

Now We're Talking: Learning by Explaining Your Reasoning to a Social Robot

FRANCES M. WIJNEN, DANIEL P. DAVISON, DENNIS REIDSMA, JAN VAN DER MEIJ, VICKY CHARISI, and VANESSA EVERS, University of Twente, The Netherlands

This article presents a study in which we explored the effect of a social robot on the explanatory behavior of children (aged 6–10) while working on an inquiry learning task. In a comparative experiment, we offered children either a baseline Computer Aided Learning (CAL) system or the same CAL system that was supplemented with a social robot to verbally explain their thoughts to. Results indicate that when children made observations in an inquiry learning context, the robot was better able to trigger elaborate explanatory behavior. First, this is shown by a longer duration of explanatory utterances by children who worked with the robot compared to the baseline CAL system. Second, a content analysis of the explanations indicated that children who worked with the robot included more relevant utterances about the task in their explanation. Third, the content analysis shows that children made more logical associations between relevant facets in their explanations when they explained to a robot compared to a baseline CAL system. These results show that social robots that are used as extensions to CAL systems may be beneficial for triggering explanatory behavior in children, which is associated with deeper learning.

CCS Concepts: • **Applied computing** → **Interactive learning environments**; • **Computer systems organization** → *Robotics*; • **Social and professional topics** → *Children*;

Additional Key Words and Phrases: Social robotics, computer-aided learning system, inquiry learning, learning by explaining

ACM Reference format:

Frances M. Wijnen, Daniel P. Davison, Dennis Reidsma, Jan van der Meij, Vicky Charisi, and Vanessa Evers. 2019. Now We're Talking: Learning by Explaining Your Reasoning to a Social Robot. *ACM Trans. Hum.-Robot Interact.* 9, 1, Article 5 (December 2019), 29 pages.
<https://doi.org/10.1145/3345508>

1 INTRODUCTION

Educators have a variety of tools available to support learners to achieve learning goals in their education. These can be as basic as books and chalkboards or more advanced tools, such as simulations or applications in which learners can observe and explore certain phenomena. Technology has become more and more important for education. There are many examples of

This research has been funded by the European Union 7th Framework Program (FP7-ICT-2013-10) EASEL under the grant agreement No 611971.

Authors' addresses: F. M. Wijnen and J. van der Meij, University of Twente, Faculty of Behavioural, Management and Social Sciences, Enschede, The Netherlands; emails: {f.m.wijnen, j.vandermeij}@utwente.nl; D. P. Davison, D. Reidsma, V. Charisi, and V. Evers, University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science, Enschede, The Netherlands; emails: {d.p.davison, d.reidsma, v.charisi, v.evers}@utwente.nl.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.

2573-9522/2019/12-ART5

<https://doi.org/10.1145/3345508>

computer-supported learning and technology-supported education. Technology can often provide assignments and give feedback (right or wrong answer), and, in some situations, the technology is able to adapt the assignment's difficulty level to the learner. In recent years we have seen an emerging role for socially capable interactive learning technologies. Technologies such as virtual agents and robots are capable of interacting and engaging with the learner on a *social* level. Engaging the learner in social interaction opens up possibilities for providing richer task-related support and scaffolding.

In this article, we are specifically interested in a robot's influence on the child's *explanatory behavior* in the context of *inquiry learning*.

1.1 Background

Explanatory behavior is associated with gaining a greater understanding of one's own ideas and knowledge. For the past 15 years, research has proven that explaining leads to a deeper understanding when learning new things [Chi et al. 1989; Coleman et al. 1997; Hayashi et al. 2012; Holmes 2007]. There are two forms of explaining: (1) in solitary learning situations, where the learner explains the subject of interest to themselves, which is called *self-explaining*, and (2) in collaborative learning situations, where a learner explains the learned subject to another person, which is called *interactive explaining* [Ploetzner et al. 1999]. On the one hand, several studies have provided successful examples of self-explanation activities [Chi et al. 1989; VanLehn and Jones 1993]. Yet, providing self-explanations has one important disadvantage: It is not very intuitive to provide detailed explanations to oneself. On the other hand, Rittle-Johnson et al. [2008] found that when children explained to another person (in this case their mother) who was only listening, it had a more positive effect on learning outcomes than self-explaining. A social partner can range from a partner who is just listening, to an interactive partner who provides support and feedback to the learner [Holmes 2007].

Inquiry learning is learning science by doing science. The National Science Foundation states that "inquiry is an approach to learning that involves a process of exploring the natural or material world, and leads to asking questions, making discoveries, and rigorously testing those discoveries in the search for new understanding. Inquiry should mirror as closely as possible the enterprise of doing real science" [NSF 2000, p. 2]. Inquiry learning is often described as a cycle or spiral, which includes formulation of a question, investigation, creation of a solution or a response, discussion, and reflection in connection with results [Bishop et al. 2004]. Inquiry learning has the purpose of engaging learners in active learning, ideally based on their own questions and is a learner-centered and learner-lead process [Saunders-Stewart et al. 2012]. Inquiry learning is characterised by open-ended tasks. Open-ended tasks are often difficult for learners who therefore need help with structuring and understanding [Kirschner et al. 2006; Lazonder 2014; Mayer 2004] to progress to their "Zone of Proximal Development" [Vygotsky 1978].

1.2 Supporting the Learning Process through Computer-aided Learning Systems

In some situations, a learner might be working on an inquiry task without assistance from a human teacher or peer collaborator. This might be the case when working on solitary, individual assignments, when engaging in self-directed learning, or when practising assignments away from school. In such cases, it can still be beneficial for the learner to receive additional support or "scaffolding." In the 1970s and 1980s, as computers were becoming more accessible, researchers discovered the potential benefits of using computers to scaffold the learning process [Kulik and Kulik 1991]. Such Computer Aided Learning (CAL) systems support the learner, for example, by offering background information about the topic, by providing templates, or by constraining the

learner's interaction with the learning environment [Lazonder 2014; van Joolingen and de Jong 1997; Van Joolingen et al. 2007].

Such support systems can be further imbued with social characteristics, such as those offered by virtual agents or robots, to provide a level of *social interaction* that support the learning process [Woolf 2010]. Social interaction in learning goes beyond the support offered by traditional solitary learning materials like books or (digital/interactive) assignment sheets [Vygotsky 1978]. Previous research has shown that social agents can enhance the learning process using various approaches. For example, Ramachandran et al. [2018] found a positive effect of using a physically embodied social robot on thinking aloud behavior on children and Baxter et al. [2017] found that children who work with a social robot that either displays personalized or non-personalized behavior can support children in their learning.

Intelligent Tutoring Systems (ITS) focus on tutoring, coaching, and teaching students through a variety of methods. In the context of inquiry learning, for instance, an ITS might *monitor* and *structure* the inquiry process, offering adequate *advice* where possible [Woolf 2010]. Initially, in such systems the *social interaction* aspects were not very pronounced. Instead, tutoring was often achieved through direct, top-down instructions, feedback, and suggestions, or through manipulations in the (textual or graphical) user interface. To strengthen the social dimension, researchers may extend the ITS with a social agent to deliver the instructions, feedback, and suggestions [Johnson et al. 2000].

Such Animated Pedagogical Agents (APAs) are embodied virtual avatars located within the learning environment. Although APAs may sometimes take on other forms, such as animals- or objects, they are often represented with a humanlike appearance. By using their familiar appearance and rich social modalities, they enrich the learning experience of the student [Gulz 2005]. In addition to scaffolding the learning process by delivering verbal instructions and feedback, APAs may use nonverbal methods of communication, such as locomotion, gaze, gestures, and facial expressions [Johnson et al. 2000]. APAs have shown promising results in various areas of research. For example, APAs can help improve students' self-efficacy in learning (e.g., [Baylor and Kim 2005]), they may reduce anxiety in learning (e.g., [Baylor et al. 2004]), or they may offer motivational scaffolds (e.g., [van der Meij et al. 2015]).

Not all APAs take on the role of tutor or mentor. For example, when an APA presents itself as a co-learner, it may explicitly ask the student for assistance or explanation. Such a Teachable Agent System (TAS) additionally invokes the learning-by-teaching paradigm [Biswas et al. 2005]. Teachable APAs offer the learner a context in which to explain their thoughts and reasoning process, placing emphasis on social interactions between the learner and the agent. For example, Leela-wong and Biswas [2008] found that students *teaching* a virtual agent learned better than those *being taught* by a virtual agent. Their teachable social agent, Betty, was given a background story in which she had to prepare for a quiz.

In summary, socially expressive CAL systems can help support the learning process in various ways. We are particularly interested in investigating how a social robot can make a meaningful contribution to the learning process by leveraging the *interactive explaining* paradigm.

1.3 CAL Systems Enhanced with Social Robots

In previous work, social robots have shown promising results in educational settings. To optimally support the learner, they often take on various roles in the learner's educational process, such as "peer" or "tutor" [Mubin et al. 2013; Okita et al. 2009; Shin and Kim 2007]. The level of involvement with the child's learning process can vary with the role of the robot.

On the one hand, a robot can be presented as a facilitator, teaching assistant or tutor [Chandra et al. 2015; Chang et al. 2010; Kanda et al. 2012; Saerbeck et al. 2010; Zhen-Jia You et al. 2006], who

guides one or more learners working on a task. In such situations the robot often takes on the role of more knowledgeable other. Operating within this role, a social other primarily intervenes in the learning process through tutoring methods such as direct instruction and explicit feedback [Wood et al. 1976]. When offering such tutoring support a robot may benefit from being physically embodied, enabling it to gaze at the user and their actions in the learning environment [Leyzberg et al. 2012].

On the other hand, robots can be presented as co-learners or peers and can have collaborative or spontaneous interactions with children. Within the context of learning, collaboration implies that the peers are more or less on the same status level, can perform the same actions, and have a common goal [Dillenbourg 1999]. Such an approach seems to work well for situations where a child interacts with a robot for multiple sessions or over an extended period of time [Kanda et al. 2004; Tanaka et al. 2007]. Peerlike robots also show promising results in contexts where a certain level of bonding is beneficial to the learning process. For example, the ALIZ-E project describes a social robot that helps diabetic children learn about their condition, through several playful interactions [Belpaeme et al. 2012; Coninx et al. 2015]. Providing the robot with a relatable background story—like the child, it, too, suffers from diabetes—enables more natural co-learning paradigms, where child and robot discover diabetic self-management routines together. These kinds of roles cannot be taken by the technology (as easily) when the learning technology does not include a social agent.

Similarly, children are naturally inclined to help a “robot in need” when it displays distressed behavior, although the extent of helping depends on how the child was introduced to the robot [Beran and Ramirez-serrano 2011; Beran et al. 2010]. Interestingly, researchers may leverage such principles by presenting a robot as a less-knowledgeable other, to whom the child provides caregiving behavior [Tanaka and Matsuzoe 2012] or whom they teach [de Greeff and Belpaeme 2015; Hood et al. 2015]. Using such approaches, students may benefit from explaining certain concepts to their peers, through the learning-by-teaching paradigm [Chase et al. 2009; Rohrbeck et al. 2003]. In addition, Powers et al. [2005] found that participants were inclined to verbalize and explain concepts differently to conversational robots adhering to various stereotypical personas and Hyun et al. [2008] found that children improve their linguistic abilities when working with an interactive robot compared to a multimedia notebook.

Concluding, a social robot has valuable *social characteristics* that can enhance educational technology, enabling it to engage in rich social interactions. Such robots and other social agents can quite naturally be given elaborate background stories and personas, they can develop and adapt their social abilities over time, can engage in interactive social (storytelling) behaviors, and can take on richer and more diverse roles in the learning process. Additionally, a robot is *physically present* in the same physical space as the learning materials. This enables the robot to more naturally interact with the learner in the context of the task, for instance, by gesturing, gazing or pointing.

In this study, we explored the effect of extending a CAL system with a social robot on the explanatory behavior of young children. We leveraged the inherent social nature of the robot to present it as an interactive co-learner with a background story. While guiding the inquiry learning task the children engaged in, the robot attempted to trigger verbal explanations from the child. Furthermore, the robot would provide feedback and ask questions about the inquiry task.

2 RESEARCH QUESTION

The purpose of the present study was to assess the value of a social robot for supporting a child’s verbalization of their knowledge and reasoning process during an inquiry learning task. As a baseline condition, we created a CAL system that guided a child through an exercise, where feedback and questions were read out loud by the system, delivered through a readily available PC speaker.

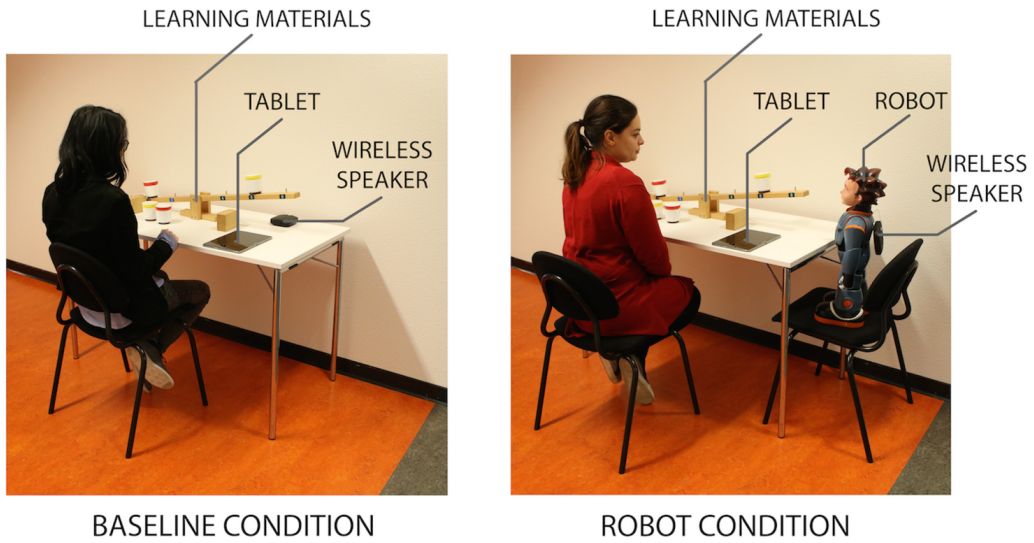


Fig. 1. The experiment setup illustrating two conditions: (1) Baseline no-robot CAL system, where feedback and questions are delivered through a wireless speaker and (2) CAL system enhanced with a social robot, where identical feedback and questions are delivered by the robot through a wireless speaker attached to its back. In both conditions the CAL system was controlled by a dialogue engine running on a laptop, hidden out of view of the children.

In an experimental condition, the same feedback and questions were delivered by a social robot in a way that leverages the embodied, social nature of a robot in the best possible way.

The main research question is as follows: *What are the effects on the explanatory behavior of children when extending a baseline CAL system with a social robot to deliver the system's feedback and questions?*

We expect that a robot with all its social expressive capabilities will be more effective at eliciting verbal explanatory behavior from children compared to the “less-social” baseline CAL system. To measure the expected effect on explanatory behavior, we focus on three dependent variables: (1) verbalisation duration, (2) number of aspects mention in the explanations (breadth), and (3) number of relations explicitly mentioned by the children in their explanations (depth). Based on the research on interactive explaining (e.g., [Ploetzner et al. 1999]) we expect that children would provide more extensive and detailed explanations if they explained to the social robot, which would lead to a longer duration of verbalization. Furthermore, we expect that children would provide more *thorough* (i.e., broader and deeper) explanations to the robot.

3 METHODOLOGY

We conducted a study that employed a between-subjects design with two conditions as illustrated in Figure 1. In the baseline *no-robot condition* children performed an inquiry learning task without a robot, using only our basic CAL system. The system provided spoken assignments, feedback, and questions that were played through a wireless speaker placed on the table. A volunteer recorded the voice of the system beforehand to ensure that the verbal utterances were articulated clearly and understandably. We chose to include such verbal support in the baseline system, since not all children were able to read fluently and to ensure that the two conditions are as similar as possible (the robot also speaks). Children were instructed to provide their answers to the stated questions verbally.

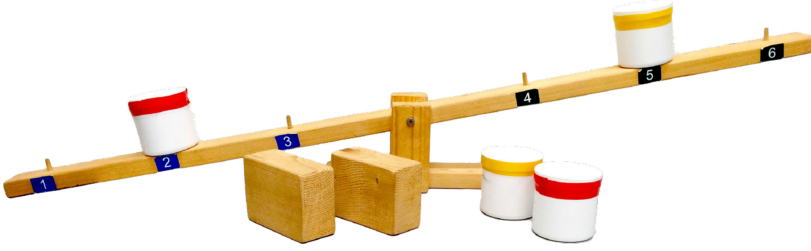


Fig. 2. The balance, colored pots, and wooden blocks used in the inquiry learning task.

In the *robot condition* children performed the same inquiry learning task, using the same baseline CAL system that was now extended with a social robot. In this condition, the robot provided the same assignments, feedback, and questions as was done by the CAL system in the baseline condition. The same audio recordings were played through the same wireless speaker that was now attached to the back of the robot. The behavior design of the robot was informed by design guidelines emerging from an extensive contextual analysis of inquiry learning tasks with our target user group [Davison et al. 2019]. While the child interacted with the robot and the learning task, the robot displayed the following behaviors: (1) *Facial expressions*: When children progressed through assignments the robot showed happy expressions, and directly after children performed the inquiry experiment the robot showed an amazed expression; (2) *Interactive gaze*: The robot gazed toward the child when speaking, the robot gazed toward the tablet when the child pressed a button or when a task appeared, and the robot gazed toward the learning materials when the objects were being manipulated; and (3) *Lifelike behaviors*: The robot blinked at random intervals, and lip-synchronization was added to give the impression that the robot was speaking. Some of the robot's motors (in particular, those in its limbs) are relatively noisy. Therefore, so as not to distract from the learning experience, we chose not to use full-body animations (such as pointing).

Furthermore, the design guidelines that followed from our contextual analysis suggested how the robot should be introduced to children [Davison et al. 2019]. The robot was introduced to the children as a peer but with well-developed inquiry skills (i.e., he knew *how* to perform the inquiry task, but did not know the correct answers yet). We presented the robot with a background story (similarly to, e.g., [Leelawong and Biswas 2008]), thereby contributing to its social characteristics. The background story of the robot was as follows: "It is a student from a planet far away. It is now on earth because it has an assignment from its teacher to study the effects of balance on earth. The robot wants to explore this phenomenon with like-minded people: children!"

3.1 Components of the CAL System

The experiment setup, as illustrated in Figure 1, consisted of the following components:

Balance. The participants' assignment was to explore the concept of balance (i.e., the moment of force) using a balance (see Figure 2). Children received a wooden balance with four pots: two red pots and two yellow pots. The color of the pots represented their weight; red was heavier than yellow. Two wooden blocks could be placed under the balance to prevent the balance from tilting when placing or removing the pots. When children had finished placing the pots on the balance, they could remove the wooden blocks and observe whether the balance tilted to the left, to the right, or remained horizontal.

Tablet. Children in both conditions used a Google Nexus 10 tablet, as shown in Figure 3. This tablet was used to display the assignments using pictures and text and was used as input for the system. Children could indicate that they wanted extra help or continue to the next assignment by



Fig. 3. The tablet interface used during both conditions.



Fig. 4. The R25 Zeno robot

pressing one of two buttons on the tablet: a question mark for help and a green arrow to continue. When children pressed the question mark, the CAL system generated a more elaborate explanation depending on the current phase of the learning task.

Wireless speaker. A small (approximately $10 \times 10 \times 3$ cm) wireless Philips BT2500B Bluetooth speaker was used as audio output for the CAL system. An identical speaker was used in both conditions to ensure uniform audio quality. In the baseline no-robot condition, the speaker was placed on the table. In the robot condition, the speaker was attached to the back of the robot.

Social Robot. Participants in the robot condition worked with the humanoid R25 Robokind robot “Zeno” (see Figure 4). This robot is specifically suitable for simulating human facial expressions such as amazement and happiness. The identical audio recordings for the spoken assignments, feedback, and questions, as used in the baseline CAL system, were now delivered by the robot. To rule out any influence of sound quality, the wireless speaker was attached to the back of the robot, replacing its built-in speaker. The robot displayed lip-synchronization movements in alignment with the prerecorded speech, making it appear as if the robot is speaking. As described above, the robot displayed several behaviors during the interactions (facial expressions, interactive gaze, and lifelike behaviors). The robot’s behaviors were specified using Behavior Markup Language (BML) [Kopp et al. 2006], which was parsed and executed by AsapRealizer [Reidsma and van Welbergen 2013]. BML behaviors were generated at appropriate times during the interaction by the system’s dialogue engine.

3.2 Inquiry Learning Assignments with the Balance

The children received four consecutive assignments. These assignments started easy and became more difficult as children progressed through the tasks. In the first assignment, children used two pots of the same weight and placed them at the same distance from the pivot point on both sides, resulting in the balance staying horizontal. In the second assignment, children used two pots of the same weight, but the distance to the pivot point was varied. As a result, the balance tilted to one side. In the third assignment, the weight of the pots was varied, but the distance to the pivot point was equal on both sides. The result was that the balance tilted to either one of the sides. In the final assignment, both the weight of the pots and the distance to the pivot point were varied. The result was that the balance remained horizontal.

In every assignment, children followed simplified steps of the inquiry cycle: prepare, predict, experiment, observe, and conclude. In the *prepare phase*, children placed the pots on the balance, according to the assignment as was displayed on the tablet. In the *predict phase*, children were asked by the system to verbally state their hypothesis about what would happen to the balance if the wooden blocks, which prevented the balance from tilting, were removed. In the *experiment phase*, children removed the wooden blocks. In the *observe phase*, children were asked by the system to tell what happened to the balance (tilted left, tilted right, or remained horizontal). In the *conclude phase*, the system asked the children to explain why they thought the balance was in this position.

3.3 Dialogue Engine

In both conditions, the CAL system was controlled by a dialogue engine running on a laptop, hidden out of view of the children. This central dialogue engine module was responsible for managing all aspects of the inquiry learning task: (1) monitoring the progression of the child in the task, (2) transitioning between the phases of the task, (3) generating images, buttons, and text to be displayed on the tablet, and (4) playing the audio files (prerecorded speech) through the wireless speaker. The dialogue engine was constructed as a Finite State Machine (FSM) that modelled the interaction of the child with the learning materials as a collection of states, input events, transitions, and outputs.

3.3.1 States. Inquiry phases of the learning process of the child were represented as collections of states in the FSM. For example, at the start of each task, in the prepare phase, children had to place weighted pots on specific pins (e.g., a yellow pot on pin 1 and a red pot on pin 5) and wooden blocks under either side of the balance. The wooden blocks prevented the balance from tipping over and had to remain in place until the experiment phase. The FSM used four states to model the preparation phase, as shown in Figure 5: (1) the initial state containing the basic assignment instructions, (2) a help state for when the child requests or needs (i.e., when they remain idle) additional help, (3) an error state for when the wooden blocks are removed prematurely, and (4) a success state for when the pots are placed correctly.

3.3.2 Events and transitions. The FSM would transition between individual states based on certain input events. Transitions between states were triggered based on the following input events: (i) the child manipulates the physical learning materials (i.e., placing/removing the weighted pots and placing/removing the wooden blocks), (ii) the child presses buttons on the tablet (i.e., pressing the help button and pressing the continue button), (iii) speech activity levels (specifically, state transitions were only triggered on an *end of speech* signal, not on speech content), and (iv) a system timer expires (i.e., the child remains idle for a certain duration, between 10 and 60 s, depending on the phase of the task). These transition triggers were chosen in this way because they do not require complex sensing by the system and are in principle feasible to detect automatically. Figure 5

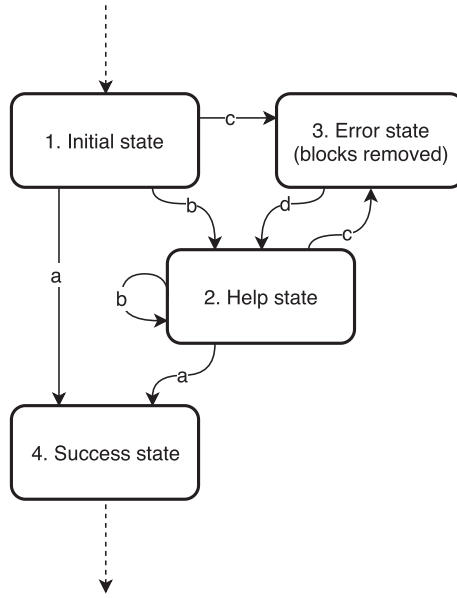


Fig. 5. Diagram of the states and transitions in the preparation phase of an inquiry learning task. Other phases of the task followed a similar structure of initial, help, error, and success states.

shows the following events and state transitions: (a) The system transitions to the success state when the pots are placed correctly according to the task; (b) the system transitions to the help state when the child uses the wrong pots, places pots on incorrect locations, presses the help button, or when the system timer expires; (c) the system transitions to an error state when the child prematurely removes the wooden blocks from under the balance; and (d) the system transitions to the help state when the child corrects an error (i.e., by placing the blocks back under the balance).

3.3.3 System outputs. The system generated an output when transitioning into a state, as a response to the child's actions. The system used the following outputs: (1) updating the interface of the tablet (i.e., showing/hiding image, text, and buttons) and (2) making a verbal utterance (i.e., reading the instructions aloud and offering feedback or encouragement). For example, during the prepare phase the system would offer the following outputs: (1) When transitioning to the initial state, the system would generate an assignment and would display an image, a help button, and a brief textual instruction on the tablet, which would be read out loud (e.g., "You can start by placing a red pot on pin 2 and a red pot on pin 5."); (2) when transitioning to the help state, the system would display more elaborate textual instructions on the tablet and read them out loud to the child (e.g., "The goal is to find out what happens when you put two pots on the balance beam. Before starting with the experiment, you have to have everything ready. Let's begin by putting a red pot on pin 2 and a red pot on pin five."); (3) when transitioning to the error state, the system would offer a textual and verbal reminder to the child to keep the wooden blocks underneath the balance during the preparation phase (e.g., "Hang on, we shouldn't remove the blocks yet. Let's put them back under the balance beam."); and, finally, (4) when transitioning to the success state, the system would display the continue button and would offer verbal encouragement (e.g., "Well done!"). Appendix A gives a comprehensive overview of the system utterances in all phases and states.

In later prototypes of our system, the dialogue engine was capable of automatically incorporating input from the child's tablet (i.e., button presses), external sensors (e.g., to measure the state

of the learning materials [Davison et al. 2016]), and perception modules (e.g., voice activity detection [Fernando et al. 2016]). Based on such inputs, the dialogue engine enabled the CAL system to follow the child's actions in the learning task and respond appropriately. However, in the prototype reported in this article we used a Wizard of Oz approach to simulate several sensors and perception modules.

3.4 Wizard of Oz

We opted for a Wizard of Oz (WoZ) approach to provide input to the system regarding the child's actions in the task. Using this approach a researcher simulated certain sensor values according to a predefined protocol. The simulated sensor values were input into the CAL system's dialogue engine, which then automatically triggered any applicable state transitions and system responses based on its FSM models.

The WoZ protocol was constructed during early pilot tests and was intended to cover the core interaction and common edge cases [Wijnen et al. 2015]. The full WoZ protocol for both conditions is included in Appendix A. The protocol is identical for the baseline no-robot condition and the robot condition. Table 1 shows a representative transcript of a children working on a task from both conditions. During the pilot tests the robustness of the protocol was validated by asking a non-Dutch speaking researcher (who did not understand the *content* of the childrens' speech) to wizard sessions of Dutch children. As part of the core interaction the wizard would simulate the following sensor data:

- (1) Have the weighted pots been removed or are they placed on correct/incorrect locations according to the task?
- (2) Are the wooden blocks removed or are they placed underneath the balance?
- (3) In the *hypothesise*, *observe*, and *conclude* phases: Has the child stopped speaking for 2 seconds?

Additionally, our dialogue engine would keep track of the time that children remained idle in the task (meaning that they did not speak, press buttons, or manipulate the learning materials). After a fixed timer duration had elapsed, the "trigger a timer event" button on the WoZ control interface would light up. Following the protocol, the wizard would press this button only if the child was actually idle but would ignore the button if the child was actively working on the task.¹

Furthermore, in rare cases the child would turn toward the researcher for assistance or would very explicitly ask the system (verbally) what they had to do. In these cases, following the protocol, the wizard would manually trigger a transition to the help state by pressing the "offer additional help" button.²

Finally, in the *conclude* phase of the task, the system would ask the child, "Can you explain why the balance is like this?" Most of the time the child would offer an explanation and the wizard would indicate when they stopped talking. Regardless of the actual content of their explanation, the system's default response was "That's very interesting!" However, on occasion children would simply answer either "No" or "I don't know." In these two situations, the system's default response would be inappropriate. Therefore, according to the protocol, the wizard could select the option "conclusion unknown," enabling the system to respond with "OK," after which the task would continue as usual.³ This was the only situation where the wizard had to interpret the content of the child's speech.

¹The trigger a timer event button was pressed a total of 10 times by the WoZ during our study.

²The offer additional help button was pressed a total of 1 time by the WoZ during our study.

³The conclusion unknown button was pressed a total of 36 times by the WoZ during our study.

Table 1. Transcripts of Sessions in the Baseline No-robot Condition and the Robot Condition

Baseline No-Robot condition	Robot condition
Prepare phase	
System: "Let's start by placing a red pot on pin 2 and a red pot on pin 6"	Robot: "Let's start by placing a red pot on pin 2 and a red pot on pin 6"
Child: Places pots	Child: Places pots
WoZ: Presses button [pots are placed correct]	WoZ: Presses button [pots are placed correct]
System: "Great!"	Robot: "Great!"
Predict phase	
System: "Later, you may remove the blocks. First, can you explain what you think will happen when you remove the blocks?"	Robot: "Later, you may remove the blocks. First, can you explain what you think will happen when you remove the blocks?"
Child: "I think the red pot on pin 6 will go down"	Child: "I think that... the right side will go down a bit, because on the right side it [the pot] is much further"
WoZ: Waits for child to finish speaking for two seconds, then presses button [hypothesis given]	WoZ: Waits for child to finish speaking for two seconds, then presses button [hypothesis given]
System: "OK"	Robot: "OK"
Experiment phase	
System: "Now you're going to do the experiment. You may remove the two blocks from under the balance."	Robot: "Now you're going to do the experiment. You may remove the two blocks from under the balance."
Child: Removes wooden blocks	Child: Removes wooden blocks
WoZ: Presses button [blocks are removed]	WoZ: Presses button [blocks are removed]
System: "Great!"	Robot: "Great!"
Observe phase	
System: "What happened to the balance?"	Robot: "What happened to the balance?"
Child: "6 went down and 2 went up"	Child: "It went to the right because it is further away"
WoZ: Waits for child to finish speaking for two seconds, then presses button [observation given]	WoZ: Waits for child to finish speaking for two seconds, then presses button [observation given]
System: "That's interesting!"	Robot: "That's interesting!"
Conclude phase	
System: "Can you explain <i>why</i> the balance is like this?"	Robot: "Can you explain <i>why</i> the balance is like this?"
Child: "Because... ehmm well first they were on the same pins and now one [pot] is on the other and the balance goes... eh... goes a bit down and then the pot, I think, goes down automatically"	Child: "Because at two it is less far from the thing [midpoint] for example... and at 4... the pot that is on pin 6 is much further away"
WoZ: Waits for child to finish speaking for two seconds, then presses button [conclusion given]	WoZ: Waits for child to finish speaking for two seconds, then presses button [conclusion given]
System: "That is very interesting!"	Robot: "That is very interesting!"

The dialogue engine's FSM models, the system's utterances, and the WoZ protocols were identical in both conditions.

3.5 Exit Interview

After children had completed the experiment, we conducted semi-structured exit interviews. These provided us with an indication whether children perceived the robot as more social compared to the baseline CAL system. In the baseline no-robot condition, from the children's perspective, the CAL system was represented by the tablet. Therefore, during the interviews in the no-robot condition, we referred to the tablet when asking about their experience. The following questions were used in both conditions:

- (1) *Did you enjoy it?* This was asked as a warm-up question for the children to become accustomed to the interview process.
- (2) *What did you like most?* This was the second warm-up question. This question was asked to trigger children to recall the activity. We expected that the most prevalent event from their experience would be mentioned here.
- (3) *Can you tell me something about the robot/tablet?* This question was asked to see what children found the most notable about the robot/tablet.
- (4) *When you go home after school, what will you tell your parents/siblings about what you did here?* Again this question was asked to see which experience was the most prevalent according to the children and worth telling other people about.
- (5) *Could the robot/tablet see what you were doing?* This question was asked to see if children thought they were visible to the technology.
- (6) *Could the robot/tablet hear you?* This question was asked to see if children thought they were heard by the technology.

In the *robot* condition the following additional questions were asked specifically about the robot:

- (7) *Did the robot help you or did you help the robot?* This question was asked to see how children related to the robot in context of the task.
- (8) *Do you think the robot is smart?* This question was asked to see how children estimated the cognitive capabilities of the robot.
- (9) *Do you think the robot is friendly?* This question was asked to see how children regarded the robot on a social level.
- (10) *Do you think the robot thinks you are friendly?* This question was asked to see how children regarded the robot on a social level.
- (11) *How old do you think the robot is?* This question was asked to see whether children would give him a "human" or "device" age. This could indicate how they perceived the robot.

Additionally, children in the baseline *no-robot* condition were given the opportunity to also do one additional assignment *with* the robot, after fully completing the experiment and interview. These children were then asked to compare their experience with the robot and the tablet:

- (7) *What do you think is better? With the tablet or with the robot?* This question was asked to see which device the children preferred in the context of this task.

3.6 Picture Task

A picture selection task was used as a second measure to gain insight into how the CAL system and the robot were perceived by the children. Children were asked to select a picture from a given set that they thought *did* or *did not* fit to the CAL system or the robot. In the no-robot condition, from the children's perspective, the CAL system was represented by the tablet. Therefore, during the picture task we referred to the tablet when asking them to make a comparison.

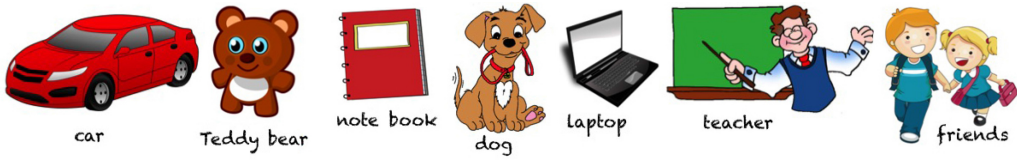


Fig. 6. The collection of pictures presented to the children in the picture task.

When starting the picture selection task a set of images were presented on the table. The images are shown in Figure 6. Children first received a picture of either the tablet or the robot. They were then asked to select a picture that, according to them, *best fitted* with the tablet/robot and explain why they chose that specific picture. This was repeated for a second picture. Then children were asked to pick a picture that, according to them, *did not fit* with the tablet/robot and explain why they chose that picture. This was also repeated for a second picture. The picture task was embedded in the exit interview and was introduced after children had answered question six. When children finished the picture task the experimenter continued the interview with question seven.

The exit interviews, including the picture task, were video-recorded. The answers children gave were first transcribed. Then, the data were structured according to categories that emerged from the data. Analysis of the categories gave insights in the responses children gave per question in the two conditions and an overview of the responses children gave during the picture task per condition. Three researchers coded these answers. An overview of the responses is presented in Section 4.4.

3.7 Annotation of Children's Speech

Every session was video-recorded from two angles to capture the full interaction scene. In addition, our dialogue system automatically collected log data that contained all input from the Wizard of Oz tablet and from the tablet used by the children and every output action from the CAL system. The log data were used to automatically annotate the videos, using the language archive program ELAN.⁴ The videos were then further manually annotated on two levels.

The first level was *child speech* (annotated manually for inquiry phases *predict*, *observe*, and *conclude*) and contained two labels: verbalization and fillers. *Verbalization* was used when children provided verbal explanations and was used directly to assess the duration of verbalization. *Fillers* was used when children voiced utterances such as “ehmm...,” which indicated that they wanted to say something. Fillers were only annotated when they occurred at the start of an utterance, preparatory to the child providing their actual explanation. We separated preparatory fillers from verbalization, because we were primarily interested in measuring the length of the actual explanation.

The second level was *system speech* and contained three labels: giving explanation, asking question, and waiting for response. System speech was annotated automatically based on generated log files from the CAL system. *Giving explanation* was used when the CAL system gave an explanation or a verbal response to the child (e.g., stating the assignments, offering more elaborate help, or giving encouragement). *Asking question* was used when the system stated a question and was also annotated automatically (e.g., asking for a hypothesis, observation, or conclusion). *Waiting for response* was used when the system had stated a question and was waiting for a response from the child. The annotations were converted to numerical data and analyzed using the statistical analysis program (R) [R Development Core Team 2008].

⁴For more information on the ELAN tool, see Sloetjes and Wittenburg [2008].

3.8 Content Coding: Children's Explanations

To gain a better understanding of the quality and thoroughness of children's explanations, we coded the content of their explanations on two levels: (1) breadth—how many different things do they talk about and (2) depth—how often they relate these things together.

First, to analyze the *breadth* of each explanation we investigated the number of facets that were mentioned. Using a grounded theory approach, we derived a coding scheme consisting of nine codes (the facets) through axial coding:

Balance position Used when a child explicitly made a remark about the current position of the balance. For example, “this side went down” or “it is in balance.”

Weight of pots Used when a child explicitly mentioned the (relative) weight of the pots. For example, “this pot is just as heavy as that one” or “there is more in this pot than in that one.”

Distance of pots Used when the child explicitly mentioned the (relative) distance of the pots to the pivot point. For example, “their distance is the same” or “this one is all the way to the end.”

Position of pots Used when a child explicitly mentioned the (absolute) position of the pots on the balance. Note: This is different from mentioning the distance from the pivot point. For example, “this pot is on pin 1 and this one is on pin 4” or “the red one is here and the yellow one is there.”

Example/counter example Used when the child gave a (counter)example to clarify his/her explanation. For example, “the balance has now tilted to the right, but if this pot were here it would remain in balance.”

Naive weight Used when the child explicitly mentioned that one side of the balance was heavier, causing it to fall down. In this case, the child did not necessarily show understanding of all principles of the balance work but did show a naive understanding of the moment of force. For example, “the pots have equal weight but this side of the balance is heavier” or “this pot is heavier, but there is the same weight on both sides.”

Circular reasoning Used when a child explained the position of the balance without mentioning one of the variables (distance from the pivot point or weight of the pots). In other words, they were *stating the obvious*, without showing understanding of the underlying principles. For example, “the left side goes down because it is heavier” or “it is balanced because it weighs the same on both sides.” As opposed to the “naive weight” code, a child using circular reasoning does *not* explicitly mention the weight or position of the pots on the balance in their explanation.

Other Used when the child talked about something that could not be coded with one of the codes described above.

No answer Used when a child did not know the answer or gave no explanation at all.

These nine codes covered all facets relevant to explaining the outcomes of our balance task. During content analysis, we coded *whether* a child talked about a specific facet during their explanation, not how long or what exactly they said about it. Consequently, each explanation was coded with respect to the facets that were mentioned in the explanation. In cases where children mentioned multiple facets in their explanation, this resulted in multiple codes. If a child repeated his/herself during an explanation, then multiple identical codes were used. For example, an explanation such as “The red pot is heavier than the yellow pot. So then the balance tipped to the right, because the red pot is heavier.” would be coded “Weight of pots” twice, and “Balance position” once.

Following the above approach, researcher A annotated all data and researcher B independently annotated 20% of the data. A Cohen's K of 0.87 showed an excellent inter-rater agreement. Some confusion between annotators was found for codes *distance of pots* and *position of pots*. Closer examination of the coded fragments revealed an ambiguity in the wording of the coding scheme, which was consistently misinterpreted by researcher B. Correction of this ambiguity resulted in an updated Cohen's K of 0.92, which indicates that the above collection of codes are a viable method for reliably coding and interpreting the content of children's explanations.

Second, to analyze the *depth* of each explanation we coded the number of logical associations that were explicitly mentioned by the child connecting those facets. Explanations that incorporate such associations (or inferences) seem to enhance students' knowledge [Ploetzner et al. 1999] and are important characteristics of scientific explaining [Kuhn and Reiser 2005]. Explicitly mentioning associations between facets may illustrate an understanding of the presence (or absence) of a relationship that may exist between the facets.

We annotated the use of logical deductive phrases to code whether an explicit association was drawn between facets of an explanation. Specifically, phrases that are causal, oppositional, or conditional in nature. Examples of such phrases are as follows: "but," "although," "because," "that's why," "and so," "therefore," "that means that," "provided that," "otherwise," and so on. On the one hand, an example of a shallow explanation could be "The balance is horizontal. The red pot is heavy and is placed on pin four. The yellow pot is light and is placed on pin one." Although this example covers three relevant facets, namely [Balance position], [Weight of pots], and [Position of pots], it does not explicitly illustrate a deeper level of understanding with regard to the logical relations that exist between the facets. On the other hand, a much deeper explanation using the same three facets could be "Although the red pot is heavy, it is placed on pin four. The yellow pot is light and is therefore placed on pin one. That's why the balance is horizontal." This phrase explicitly draws logical associations between the mentioned facets with the keywords "although," "therefore," and "that's why." For this example, we would annotate three associations.

3.9 Procedure

Prior to the experiment, ethical approval was granted by the ethical boards of the Behavioural, Management and Social sciences (BMS) and the Electrical Engineering, Mathematics and Computer Science (EEMCS) faculties of the University of Twente. Permission forms were distributed to the children's parents through the school's regular communication channels. Sessions took place in a separate room in the school during regular school hours. The children first received an introduction by the experimenter, which took approximately 5 minutes. In the robot condition, the experimenter first introduced the Zeno robot and its background story. In both conditions the experimenter told the children that they should provide their answers by talking out loud. The experimenter then showed how the pots were placed on the balance and explained the function of the two buttons on the tablet (the question mark and the green arrow). After this introduction the experimenter walked away and sat somewhere out of sight in the same room. The children then started with the first assignment. During the experiment, the experimenter served as Wizard of Oz controller. It took children approximately 20 minutes to complete all four assignments. Then the experimenter started the exit interview, including the picture task. Children from the baseline no-robot condition needed approximately 5 minutes to complete the interview and picture task. Children from the robot condition needed approximately 10 minutes.

To clarify how sessions went in practice we added two videos as supplementary materials to this publication. In the first video we highlight the setup and procedure of the experiment and the function of the Wizard of Oz. In the second video an actor reenacts a transcript of a representative session from each condition, to illustrate how real sessions with children went in practice.

Table 2. Participant Demographics per Condition

	Age	Nr. of participants
	Mean (SD)	Total (girls/boys)
No-Robot	8.2 (1.5)	22 (15/7)
Robot	8.1 (1.3)	24 (12/12)

3.10 Participants

In total, 53 Dutch elementary school children participated in the study. Sessions that were incomplete or that contained missing data were removed, resulting in 46 sessions suitable for further analysis. The demographics for each condition are given in Table 2. We recruited two composite classes (grades 1-2-3) from two different schools in the same city to participate in this study. Children from these classes were randomly assigned to either the baseline no-robot condition ($n = 22$) or the robot condition ($n = 24$).

4 RESULTS

4.1 Duration of Verbalization

Analysis for normality showed that the data for duration significantly deviated from a normal distribution in both conditions (Shapiro-Wilk $p < 0.001$). Therefore, Mann-Whitney U tests were used to determine the differences in duration between the two conditions.

The *response duration* of the child was measured as the duration (in seconds) of their responses to questions stated by the system (minus any leading fillers like “Uuuhm”). The Mann-Whitney analysis showed that, overall, children verbalized for a significantly longer duration in the robot condition ($M = 5.7$, $SD = 2.5$) than in the baseline no-robot condition ($M = 4.2$, $SD = 2.4$), $U = 172.00$, $z = -2.019$, $p = 0.043$ (two-tailed).

In addition to the analysis of the overall results, we looked more in depth into the differences per inquiry phase (with Bonferroni correction: $\alpha = 0.0167$). The Mann-Whitney U test showed no significant differences in the *predict phase* for duration (robot: $M = 4.6$, $SD = 2.6$, no-robot: $M = 3.7$, $SD = 2.2$), $U = 194.00$, $z = -1.525$, $p = 0.127$ (two-tailed). In the *observe phase*, the Mann-Whitney U test showed no significant difference between the baseline no-robot and the robot condition for the duration of the verbalization (robot: $M = 5.5$, $SD = 3.4$, no-robot: $M = 3.4$, $SD = 1.8$), $U = 161.00$, $z = -2.270$, $p = 0.023$ (two-tailed). In the *conclude phase* there was no significant difference between the conditions for duration (robot: $M = 7.1$, $SD = 3.7$, no-robot: $M = 5.5$, $SD = 4.1$): $U = 186.00$, $z = -1.704$, $p = 0.088$ (two-tailed).

These results show that children verbalized for a longer duration to the robot. From this we conclude that in this study questions asked by a social robot yielded longer verbalizations than those asked by a baseline CAL system.

4.2 Explanation Breadth

We carried out a content analysis of the children’s verbalizations by annotating the various facets that the children addressed in their responses. A preliminary analysis of children’s responses revealed that, in both conditions, the most elaborate explanations were given in the *observe phase* and *conclude phase*. We therefore focused our annotation on these two phases. In the *predict phase* children primarily stated their hypotheses, often without providing additional justification or explanation.

The results of the content analysis are shown in Figure 7 and summarized in Table 3. A Mann-Whitney U test was used to investigate whether the differences between the number of codes in

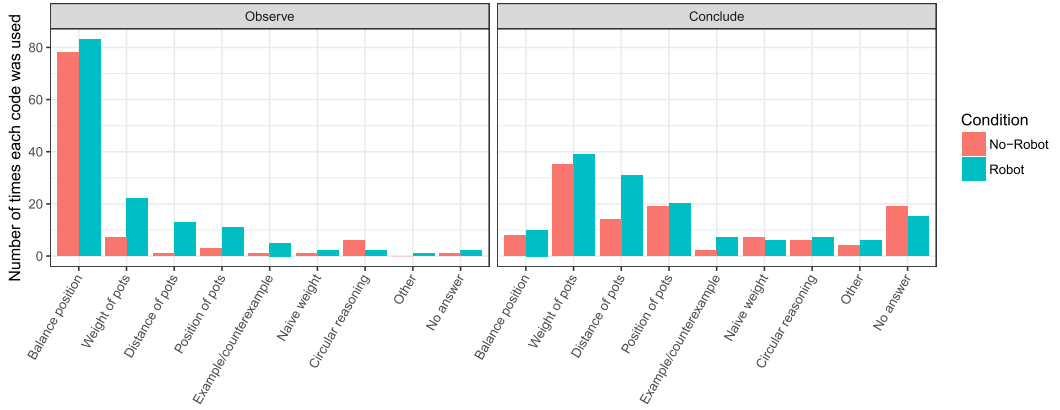


Fig. 7. Explanation breadth: The number of times each code was used per condition for the Observe and Conclude phases.

Table 3. Explanation Breadth: Number of Times Each Code Was Used During Annotation in Each Phase in the Two Conditions

Code	No-Robot		Robot	
	Total	Mean (SD) per child	Total	Mean (SD) per child
Observe phase				
Balance position	78	3.55 (0.96)	83	3.46 (1.02)
Weight of pots	7	0.32 (0.57)	22	0.92 (1.06)
Distance of pots	1	0.05 (0.21)	13	0.54 (0.72)
Position of pots	3	0.14 (0.47)	11	0.46 (0.72)
Example/counterexample	1	0.05 (0.21)	5	0.21 (0.72)
Naive weight	1	0.05 (0.21)	2	0.08 (0.28)
Circular reasoning	6	0.27 (0.88)	2	0.08 (0.28)
Other	0	0 (0)	1	0.04 (0.2)
No answer	1	0.05 (0.21)	2	0.08 (0.41)
<i>Sub-total</i>	98	4.45 (0.8)	141	5.88 (2.31)
Conclude phase				
Balance position	8	0.36 (0.66)	10	0.42 (0.5)
Weight of pots	35	1.59 (1.18)	39	1.62 (0.97)
Distance of pots	14	0.64 (0.9)	31	1.29 (1.16)
Position of pots	19	0.86 (1.08)	20	0.83 (1.01)
Example/counterexample	2	0.09 (0.29)	7	0.29 (0.86)
Naive weight	7	0.32 (0.57)	6	0.25 (0.53)
Circular reasoning	6	0.27 (0.63)	7	0.29 (0.75)
Other	4	0.18 (0.39)	6	0.25 (0.44)
No answer	19	0.86 (1.49)	15	0.62 (1.06)
<i>Sub-total</i>	114	5.18 (2.11)	141	5.88 (1.75)
Total	212	9.64 (2.63)	282	11.75 (3.35)

For each code, we report the mean and its standard deviation of the number of times that code was used during the annotation of each child.

Table 4. Examples of Child Utterances Given in Both Conditions in the Observe Phase

Baseline No-Robot condition	Robot condition
Observe phase	
Child 1: “It’s still the same [balanced]”	Child 4: “The balance was the same, because the two pots have the same weight, so then the balance becomes like a scale.”
Child 2: “It remains the same [balanced]”	Child 5: “If you put one pot on 2 and the other on 5, then they are both in the middle, and if you then remove the blocks then the pots are balanced.”
Child 3: “It stays the same [balanced]”	Child 6: “They are the same weight”

For each condition, these examples reflect the three children who mentioned the most facets in their explanation in the first task.

Table 5. Explanation Depth: Number of Logical Association between Facets in Each Phase per Condition

	No-Robot		Robot	
	Total	Mean (SD) per child	Total	Mean (SD) per child
Observe phase	6	0.07 (0.14)	46	0.48 (0.53)
Conclude phase	73	0.84 (0.63)	94	1.01 (0.50)
Total	79	0.45 (0.35)	140	0.74 (0.40)

each condition was significantly different. First, for each child the aggregated number of codes was calculated over the whole session consisting of four assignments. We found that, overall, in the robot condition ($M = 11.75$, $SD = 3.35$) children mentioned more facets in their explanations than in the baseline no-robot condition ($M = 9.64$, $SD = 2.63$), $U = 143.50$, $z = -2.675$, $p = 0.007$ (two-tailed). In the *observe phase* we found a significant difference (with Bonferroni correction $\alpha = 0.025$) between the no-robot condition ($M = 4.45$, $SD = 0.8$) and the robot condition ($M = 5.88$, $SD = 2.31$), $U = 157$, $z = -2.60$, $p = 0.009$. We found no overall significant difference in the *conclude phase*, $U = 186.00$, $z = -1.758$, $p = 0.079$. When we explore Table 3 we see that the facets balance position, weight of pots, distance of pots and position of pots are the most often mentioned facets in both conditions. This is interesting, because these facets are crucial to mention if one wishes to fully explain the function of the balance.

These results indicate that the children not only verbalized longer but also mentioned more relevant facets in their explanations. From this, we conclude that when questions were delivered by a social robot, as opposed to a baseline CAL system, children were prone to give more detailed and relevant explanations.

4.3 Explanation Depth

To investigate the depth of the children’s explanations we coded how often children drew logical associations between facets through the use of causal, oppositional, or conditional deductive phrases. Table 4 provides some examples of verbal utterances given by children in the observe phase for both conditions.

We annotated a total of 140 logical associations in the robot condition and a total of 79 in the baseline no-robot condition. Table 5 lists the details per condition and per phase. A Mann-Whitney U test showed a significant overall difference between conditions, $U = 156$, $z = -2.364$, $p = 0.018$

(two-tailed). For explanations given in the *observe phase*, we found a significant difference between the robot condition and the baseline no-robot condition, $U = 141$, $z = -3.026$, $p = 0.002$ (two-tailed) (with Bonferroni correction alpha is 0.025). No significant difference was found in the *conclude phase*.

This indicates that children who interacted with the robot made more (explicit) logical associations between relevant facets compared to children who worked with the no-robot baseline condition.

4.4 Exit Interviews: Was the Robot Perceived as Social?

In the baseline no-robot condition, there were three sessions for which the log file data was not properly stored and therefore excluded in the analysis of the verbalization content and duration. Otherwise, these sessions were completed and were therefore included in the exit interview and picture task analysis resulting in the analysis of responses from 25 children in the baseline no-robot condition and 24 children in the robot condition.

The first two questions in the interview were warming-up questions and were asked primarily so the child could get used to the interview process. Responses to these two questions are reported as such and will not be further interpreted. To the first question, *Did you enjoy it?*, most children (44 of 49) answered that they enjoyed the experiment. To the second question, *What did you like most?*, children in both conditions reported that they enjoyed some part of the inquiry, such as "I enjoyed placing the pots on the balance." Some children also mentioned the interaction with the tablet/robot. For example, "I can just talk to the robot" or "that thing [tablet] talked to me, which was funny."

The third question, *Can you tell me something about the robot/tablet?*, was asked to see what children found most notable about the robot or tablet. Most children in the robot condition gave a response regarding the interactive behavior of the robot (10 of 24), for example, "he [the robot] can explain things to me." Furthermore, some children mentioned the embodiment of the robot (8 of 24). Children focused on two things: that the robot moved and the size of the robot. Interestingly, children did not mention the facial expressions that the robot displayed. Additionally, children in the robot condition mentioned the role of the robot in the task (6 of 24), for example, "the robot taught me a lot." In the baseline no-robot condition, children mostly talked about the interaction with the tablet (10 of 25). For example, "it was interesting that it could talk." Instead, children in the baseline condition mentioned the embodiment/design (5 of 25) of the tablet, for example, "there are some pictures here for if you don't quite understand."

To the fourth question, *When you go home after school, what will you tell your parents/siblings about what you did here?*, the responses of the children in both conditions were mostly related to the description of the task (baseline: 11 of 25, robot: 15 of 24), for example, "that I have worked with the balance." In the robot condition, the children also talked about the embodiment of the robot (7 of 24), for example, "I would tell that the robot could move and talk" or "that he is small and has a pleasant voice." In addition, again several children in the robot condition mentioned the interactive behavior of the robot (5 of 24), for example, "He [the robot] can talk." In the no-robot condition, only one child mentioned the embodiment of the tablet. Children were more focused on the interaction with the tablet (8 of 25). For example, "he [the tablet] talked to me and I didn't have to type, but I could just talk to it."

To the fifth question, *Could the robot/tablet see what you were doing?*, most children in the robot condition answered that they thought the robot could see them (19 of 24). In the no-robot condition, about half of the children said that the tablet could see them (12 of 25). We also asked *why* children thought the robot/tablet could (or could not) see them? In both conditions, responses were related either to some technical aspect, such as, "I think he (the robot) has a camera in his eyes" or

Table 6. Pictures That Were Chosen That *Did* Fit with the Tablet or Robot for Each Condition

Picture	No-robot		Robot	
	1st choice	2nd choice	1st choice	2nd choice
Car	0.0%	30.4%	12.5%	8.7%
Teddy bear	0.0%	4.3%	0.0%	4.3%
Note book	8.0%	26.1%	0.0%	0.0%
Dog	4.0%	4.3%	0.0%	0.0%
Laptop	80%	0.0%	50.0%	26.1%
Teacher	8.0%	34.8%	16.7%	39.1%
Friends	0.0%	0.0%	20.8%	21.7%

the reactive behavior of either the tablet or robot, such as “if I put the blocks under the balance, the green arrow appeared.” To the sixth question, *Could the robot/tablet hear you?*, most children in the robot condition thought that the robot was able to hear them (19 of 24). In the no-robot condition, 18 of 25 children thought that the tablet could hear them. When asked to explain their answers, most explanations were related to the responsive behavior of the system. However, occasionally children in both conditions also mentioned that it was some external factor that caused the robot/tablet to give a response, such as “you heard me and controlled the robot.”

In the robot condition some additional questions were asked. To the question *Did the robot help you or did you help the robot?*, the majority of children responded that he/she and the robot helped each other. To the question *Do you think the robot is smart?*, most children responded that they thought the robot was smart. To the question *Do you like the robot?*, almost all children answered “yes.” To the question *Do you think the robot likes you?*, Almost all children answered “yes.” To the last question, *How old do you think the robot is?*, 20 children gave the robot a “human age” between 5 and 12 years old, which is around the same age as most of the children. There were only 4 children who asked for clarification or responded in a way that indicated they were considering giving the robot a “device age,” for example, “I think he is not so old. We didn’t have robots yet in the year 2000.”

In the baseline no-robot condition, children had the opportunity to do one assignment with the robot after they finished the experiment. When children chose to do this, we asked them: *What do you think is better? With the tablet or with the robot?* Eighteen of the 22 children indicated that they preferred the robot. In most cases, the children mentioned the appearance of the robot. For example, “he can move and he looks at you and he really says something!” There was only one child who said that he/she preferred working with only the tablet, although he/she did not explain why. Three children said they enjoyed working with both the robot and the baseline CAL system but did not elaborate on this.

4.5 Picture Task: Was the Robot Perceived as Social?

Children were asked to choose a *first* picture that best fitted with the robot/tablet and explain their choice. They were then asked to choose a different *second* picture and explain their choice. Table 6 gives an overview of the pictures children chose in each condition.

In both conditions, the majority of children first selected the laptop as the best fit for the robot/tablet. However, there was some more variation in the selection of pictures in the robot condition compared to the no-robot condition. Interestingly, some children in the robot condition chose the picture of the children, while no one in the baseline no-robot condition did.

Table 7. Pictures That Were Chosen That *Did Not* Fit with the Tablet or Robot for Each Condition

Picture	No-robot		Robot	
	1st choice	2nd choice	1st choice	2nd choice
Car	9.5%	9.5%	30.4%	17.4%
Teddy bear	33.3%	52.4%	13.0%	30.4%
Note book	0.0%	4.8%	13.0%	0.0%
Dog	47.6%	33.3%	30.4%	39.0%
Laptop	4.8%	0.0%	0.0%	0.0%
Teacher	0.0%	0.0%	0.0%	13.0%
Friends	4.8%	0.0%	13.0%	0.0%

Table 8. Frequency of Annotation Labels Coded in Participants' Explanations for Choosing a Picture That *Did* or *Did Not* Fit with the Robot or Tablet

Annotation label	Picture <i>did</i> fit		Picture <i>did not</i> fit	
	No-robot	Robot	No-robot	Robot
Technology/design	48.0%	45.0%	30.4%	15.4%
Tool/function	28.0%	2.5%	34.8%	11.5%
Social characteristics	0.0%	20.0%	34.8%	53.8%
I learned from the system	24.0%	12.5%	0.0%	11.5%
I taught the system	0.0%	20.0%	0.0%	7.7%

Additionally, children were asked to select a first and then a second picture that *did not* fit with the tablet/robot. Children in the robot condition most often chose the picture of the car or the dog as a picture that did not fit with the robot, as shown in Table 7. In the baseline no-robot condition most children chose the dog or the teddy bear as a picture that did not fit.

Children were asked to explain in their own words why they chose a specific picture. Their explanations were annotated on several labels that emerged from the data. Table 8 provides an overview of how often the answers were coded according to the annotation labels. These annotation labels were based on the topics children described in their explanations. Children in both conditions mainly focused on the technology or design of the robot/tablet when choosing a picture that *did* fit. For example, "The robot is controlled by a computer." Interestingly, children who worked with the robot also talked about the social characteristics of the robot. For example, "he [the robot] is friendly." Some of them talked about them teaching something to the system, indicating that they perceive the robot as being capable of learning. For example, "I am actually the teacher now." Similarly, when explaining why a picture *did not* fit, children in the robot condition mainly focused on describing the social characteristics of the robot, for example, "[the car] is similar to a robot, but it does not bring you company." In the baseline no-robot condition, children mainly focused on describing the tool/function of the tablet and its social characteristics.

5 DISCUSSION

In educational literature, several studies show that making your reasoning explicit by explaining something either to yourself or another person can contribute to deeper understanding and better learning [Chi et al. 1989; Coleman et al. 1997; Holmes 2007; Ploetzner et al. 1999]. Therefore, prompting children to engage in explanatory behavior is associated with better learning. The results of our study show that children *verbalized for a longer duration* to the robot. From this, we

conclude that questions asked by a social robot yielded *longer* verbalizations than those asked by a baseline CAL system.

To gain a better insight in the content of children's verbalizations, we looked specifically at the *breadth* and *depth* of their explanations. First, regarding the explanation breadth, we found that children touched upon more relevant facets in the observe phase when explaining to the robot compared to the baseline no-robot CAL system. More specifically, children in the robot condition mentioned the facets *weight of pots* and *distance of pots* significantly more often. These facets are crucial for explaining the tilt of the balance and should therefore be included in a good explanation of the working of the balance. The majority of children who worked with the robot started explaining the working of the balance in the observe phase, without being prompted to do so, while we only occasionally saw this for children who worked with the baseline CAL system. We found that 15 of 24 (62.5%) children who worked with the robot gave at least once a more elaborate explanation than just an observation about the balance position. In the baseline CAL condition, this occurred for 8 of 22 (36.4%) children. This could indicate that children are more eager to explain to the robot how the balance works compared to explaining to the baseline CAL system. Second, regarding the explanation depth, we found that children draw more logical associations between the various facets of their explanations in the observe phase, showing a deeper level of deductive reasoning. From this, we conclude that the CAL system extended with the robot is better able in motivating children to give more elaborate and detailed explanations, which, according to the literature, is an indication of deeper learning.

There are several possibilities that might help explain why the robot yielded such results. By adding the robot, several aspects were introduced that were not present in the baseline CAL condition. First, the robot is physically present. Second, the robot displayed facial expressions. Third, the robot displayed interactive gaze. Fourth, the robot displayed lifelike behaviors such as blinking its eyes and moving its mouth appearing as if he was speaking. Last, the robot was introduced with a background story. Each of these factors might motivate children to provide more elaborate explanations.

For example, Leyzberg et al. [2012] found that the physical presence and embodiment of robots help improve cognitive learning gains during tutoring and Ramachandran et al. [2018] found a positive effect of a physically embodied social robot on thinking aloud behavior in children. Furthermore, Zaga et al. [2017] found that social gaze movements increase children's perception of animacy and likeability. In addition, a robot that displays both social gaze and deictic gaze (which the robot in this study did) is perceived as more helpful and this might engage learners more. Furthermore, Powers et al. [2005] found that the way a robot is introduced (e.g., background story) impacts the way participants verbalize and explain concepts. Finally, facial expressions are known to impact how people perceive a robot and these facial expressions can complement or enhance an agents message [Breazeal 2003].

Although we cannot draw grounded conclusions about which of these factors contribute to the observed effect, we did conduct exit interviews to gain insight in whether the children perceived the robot as more social. To gain insight in what children found most notable about either the robot or the tablet, we asked them whether they could tell something about the robot/tablet. In their answers, children from both conditions talked most about the interaction with the robot/tablet. However, slightly more children in the robot condition talked about the embodiment of the robot (8 of 24) than children in the tablet condition (5 of 25). In the robot condition, they hereby focused on two things: that the robot moved and the size of the robot. Interestingly, children did not mention the facial expressions the robot displayed. In addition, we asked children what they would tell other people (in this case, their parents) about the experiment. In the robot condition, children again mentioned the embodiment of the robot (7 of 24) more often than children in baseline condi-

tion (1 of 25). Children mentioned the fact that the robot could move, but they also mentioned the voice of the robot (5 of 24), which they perceived as pleasant and similar to a children's voice. In the tablet condition, none of the children mentioned the voice, although this was the same voice as the robot. In addition to the exit interview, we conducted a picture task. Interestingly, 20.8% of the children who worked with the robot chose the picture with the friends as a first picture and 21.7% as a second picture that *did* fit the robot. None of the children in the baseline condition chose this picture to fit the tablet. This seems to indicate that some children associate the robot with friends, while none of the children associate the tablet with friends. When explaining why they chose a specific picture (either that fit or that did not fit) children in the robot condition mentioned the social characteristics more often than children in the no-robot baseline condition.

As mentioned above, we cannot draw grounded conclusions about what factors explain our findings, but the exit interview and picture task seem to indicate that children perceive both systems as somewhat social. Children in the robot condition focused more often on the embodiment of the robot where they noticed that the robot moved, the size of the robot, and its voice. Also, children seem to associate the robot more often with "friends" than the tablet and focused on the social characteristics more often when choosing a picture.

Although it might be interesting from a theoretical perspective to disentangle the contribution of each of these factors, this is difficult in a real-life study, since an anthropomorphic robot that does not move or display (friendly) facial expressions might be perceived as unnatural or even scary. It seems, however, that a robot that has such a collection of aspects motivates children to provide more elaborate explanations, which is an indication of a better understanding of the learning content.

6 LIMITATIONS AND FUTURE WORK

Although we conducted this study as thoroughly as possible, there are some limitations to the results presented here.

First, in our analysis of the logical associations between facets mentioned in the children's explanations (i.e., the *depth* of their explanations), we focused only on objectively observable *explicit* verbal associations. This limited us to draw conclusions regarding how children explicitly externalise their thoughts and reasoning processes. However, it may be that children also make such associations implicitly in their internal thought processes, without expressing this verbally to the CAL system.

Second, regarding the interview, this study was done with Dutch-speaking children. In the Dutch language, it is common to refer to inanimate objects (such as the tablet) as "he" instead of "it." It is therefore difficult to interpret such responses from the children as a measure for their social reference toward the tablet or robot. This also applies to the introduction of the robot at the start of the session. Although we used the male pronoun, the robot was not strongly introduced as being male. From previous experience, we know that most children see this particular robot as male, even when it is introduced in a fully gender-neutral way [Cameron et al. 2016].

Third, the children were interviewed by the experimenter directly after the session. It is likely that children saw the experimenter as someone with a higher social status and were therefore inclined to give socially desirable answers. This could explain the high number of positive answers to the warm-up question: *Did you enjoy it?* In future research, we aim to explore additional aspects of interactive learning where the social nature of a robot may make a valuable contribution to the learning process. To make more definitive statements about the perceived social nature of the system, it is important to improve the sensitivity and validity of the measurement tools. Developing such tools specifically for young children demands an iterative approach. This article described the first iteration in the development of such a tool, consisting of a semi-structured interview and

accompanying picture selection task. Since then, based on the insights from this first iteration, a second iteration of the tool has been designed and validated to measure perceived social competence of a robot, as described in Charisi et al. [2017].

7 CONCLUSION

We conducted a study in which children carried out an inquiry learning task with either a baseline CAL system or a CAL system extended with a social robot. Children were asked to verbalize their hypotheses, observations, and conclusions out loud to the system. The main research question was *What are the effects on the explanatory behavior of children when extending a baseline CAL system with a social robot to deliver the system's feedback and questions?* Our expectation was that a robot would be more effective in eliciting verbalizations, because it is more intuitive to explain something to a social partner and the robot is a "social" entity. Results show that:

- Children verbalized longer when they explained to a robot compared to a baseline CAL system
- Children who worked with the robot touched upon more relevant facets (explanation breadth)
- Children who worked with the robot made more logical associations between the relevant facets in their explanations (explanation depth)

These results show that extending a CAL system with a social robot can have a positive effect on children's explanatory behavior and might therefore have potential for influencing children's learning behavior in new and positive ways.

APPENDIX

A WIZARD OF OZ PROTOCOL FOR BOTH CONDITIONS

Preparation Phase

Initiate state → System speech: "You can start by placing a red pot on pin 2 and a red pot on pin 5."

Response options:

- (1) Child removes blocks → Wizard presses button: "Blocks removed" → System goes to error state → System speech: "Hang on, we shouldn't remove the blocks yet. Let's put them back under the balance beam."
- (2) Child does nothing → After the timer event has run out, the button "trigger a timer event" appears. The Wizard can press this button (if child is indeed idle) → System goes to help state → System speech: "The goal is to find out what happens when you put two pots on the balance beam. Before starting with the experiment, you have to have everything ready. Let's begin by putting a red pot on pin 2 and a red pot on pin 5."
- (3) Child asks researcher or system verbally for help → Wizard presses button "offer additional help" → System goes to help state → System speech: same as (2).
- (4) Child places pots incorrectly → Wizard presses button: "Pots placed incorrectly" → System goes to initial state → System speech: repeat assignment.
- (5) Child places pots correctly → Wizard presses button: "Pots placed correctly" → System goes to success state (green continue button appears on the child's tablet) → System speech: "Good job!"

Predict Phase

Initiate state → System speech: "In a few moments you can remove the wooden blocks. But first, can you say what you think will happen to the balance beam after you take the blocks away?"

Response options:

- (1) Child removes blocks → Wizard presses button: "Blocks removed" → System goes to error state → System speech: "Hang on, we shouldn't remove the blocks yet. Let's put them back under the balance beam."
- (2) Child does nothing → After the timer event has run out, the button "trigger a timer event" appears. The Wizard can press this button (if child is indeed idle) → System goes to help state → System speech: "In research you always start with a prediction. A prediction is an explanation about what you think that will happen. So you explain what you think will happen with the balance beam after you remove the blocks. Can you say what you think will happen to the balance beam after you remove the blocks?"
- (3) Child asks researcher or system verbally for help → Wizard presses button "offer additional help" → System goes to help state → System speech: same as (2).
- (4) Child places pots incorrectly → Wizard presses button: "Pots placed incorrectly" → System goes to initial state → System speech: repeat assignment.
- (5) Child provides a response (content irrelevant) → After 2 seconds Wizard presses button: "hypothesis given" → System goes to success state (green continue button appears on the child's tablet) → System speech: "OK!"

Experiment Phase

Initiate state → System speech: "Now you are going to do the experiment. You may remove the two wooden blocks from under the balance beam".

Response options:

- (1) Child removes blocks → Wizard presses button: "Blocks removed" → System goes to success state (green continue button appears on the child's tablet) → System speech: "Well done!"
- (2) Child does nothing → After the timer event has run out, the button "trigger a timer event" appears. The Wizard can press this button (if child is indeed idle) → System goes to help state → System speech: repeat assignment.
- (3) Child asks researcher or system verbally for help → Wizard presses button "offer additional help" → System goes to help state → System speech: repeat assignment.

Observe Phase

Initiate state → System speech: "Can you say, what happened to the balance beam?"

Response options:

- (1) Child does nothing → After the timer event has run out, the button "trigger a timer event" appears. The Wizard can press this button (if child is indeed idle) → System goes to help state → System speech: "Now you have done the experiment and we have to see what happened with the balance beam. Can you say, what happened to the balance beam?"
- (2) Child asks researcher or system verbally for help → Wizard presses button "offer additional help" → System goes to help state → System speech: same as (1).
- (3) Child provides response (content irrelevant) → After 2 seconds Wizard presses button "Observation given" → System goes to success state (green continue button appears on the child's tablet) → System speech: "That's interesting!"

Conclude Phase

Initiate state → System speech: "Can you explain why the balance beam is in this position?"

Response options:

- (1) Child does nothing → After the timer event has run out, the button “trigger a timer event” appears. The Wizard can press this button (if child is indeed idle) → system goes to help state → System speech: “Now you have done an experiment. First, you put the pots on the balance beam, this was the preparation. Next, you said what you thought would happen, this was your prediction. Then, you took the wooden blocks away, that was the experiment. Finally, you saw what happened with the balance beam, that was your observation. Now we are going to try and understand what really happened. Can you explain why the balance beam is in this position?”
- (2) Child asks researcher or system verbally for help → Wizard presses button “offer additional help” → System goes to help state → System speech: same as (1)
- (3) Child provides negative response (No or I don’t know) → After 2 seconds Wizard presses button “Conclusion unknown” → System goes to success state (green continue button appears on the child’s tablet) → System speech: “OK!”
- (4) Child provides positive response (gives conclusion) → After 2 seconds Wizard presses button “Conclusion given” → System goes to success state (green continue button appears on the child’s tablet) → System speech: “That’s very interesting!”

ACKNOWLEDGMENTS

We thank the participating schools and teachers for the pleasant collaboration throughout this study, we thank the children for their enthusiasm and motivation during the activity, and we thank Marjolein van der Valk for volunteering to record the spoken voice of the system.

REFERENCES

- Paul Baxter, Emily Ashurst, Robin Read, James Kennedy, and Tony Belpaeme. 2017. Robot education peers in a situated primary school study: Personalisation promotes child learning. *PLoS One* 12, 5 (2017), e0178126.
- Amy L. Baylor and Yanghee Kim. 2005. Simulating instructional roles through pedagogical agents. *Int. J. Artif. Intell. Educ.* 15, 2 (2005), 95–115. DOI: <https://doi.org/10.1007/BF02504991>
- Amy L. Baylor, E. Shen, Daniel Warren, and P. Freire. 2004. Supporting learners with math anxiety: The impact of pedagogical agent emotional and motivational support. In *Proceedings of the Workshop on Social and Emotional Intelligence in Learning Environments at the International Conference on Intelligent Tutoring Systems (ITS’04)*. Springer, Berlin, 6–12.
- Tony Belpaeme, Paul E. Baxter, Robin Read, Rachel Wood, Heriberto Cuayáhuil, Bernd Kiefer, Stefania Racioppa, Ivana Kruijff-Korbayová, Georgios Athanasopoulos, Valentin Enescu, Rosemarijn Looije, Mark Neerinx, Yiannis Demiris, Raquel Ros-Espinoza, Aryel Beck, Lola Cañamero, Antione Hiole, Matthew Lewis, Ilaria Baroni, Marco Nalin, Piero Cosi, Giulio Paci, Fabio Tesser, Giacomo Sommariva, and Remi Humbert. 2012. Multimodal child-robot interaction: Building social bonds. *J. Hum.-Robot Interact.* 1, 2 (Dec. 2012), 33–53. DOI: <https://doi.org/10.5898/JHRI.1.2.Belpaeme>
- Tanya N. Beran and Alejandro Ramirez-Serrano. 2011. Can children have a relationship with a robot? In *Human-Robot Personal Relationships*, Maarten H. Lamers and Fons J. Verbeek (Eds.). Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Vol. 59. Springer, Berlin, 49–56. DOI: <https://doi.org/10.1007/978-3-642-19385-9>
- Tanya N. Beran, Alejandro Ramirez-Serrano, Roman Kuzyk, Sarah Nugent, and Meghann Fior. 2010. Would children help a robot in need? *International J. Soc. Robot.* 3, 1 (Sep. 2010), 83–93. DOI: <https://doi.org/10.1007/s12369-010-0074-7>
- Ann Peterson Bishop, Bertram C. Bruce, Karen J. Lunsford, M. Cameron Jones, Muzhgan Nazarova, David Linderman, Mihye Won, P. Bryan Heidorn, Rajeev Ramprakash, and André Brock. 2004. Supporting community inquiry with digital resources. *J. Dig. Inf.* 5, 3 (2004).
- Gautam Biswas, Krittaya Leelawong, Daniel Schwartz, Nancy Vye, and The Teachable Agents Group at Vande. 2005. Learning by teaching: A new agent paradigm for educational software. *Appl. Artif. Intell.* 19, 3–4 (Mar. 2005), 363–392. DOI: <https://doi.org/10.1080/08839510590910200>
- Cynthia Breazeal. 2003. Emotion and sociable humanoid robots. *Int. J. Hum.-Comput. Stud.* 59, 1 (2003), 119–155. <http://www.sciencedirect.com/science/article/pii/S1071581903000181>.
- David Cameron, Samuel Fernando, Abigail Millings, Michael Szollosy, Emily Collins, Roger Moore, Amanda Sharkey, and Tony Prescott. 2016. Congratulations, it’s a boy! Bench-marking children’s perceptions of the robokind Zeno-R25. In

- Proceedings of the Towards Autonomous Robotic Systems Conference (TAROS'16)*, L. Alboul, D. Damian, and J. Aitken (Eds.), Vol. 9716. Springer, Cham, 33–39. DOI : https://doi.org/10.1007/978-3-319-40379-3_4
- Shruti Chandra, Patricia Alves-Oliveira, Severin Lemaignan, Pedro Sequeira, Ana Paiva, and Pierre Dillenbourg. 2015. Can a child feel responsible for another in the presence of a robot in a collaborative learning activity? In *Proceedings of the 2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'15)*. IEEE, 167–172. DOI : <https://doi.org/10.1109/ROMAN.2015.7333678>
- Chih-Wei Chang, Jih-Hsien Lee, Po-Yao Chao, Chin-Yeh Wang, and Gwo-dong Chen. 2010. Exploring the possibility of using humanoid robots as instructional tools for teaching a second language in primary school. *Educ. Technol. Soc.* 13, 2 (Apr. 2010), 13–24. <https://www.questia.com/library/journal/1G1-232945833/exploring-the-possibility-of-using-humanoid-robots> <http://www.editlib.org/p/75345/>
- Vicky Charisi, D. P. Davison, F. M. Wijnen, Dennis Reidsma, and Vanessa Evers. 2017. Measuring children's perceptions of robots' social competence: design and validation. In *Proceedings of the International Conference on Social Robotics (ICSR'17)*. Lecture Notes in Computer Science, Vol. 10652. Springer, Cham, 676–686. DOI : https://doi.org/10.1007/978-3-319-70022-9_67
- Catherine C. Chase, Doris B. Chin, Marily A. Oppezzo, and Daniel L. Schwartz. 2009. Teachable agents and the protégé effect: Increasing the effort towards learning. *J. Sci. Educ. Technol.* 18, 4 (Aug. 2009), 334–352. DOI : <https://doi.org/10.1007/s10956-009-9180-4> arxiv:arXiv:1002.2562v1
- Micheline T. H. Chi, Mirloi Miriam Bassok, Matthew W. Lewis, Peter Reimann, and Robert Glaser. 1989. Self-explanations: How students study and use examples in learning to solve problems. *Cogn. Sci.* 13, 2 (Apr. 1989), 145–182. http://doi.wiley.com/10.1207/s15516709cog1302_1
- Elaine B. Coleman, Ann L. Brown, and Inna D. Rivkin. 1997. The effect of instructional explanations on learning from scientific texts. *J. Learn. Sci.* 6, 4 (Oct. 1997), 347–365. DOI : https://doi.org/10.1207/s15327809jls0604_1
- Alexandre Coninx, Paul Baxter, Elettra Oleari, Sara Bellini, Bert Bierman, Olivier Blanson Henkemans, Lola Cañamero, Piero Cosi, Valentin Enescu, Raquel Ros Espinoza, Antoine Hiolle, Rémi Humbert, Bernd Kiefer, Ivana Kruijff-Korbayová, Rosmarijn Looije, Marco Mosconi, Mark Neerincx, Giulio Paci, Georgios Patsis, Clara Pozzi, Francesca Sacchitelli, Hichem Sahli, Alberto Sanna, Giacomo Sommavilla, Fabio Tesser, Yiannis Demiris, and Tony Belpaeme. 2015. Towards long-term social child-robot interaction: Using multi-activity switching to engage young users. *J. Hum.-Robot Interact.* 5, 1 (Aug. 2015), 32. DOI : <https://doi.org/10.5898/JHRI.5.1.Coninx>
- Daniel Patrick Davison, Vasiliki Charisi, Frances Martine Wijnen, Andrea Papenmeier, Jan van der Meij, Dennis Reidsma, and Vanessa Evers. 2016. Design challenges for long-term interaction with a robot in a science classroom. In *Proceedings of the RO-MAN'16 Workshop on Long-term Child-robot Interaction*. IEEE. <https://research.utwente.nl/en/publications/design-challenges-for-long-term-interaction-with-a-robot-in-a-sci>.
- Daniel P. Davison, Frances M. Wijnen, Jan van der Meij, Dennis Reidsma, and Vanessa Evers. 2019. Designing a social robot to support children's inquiry learning: A contextual analysis of children working together at school. *Int. J. Soc. Robot.* (15 May 2019). 1–25. DOI : <https://doi.org/10.1007/s12369-019-00555-6>
- Joachim de Greeff and Tony Belpaeme. 2015. Why robots should be social: Enhancing machine learning through social human-robot interaction. *PloS one* 10, 9 (Jan. 2015), e0138061. DOI : <https://doi.org/10.1371/journal.pone.0138061>
- Pierre Dillenbourg. 1999. What do you mean by collaborative learning? *Collaborative Learning: Cognitive and Computational Approaches*, Pierre Dillenbourg (Ed.). Elsevier, 1–99. <https://telearn.archives-ouvertes.fr/hal-00190240/document>.
- Samuel Fernando, Roger K. Moore, David Cameron, Emily C. Collins, Abigail Millings, Amanda J. Sharkey, and Tony J. Prescott. 2016. Automatic recognition of child speech for robotic applications in noisy environments. (Nov. 2016) arxiv:1611.02695. Retrieved from <http://arxiv.org/abs/1611.02695>.
- A. Gulz. 2005. Social enrichment by virtual characters - differential benefits. *J. Comput. Assist. Learn.* 21, 6 (Nov. 2005), 405–418. DOI : <https://doi.org/10.1111/j.1365-2729.2005.00147.x>
- Y. Hayashi, M. Matsumoto, and H. Ogawa. 2012. Pedagogical agents that support learning by explaining: Effects of affective feedback. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society*. 1650–1655. <http://escholarship.org/uc/item/3q92z8bz.pdf>, <http://palm.mindmodeling.org/cogsci2012/papers/0291/paper0291.pdf>.
- Jeffrey Holmes. 2007. Designing agents to support learning by explaining. *Comput. Educ.* 48, 4 (May 2007), 523–547. DOI : <https://doi.org/10.1016/j.compedu.2005.02.007>
- Deanna Hood, Séverin Lemaignan, and Pierre Dillenbourg. 2015. When children teach a robot to write: an autonomous teachable humanoid which uses simulated handwriting. In *Proceedings of the International Conference for Human-Robot Interaction (HRI'15)*, 83–90. DOI : <https://doi.org/10.1145/2696454.2696479>
- Eun-ja Hyun, So-yeon Kim, Siekyung Jang, and Sungju Park. 2008. Comparative study of effects of language instruction program using intelligence robot and multimedia on linguistic ability of young children. In *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'08)*. IEEE, 187–192.
- W. Lewis Johnson, Jeff W. Rickel, and James C. Lester. 2000. Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *Int. J. Artif. Intell. Educ.* 11, 1 (2000), 47–78. <https://pdfs.semanticscholar.org/d5ca/5efdecba2707aab57d4cb2df008a7ecdd0dd.pdf>.

- Takayuki Kanda, Takayuki Hirano, Daniel Eaton, and Hiroshi Ishiguro. 2004. Interactive robots as social partners and peer tutors for children: A field trial. *Hum.-Comput. Interact.* 19, 1 (Jun. 2004), 61–84. <http://dl.acm.org/citation.cfm?id=1466547.1466551>.
- Takayuki Kanda, Michihiro Shimada, and Satoshi Koizumi. 2012. Children learning with a social robot. In *Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI'12)*. ACM Press, New York, NY, 351. DOI : <https://doi.org/10.1145/2157689.2157809>
- Paul A. Kirschner, John Sweller, and Richard E. Clark. 2006. Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educ. Psychol.* 41, 2 (Jun. 2006), 75–86. DOI : https://doi.org/10.1207/s15326985ep4102_1 arxiv:0710.4428v1
- Stefan Kopp, Brigitte Krenn, Stacy Marsella, A. Marshall, Catherine Pelachaud, Hannes Pirker, K. Thórisson, and Hannes Vilhjálmsson. 2006. Towards a common framework for multimodal generation: The behavior markup language. *Intell. Virt. Agents* 52, 6 (2006), 205–217. DOI : <https://doi.org/10.1007/11821830>
- Leema Kuhn and Brian Reiser. 2005. Students constructing and defending evidence-based scientific explanations. In *Proceedings of the Annual Meeting of the National Association for Research in Science Teaching*. Citeseer, 1–35.
- Chen Lin C. Kulik and James A. Kulik. 1991. Effectiveness of computer-based instruction: An updated analysis. *Comput. Hum. Behav.* 7, 1–2 (Jan. 1991), 75–94. DOI : [https://doi.org/10.1016/0747-5632\(91\)90030-5](https://doi.org/10.1016/0747-5632(91)90030-5) arxiv:arXiv:1011.1669v3
- Ard W. Lazonder. 2014. Inquiry learning. In *Handbook of Research on Educational Communications and Technology*, J. Michael Spector, M. David Merrill, Jan Elen, and M. J. Bishop (Eds.). Springer, New York, NY, 453–464. DOI : <https://doi.org/10.1007/978-1-4614-3185-5>
- Krittaya Leelawong and Gautam Biswas. 2008. Designing learning by teaching agents: The betty’s brain system. *Int. J. Artif. Intell. Educ.* 18, 3 (2008), 181–208. DOI : <https://dl.acm.org/citation.cfm?id=1454280>
- Daniel Leyzberg, Samuel Spaulding, Mariya Toneva, and Brian Scassellati. 2012. The physical presence of a robot tutor increases cognitive learning gains. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 34. <https://cloudfront.escholarship.org/dist/prd/content/qt7ck0p200/qt7ck0p200.pdf>.
- Richard E. Mayer. 2004. Should there be a three-strikes rule against pure discovery learning? *Am. Psychol.* 59, 1 (2004), 14–19. DOI : <https://doi.org/10.1037/0003-066X.59.1.14>
- Omar Mubin, Catherine J. Stevens, Suleman Shahid, Abdullah Al Mahmud, and Jian-Jie Dong. 2013. A review of the applicability of robots in education. *Technol. Educ. Learn.* 1, 1 (Jan. 2013). DOI : <https://doi.org/10.2316/Journal.209.2013.1.209-0015>
- NSF. 2000. An introduction to inquiry. In *Inquiry: Thoughts, Views and Strategies for K-5 Classroom 2* (2000), 1–5.
- Sandra Y. Okita, Victor Ng-Thow-Hing, and Ravi Sarvadevabhatla. 2009. Learning together: ASIMO developing an interactive learning partnership with children. In *Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'09)*. IEEE, 1125–1130. DOI : <https://doi.org/10.1109/ROMAN.2009.5326135>
- R Ploetzner, P. Dillenbourg, M. Preier, and D. Traum. 1999. Learning by explaining to oneself and to others. *Collab. Learn.: Cogn. Comput. Appr.* 1 (1999), 103–121. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.78.2701&rep=rep1&type=pdf>.
- Aaron Powers, Adam D. I. Kramer, Shirlene Lim, Jean Kuo, Sau Lai Lee, and Sara Kiesler. 2005. Eliciting information from people with a gendered humanoid robot. In *Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication*, Vol. 2005. IEEE, 158–163. DOI : <https://doi.org/10.1109/ROMAN.2005.1513773>
- R Development Core Team. 2008. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- Aditi Ramachandran, Chien-Ming Huang, Edward Gartland, and Brian Scassellati. 2018. Thinking aloud with a tutoring robot to enhance learning. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 59–68.
- Dennis Reidsma and Herwin van Welbergen. 2013. AsapRealizer in practice – A modular and extensible architecture for a BML realizer. *Entertain. Comput.* 4, 3 (Aug. 2013), 157–169. DOI : <https://doi.org/10.1016/j.entcom.2013.05.001>
- Bethany Rittle-Johnson, Megan Saylor, and Kathryn E. Swygert. 2008. Learning from explaining: Does it matter if mom is listening? *J. Exp. Child Psychol.* 100, 3 (Jul. 2008), 215–224. DOI : <https://doi.org/10.1016/j.jecp.2007.10.002>
- Cynthia A. Rohrbeck, Marika D. Ginsburg-Block, John W. Fantuzzo, and Traci R. Miller. 2003. Peer-assisted learning interventions with elementary school students: A meta-analytic review. *J. Educ. Psy.* 95, 2 (June 2003), 240–257. DOI : <https://doi.org/10.1037/0022-0663.95.2.240>
- Martin Saerbeck, Tom Schut, Christoph Bartneck, and Maddy D. Janse. 2010. Expressive robots in education. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems (CHI'10)*. ACM Press, New York, NY, 1613. DOI : <https://doi.org/10.1145/1753326.1753567>
- Katie S. Saunders-Stewart, Petra D. T. Gyles, and Bruce M. Shore. 2012. Student outcomes in inquiry instruction. *J. Adv. Acad.* 23, 1 (Feb. 2012), 5–31. DOI : <https://doi.org/10.1177/1932202X11429860>

- Namin Shin and Sangah Kim. 2007. Learning about, from, and with robots: Students' perspectives. In *Proceedings of the 16th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'07)*. IEEE, 1040–1045. DOI: <https://doi.org/10.1109/ROMAN.2007.4415235>
- Han Sloetjes and Peter Wittenburg. 2008. Annotation by category—ELAN and ISO DCR. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC'08)*, 816–820. DOI: <https://doi.org/10.3758/BRM.41.3.591>
- Fumihide Tanaka, Aaron Cicourel, and Javier R. Movellan. 2007. Socialization between toddlers and robots at an early childhood education center. *Proc. Natl. Acad. Sci. U.S.A.* 104, 46 (Nov. 2007), 17954–8. DOI: <https://doi.org/10.1073/pnas.0707769104>
- Fumihide Tanaka and Shizuko Matsuzoe. 2012. Children teach a care-receiving robot to promote their learning: field experiments in a classroom for vocabulary learning. *Hum-Rob Inter.* 1, 1 (July 2012), 78–95. <http://hri-journal.org/index.php/HRI/article/view/12>.
- Hans van der Meij, Jan van der Meij, and Ruth Harmsen. 2015. Animated pedagogical agents effects on enhancing student motivation and learning in a science inquiry learning environment. *Educ. Technol. Res. Dev.* 63, 3 (Jun. 2015), 381–403. DOI: <https://doi.org/10.1007/s11423-015-9378-5>
- Wouter R. van Joolingen and Ton de Jong. 1997. An extended dual search space model of scientific discovery learning. *Instruct. Sci.* 25, 5 (1997), 307–346. DOI: <https://doi.org/10.1023/A:1002993406499>
- W. R. Van Joolingen, T. De Jong, and A. Dimitrakopoulou. 2007. Issues in computer supported inquiry learning in science. *J. Comput. Assist. Learn.* 23, 2 (Jan. 2007), 111–119. DOI: <https://doi.org/10.1111/j.1365-2729.2006.00216.x>
- K. VanLehn and R. M. Jones. 1993. Learning by explaining examples to oneself: A computational model. In *Foundations of Knowledge Acquisition: Cognitive Models of Complex Learning*, Susan Chipman and Alan L. Meyrowitz (Eds.). Vol. 194. Springer USA, Boston, MA, 25–82. DOI: <https://doi.org/10.1007/978-1-4615-3172-2>
- L. S. Vygotsky. 1978. *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press, Cambridge, MA. <https://books.google.com/books?hl=nl&lr=&id=Irq913IEZ1QC&pgis=1>.
- Frances Wijnen, Vicky Charisi, D. P. Daniel Davison, J. van der Meij, Dennis Reidsma, and Vanessa Evers. 2015. Inquiry learning with a social robot: Can you explain that to me? In *Proceedings of the 1st International Conference on Social Robotics in Therapy and Education (New Friends'15)*. Windesheim Flevoland. <http://eprints.eemcs.utwente.nl/26625/>.
- David Wood, Jerome S. Bruner, and Gail Ross. 1976. The role of tutoring in problem solving. *J. Child Psychol. Psychiat.* 17, 2 (Apr. 1976), 89–100. DOI: <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>
- Beverly Park Woolf. 2010. *Building Intelligent Interactive Tutors: Student-centered Strategies for Revolutionizing e-learning*. Morgan Kaufmann Publishers/Elsevier. 480 pages. DOI: <https://doi.org/10.1007/BF02680460>
- Cristina Zaga, Roelof A. J. de Vries, Jany Li, Khiet P. Truong, and Vanessa Evers. 2017. A simple nod of the head: The effect of minimal robot movements on children's perception of a low-anthropomorphic robot. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 336–341.
- Zhen-Jia You, Chi-Yuh Shen, Chih-Wei Chang, Baw-Jhiune Liu, and Gwo-Dong Chen. 2006. A robot as a teaching assistant in an english class. In *Proceedings of the 6th IEEE International Conference on Advanced Learning Technologies (ICALT'06)*. IEEE, 87–91. DOI: <https://doi.org/10.1109/ICALT.2006.1652373>

Received June 2018; revised May 2019; accepted July 2019