond to a contribution of another child, i.e. a comment, question or action. The action and interaction indicators served to provide us with an estimation of the level of the children’s engagement with the activity and with each other during the experiment. A third indicator for participation in the framework is *task completion* [11]. This indicator was not applicable in our case, since children were only tasked to provide feedback and the robot decided when the task was completed.

Perspective Taking. The indicators of this class are *adaptive responsiveness* and *audience awareness*. It is considered adaptive responsiveness when a child

accepts or adapts another child’s point of view. We defined *audience awareness* as a child sharing information that was not available to the other child. For example, if one child is the only one able to see the surfacebot’s screen, and tells the other child what action the bear wants to take, this shows that the child is aware of the other’s perspective, and shares the information accordingly.

Social Regulation. This class has as its indicators *negotiation*, *self-evaluation*, *trans-active memory* and *responsibility initiative*. We defined negotiation as an attempt by the children to reach a common understanding, achieve a solution, or reach a compromise. An example of how children could negotiate using the prototype:

Child A: “Item x is super wrong”, and sets slider to absolute negative.

Child B: “No, it is a bit wrong, but not super wrong.”

Child A: Adjusts the feedback slider in accordance to feedback of Child B.

Self-evaluation concerns any comments of a child on their own performance in terms of appropriateness or adequacy in context of (inter)actions during the activity. It indicates a child’s recognition of their own strengths and weaknesses. To illustrate, a child could say: “I was too late, now the robot wears the wrong jacket.” Conversely, trans-active memory is when one of the children comments on the performance of the other in terms of appropriateness or adequacy. For example: “Let me operate the tablet, you were not fast enough!” Lastly, responsibility initiative is about involving others in the task of learning the robot. An indication is the use of first-person plural in communication regarding the activity, for example: “We should let the bear know that it is the wrong item!” In this study, taking responsibility was also annotated when one child encouraged the other to take action or share information. An example of responsibility initiative:

Child A: “Item x is OK, right?”

Child B: “Yes, send the feedback!”

6.5 Identifying Collaborative Behaviour

The annotation scheme described above mostly focuses on the analysis of children’s individual utterances. However, we felt that collaboration can also take the form of certain types of more overarching joint behaviour between children.

First of all, we considered a *division of roles* to be a form of collaboration. It happens when children divide responsibilities and a form of interdependence arises in order to successfully complete the task. An example of a division of roles we observed in the pilot study is that one child operated the tablet while the other child tracked the robot and provided updates on the robot’s actions.

Another form of collaboration is *shared planning*. A shared planning is established when the problem at hand is analyzed and a mutual agreement is reached on how to approach it. A case of shared planning in the context of the activity would be: an agreement between children about how an action of the surfacebot should be judged, as in the following example:

Child A: "The bear should get this jacket... and then go to the hallway."

Child B: "No, it should get the sweater first then go to the hallway."

Child A: "OK, jacket, sweater and then it should get these shoes in the hallway."

Besides a shared planning, children can also build *shared knowledge* by getting an understanding of another child's opinion or preferences, or establishing a shared understanding of (an element in) the activity. It applies when a child provides information or shares an opinion, in response to a question or statement of the other child. An example of shared knowledge is:

Child A: "What items do you think the bear should get?"

Child B: "The jacket and the blue jeans!"

Or when roles have been divided:

Child A: "What is the item that the robot displays?"

Child B: "It is the red jacket."

Lastly, we considered whether there was any *turn taking* between the children. The pilot study showed that children occasionally took turns in operating the tablet, or even played the game individually. Taking turns can only occur after establishing an agreement and can therefore be seen as a result of collaboration. However, it is not necessarily a wanted outcome. In the pilot study, turn taking led to situations where one child temporarily did not actively participate in the activity while the other did everything. It means they did not discover the benefit of collaboration, but made a kind of compromise to each be in full control of the activity for a while. On the other hand, taking turns could also occur while maintaining a division of roles. For example, one operates the tablet, while the other follows the surfacebot from location to location. After a while, they might decide to switch roles. This would be a coordinated effort that maintains the situation where children depend on each other while they both actively participate.

6.6 Measurements

Results were obtained via (1) observations of the video recordings, annotated using the annotation scheme, (2) logged data from interactions with the reward interface, and (3) the clothing preferences that were filled in individually by each participant before the start of the activity.

The video recordings were used for assessing the level of collaboration using the annotation scheme. They were annotated from the point the robot started taking action until the last iteration, where the robot went ‘outside’. Our annotation method was inspired by the work of Huskens et al. [12], who evaluated children’s collaborative play by reviewing 10 s fragments of videotaped play sessions for a fixed set of behaviors. A behavior that was present in the fragment was recorded as a plus. A behavior that was absent was recorded as a minus.

We adopted a similar approach by reviewing 30 s intervals of the recorded sessions using the annotation scheme. Since we wanted to observe the collaboration between children, we opted for a broader interval than 10 s. Each interval was annotated for the presence of any of the collaboration indicators described above. When an indicator was observed, it was marked as positive (+), otherwise as negative (−).

Subsequently, we determined per session a score by summing the positive annotations of the intervals for each category of the annotation scheme. This resulted in 8 indicator scores of which we took the average as an overall collaboration score. Since the sessions differed in number of annotated intervals, we computed both the indicator score (Eq. 1) and the collaboration score (Eq. 2) proportionally to the total number of intervals.

$$\text{indicator score} = \frac{\sum \text{annotated}_{\text{positive}}}{n_{\text{intervals}}} \quad (1)$$

$$\text{collaboration score} = \frac{\sum \text{indicator score}}{n_{\text{indicator scores}}} \quad (2)$$

In the same way, we calculated class (e.g. perspective taking) scores by averaging over the scores of the associated indicators. We then used these scores to compare the level of collaboration between pairs of children.

To check the reliability of the annotation scheme, one recording was annotated by two of the authors. The annotations of both authors were almost the same, except for a few minor differences due to differing interpretations of some indicators. Based on this, the definitions of each indicator were refined and example dialogues were added. This resulted in the indicators described in the previous section. Which general forms of collaborative behaviour (e.g., turn taking) took place was determined per recording, instead of per interval.

6.7 Results

The annotations of the video recordings showed that four pairs of children established a role division, where one operated the tablet and the other tracked the

surfacebot. In this respect, the second prototype encouraged collaboration more effectively, compared to the single observation of a role division during the pilot. The biggest trigger of a division of roles was the fixed tablet, which made following the surfacebot and giving feedback at the same time difficult for one individual, providing an incentive to divide roles. Figure 7 gives an impression of the children's positioning at the start of the activity and when roles were divided. One child communicates feedback to the surfacebot based on the input of the child that follows the surfacebot, who either shares information about the actions of the robot or shares his/her opinion about the action of the robot. A shared planning was observed for five pairs. Building shared knowledge occurred for three pairs. Children taking turns happened for six of the nine pairs. Each pair that established a role division also took turns. The role of operating the tablet seemed favorable to the children, so they switched roles occasionally.

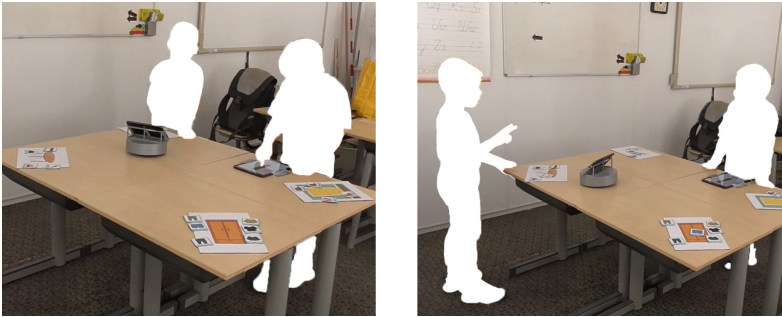


Fig. 7. The image on the left shows an impression of children at the start of the activity. The image on the right shows that they have established a division of roles.

We annotated the recordings of the 9 sessions using the scheme described in Sect. 6.4, with an average of 17.78 ± 3.55 intervals of 30 s per session. For each pair of children, we calculated the average score for each indicator, as well as average scores for the three classes of indicators: participation, perspective taking and social regulation. We also calculated an overall collaboration score based on the average indicator scores.

Because we suspected that the presence or absence of a role division might have an effect on other collaboration aspects, we compared indicator and class scores of pairs based on whether the children had established a role division or not, see Fig. 8. Across the indicators, the pairs that had a division of roles scored higher on other aspects of collaboration as well. This is reflected in the average scores of the three classes and the overall collaboration score, see Fig. 9. Most notably, the four pairs with an observed role division had a considerably larger score for audience awareness compared to the five pairs that did not. The way audience awareness was annotated was to a large extent aligned with how children divided roles. Each role division resulted in one child describing

the surfacebot’s action in situations when the other could not see the character display, which was annotated as a form of audience awareness.

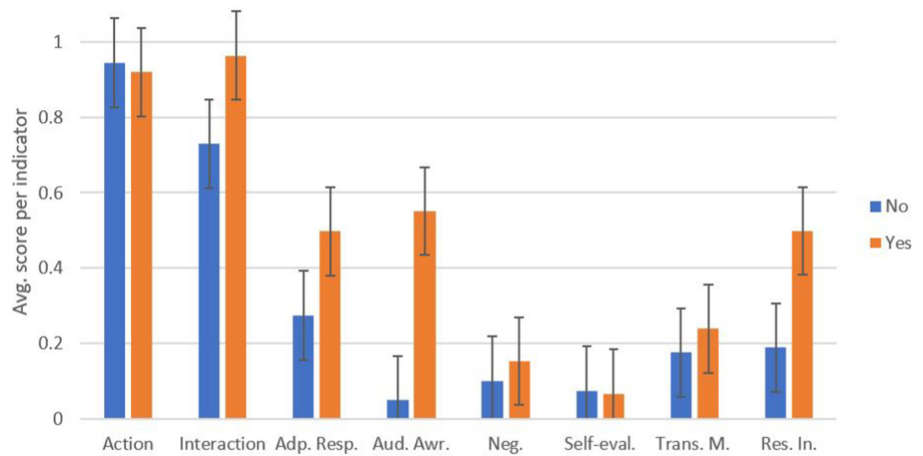


Fig. 8. The average collaboration indicator scores for pairs that established a role division (yes, n = 4) compared to the pairs that did not (no, n = 5).

Two pairs had a very low total collaboration score, 0.14 and 0.25, compared to scores between .38 and .54 for the other pairs. These two pairs communicated less regarding the surfacebot’s actions in comparison to the other groups, and did not establish a division of roles. Remarkable is that all four children in these two pairs were 9 years old, and the oldest among the participating children. Perhaps they found the activity too simple or childish, which resulted in less commitment and engagement compared to the younger children. It could indicate that the current prototype is indeed most suitable for children aged 5 to 7 years.

The use of the slider remained unchanged compared to the pilot study, in spite of the ambiguity introduced in the task and the redesign of the slider. Given the value range of 0 to 1 for the feedback, 78.7% of the communicated feedback was an extreme value, either between 0 and 0.1 (37.7%) or between 0.9 and 1.0 (41%). Therefore, providing an incentive to negotiate by leaving room for disagreement did not lead to a more sophisticated use of the slider. However, forms of negotiation were observed where children tried to determine the slider value together. In that sense, the slider still had added value.

We examined to which extent the children’s initial clothing preferences overlapped. None of the pairs had exactly the same preferences, which suggests there was room for discussion and negotiation. However, we found no indication that children who had little agreement beforehand showed more collaboration, neither in the overall collaboration score, nor in the ‘negotiation’ indicator.

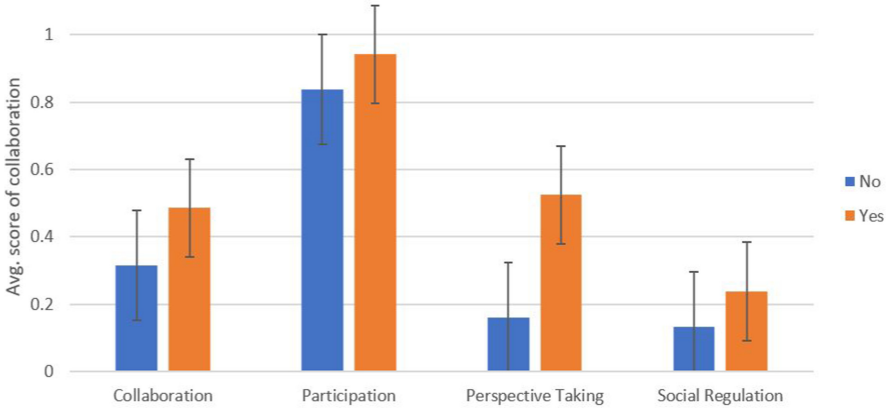


Fig. 9. The average scores of each class for the pairs that established a role division (yes, $n = 4$) compared to the pairs that did not (no, $n = 5$).

7 Discussion

Judging from the results of the main study, the second prototype seemed to encourage collaboration more effectively compared to the first prototype. Analysis of the recordings showed a higher average collaboration score for the pairs of children who established a role division, compared to the average score of the couples who did not. This difference is mainly due to higher scores for the ‘audience awareness’ and ‘responsibility initiative’ indicators. These indicators relate to sharing new information and encouraging others to take an action, which manifested itself when children divided roles.

The collaboration scores of the pairs of children appear to be mostly influenced by the presence of a role division, characterised by interdependence and communication between children. Interdependence is an influencing factor of collaboration [14]. Communication is described as an integral element of collaboration [11] and an interpersonal skill that will develop when children are provided with the opportunities for social interaction [6]. Rather than explicitly assigning roles to the children as in [27], we left it up to the children whether they established a role division or not, and in what way, thus allowing them to adopt their own collaboration style. Supporting diversity in group dynamics has been argued to be beneficial to collaborative learning [17].

However, there are other factors that may have played a role. The children were in a phase where social skills develop, and age may have influenced the level of collaboration. Another factor that we did not take into account was the level of closeness between the pairs of children, who might have been friends or just classmates. It is therefore difficult to attribute any differences in collaboration purely to changes made to the prototype or to an established division of roles.

Unlike earlier studies of collaboration between children [19,27], which were purely qualitative in nature, we attempted to carry out a quantitative analysis

based on observations. The validity of the scoring system we used still requires further investigation. An overall score for collaboration was calculated by weighting each indicator equally. However, it can be argued that some indicators are more indicative or more important for measuring collaboration than others. For example, ‘action’ only says something about the participation of children while ‘negotiation’ is a profound expression of collaboration that requires communication and a certain willingness of both children to listen to each other. In future research, a method could be developed to achieve a more sophisticated collaboration score. The indicators could, for example, be weighted according to their importance or contribution to measuring collaboration.

Children aged 3 to 7 are in a developmental phase where they enjoy fantasy [16], so an activity with a robotic character and a story can be appealing to them. However, the second prototype seems to be mainly suited for children in the age 5–7, since in the pilot study children younger than 5 showed limited understanding of the activity, while in the main study children aged 9 had much lower collaboration scores and seemed less engaged in the activity. In order to reliably compare new versions of the prototype regarding the level of collaboration displayed by children, it would be better to keep the age difference between the (pairs of) children as small as possible.

Finally, due to the small numbers of participants we cannot draw any strong conclusions at this stage; studies with more pairs of children are needed to confirm our preliminary conclusions.

8 Conclusion and Future Work

We iteratively developed a prototype of the surfacebot as a teachable robot designed to encourage collaboration. A pilot study showed that our first prototype only brought about a limited form of collaboration, as most children took turns with only one child actively teaching the robot. In the main study with a revised prototype, multiple pairs of children established collaboration through a spontaneous division of roles, with one of them operating the tablet and the other providing information about the robot’s actions or sharing opinions about them.

Although children showed enthusiasm and established collaboration while interacting with the prototype, it was outside the scope of this research to determine if children learned or developed their collaboration skills from participating in the activity. A recommendation for further research is to conduct a study into the long term effect of participating in activities with the surfacebot (that are aimed to encourage collaboration) on the collaborative skills of primary school children. Do children who regularly participate in such an activity with the surfacebot show improved collaboration compared to children who did not use the surfacebot? And if children improve their collaborative skills through the activities with the surfacebot, are the knowledge and skills transferred to other collaborative activities without the surfacebot? Longitudinal studies could reveal the educative contribution of activities with the surfacebot when introduced in the classroom.

Secondly, we recommend to redesign the activity towards a more ambiguous and complex task that invites using the slider more and gives rise to more discussion. In the current activity, children showed an understanding of the interface and the slider's purpose and effect. However, they mainly provided unilateral feedback with only completely negative or completely positive values. There was low uncertainty or disagreement about the correctness of the surfacebot's actions, limiting the need for negotiation.

Finally, we recommend exploring how learning material from children's regular school subjects can be incorporated into the concept in order to optimally exploit the learning-by-teaching paradigm, while maintaining an activity design that encourages collaboration. Curriculum-focused design techniques could be used to determine which educational topics and activities could benefit from our technology [18].

Acknowledgements. We thank Alejandro Catalá Bolós for the use of the surfacebot. We also thank the children who participated in our experiments, the daycare and the school for facilitating the experiments, and the children's parents for giving their consent.

References

1. Bellanca, J., Brandt, R.: 21st Century Skills: Rethinking How Students Learn. Solution Tree Press, Leading Edge (2010)
2. Benford, S., et al.: Designing storytelling technologies to encouraging collaboration between young children. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 556–563 (2000)
3. Biswas, G., Leelawong, K., Schwartz, D., Vye, N.: Learning by teaching: a new agent paradigm for educational software. *Appl. Artif. Intell.* **19**, 363–392 (2005)
4. Bossert, S.T.: Cooperative activities in the classroom. *Rev. Res. Educ.* **15**, 225–250 (1988)
5. Care, E., Griffin, P.: An approach to assessment of collaborative problem solving. Special issue: assessment in computer supported collaborative learning. *Res. Pract. Technol. Enhanced Learn.* **9**, 367–388 (2014)
6. Catala, A., Theune, M., Gijlers, H., Heylen, D.: Storytelling as a creative activity in the classroom. In: C&C 2017: Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition 2017, pp. 237–242 (2017)
7. Catala, A., Theune, M., Heylen, D.: My tablet is moving around, can I touch it? In: Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction, pp. 495–502. TEI 2017 (2017)
8. Chandra, S., Dillenbourg, P., Paiva, A.: Classification of children's handwriting errors for the design of an educational co-writer robotic peer. In: Proceedings of the 2017 Conference on Interaction Design and Children (IDC), pp. 215–225 (2017)
9. Dillenbourg, P.: What do you mean by collaborative learning? *Collaborative-learning: Cognitive and Computational Approaches*, pp. 1–19 (1999)
10. Garzotto, F.: Interactive storytelling for children: a survey. *Int. J. Arts Technol. (IJART)* **7**(1), 5–16 (2014)

11. Hesse, F., Care, E., Buder, J., Sassenberg, K., Griffin, P.: A framework for teachable collaborative problem solving skills. In: Griffin, P., Care, E. (eds.) *Assessment and Teaching of 21st Century Skills*. EAIA, pp. 37–56. Springer, Dordrecht (2015). https://doi.org/10.1007/978-94-017-9395-7_2
12. Huskens, B., Palmen, A., Van der Werff, M., Lourens, T., Barakova, E.: Improving collaborative play between children with autism spectrum disorders and their siblings: the effectiveness of a robot-mediated intervention based on lego therapy. *J. Autism Dev. Disord.* **45**, 3746–3755 (2014). <https://doi.org/10.1007/s10803-014-2326-0>
13. Jamet, F., Masson, O., Jacquet, B., Stilgenbauer, J.L., Baratgin, J.: Learning by teaching with humanoid robot: a new powerful experimental tool to improve children’s learning ability. *J. Robot.* **2018**, 11 (2018). Article ID 4578762
14. Lai, E.R.: *Collaboration: A literature review*. Pearson Research Report, vol. 2 (2011)
15. Leversund, A.H., Krzywinski, A., Chen, W.: Children’s collaborative storytelling on a tangible multitouch tabletop. In: Streitz, N., Markopoulos, P. (eds.) *DAPI 2014*. LNCS, vol. 8530, pp. 142–153. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07788-8_14
16. Markopoulos, P., Bekker, M.: Interaction design and children. *Interacting Comput.* **15**(2), 141–149 (2003)
17. Rick, J., Marshall, P., Yuill, N.: Beyond one-size-fits-all: how interactive tabletops support collaborative learning. In: *Proceedings of the 10th International Conference on Interaction Design and Children (IDC)*, pp. 109–117 (2011)
18. Rode, J.A., Stringer, M., Toye, E.F., Simpson, A.R., Blackwell, A.F.: Curriculum focused design. In: *Proceedings of Interaction Design and Children, IDC 2003*, pp. 119–126 (2003)
19. Rogat, T.K., Linnenbrink-Garcia, L.: Socially shared regulation in collaborative groups: an analysis of the interplay between quality of social regulation and group processes. *Cogn. Instr.* **29**(4), 375–415 (2011)
20. Roschelle, J., Teasley, S.: *The construction of shared knowledge in collaborative problem solving*. Computer Supported Collaborative Learning (1995)
21. Slavin, R.E.: *Cooperative Learning: Theory, Research, and Practice*. Prentice Hall, Hoboken (1990)
22. Thomaz, A.L., Breazeal, C.: Teachable robots: understanding human teaching behavior to build more effective robot learners. *Artif. Intell.* **172**, 716–737 (2008)
23. Tokic, M., Palm, G.: Value-difference based exploration: adaptive control between epsilon-greedy and softmax. In: Bach, J., Edelkamp, S. (eds.) *KI 2011*. LNCS (LNAI), vol. 7006, pp. 335–346. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-24455-1_33
24. Verhoeven, G., Catala, A., Theune, M.: Designing a playful robot application for second language learning. In: Brooks, A.L., Brooks, E., Sylla, C. (eds.) *ArtsIT/DLI -2018*. LNICST, vol. 265, pp. 385–394. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-06134-0_42
25. Wood, D., O’Malley, C.: Collaborative learning between peers. *Educ. Psychol. Pract.* **11**(4), 4–9 (1996)
26. Wood, D., Wood, H., Ainsworth, S., O’Malley, C.: On becoming a tutor: toward an ontogenetic model. *Cogn. Instr.* **13**(4), 565–581 (1995)

27. Woodward, J., Esmaili, S., Jain, A., Bell, J., Ruiz, J., Anthony, L.: Investigating separation of territories and activity roles in children's collaboration around tabletops. In: Proceedings ACM Human-Computer Interaction (CSCW), vol. 2 (2018)
28. Zancanaro, M., et al.: Children in the museum: an environment for collaborative storytelling. In: Stock, O., Zancanaro, M. (eds.) COGTECH. Springer, Berlin, Heidelberg (2007). https://doi.org/10.1007/3-540-68755-6_8