

Working with a Social Robot in School: A Long-Term Real-World Unsupervised Deployment

Daniel P. Davison
University of Twente
Enschede, The Netherlands
d.p.davison@utwente.nl

Frances M. Wijnen
University of Twente
Enschede, The Netherlands
f.m.wijnen@utwente.nl

Vicky Charisi
University of Twente
Enschede, The Netherlands
v.charisi@utwente.nl

Jan van der Meij
Het Erasmus
Almelo, The Netherlands
j.v.d.meij@het-erasmus.nl

Vanessa Evers
Nanyang Technological University
Singapore, Singapore
vanessa.evers@ntu.edu.sg

Dennis Reidsma
University of Twente
Enschede, The Netherlands
d.reidsma@utwente.nl

ABSTRACT

Interactive learning technologies, such as robots, increasingly find their way into schools. However, more research is needed to see how children might work with such systems in the future. This paper presents the unsupervised, four month deployment of a Robot-Extended Computer Assisted Learning (RECAL) system with 61 children working in their own classroom. Using automatically collected quantitative data we discuss how their usage patterns and self-regulated learning process developed throughout the study.

CCS CONCEPTS

• **Applied computing** → **Interactive learning environments**;
• **Computer systems organization** → *Robotics*; • **Human-centered computing** → *Field studies*; • **Social and professional topics** → *Children*; • **Hardware** → Sensors and actuators.

KEYWORDS

Child-robot interaction, long-term, in the wild, unsupervised interaction, user study, interaction design, social robot, inquiry learning, sensorised learning materials, computer assisted learning system

ACM Reference Format:

Daniel P. Davison, Frances M. Wijnen, Vicky Charisi, Jan van der Meij, Vanessa Evers, and Dennis Reidsma. 2020. Working with a Social Robot in School: A Long-Term Real-World Unsupervised Deployment. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI'20)*, March 23–26, 2020, Cambridge, United Kingdom.. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3319502.3374803>

1 INTRODUCTION

Robots increasingly find their way into classrooms. To support long-term interactions they should be capable of offering learning tasks that move and grow with the development of the child and therefore can be accessed across recurring interactions while remaining relevant to the children, while the robot behaviour should

Permission to make digital or hard copies of part or all 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 third-party components of this work must be honored. For all other uses, contact the owner/author(s).

HRI '20, March 23–26, 2020, Cambridge, United Kingdom.

© 2020 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6746-2/20/03.

<https://doi.org/10.1145/3319502.3374803>

stay relevant and engaging over time [6]. To understand the exact design parameters it is important to gain insight into how children interact with HRI systems, day-to-day in their own classrooms. However, longitudinal in-the-wild studies with educational robots are relatively scarce in comparison to the large number of short term studies. A recent review on educational robots found 101 papers with 309 study results [2], but a survey on only longitudinal interactions [24] reported a mere 8 studies. We discuss an additional 6 studies, but in a review on evaluation methods we also observed a substantial lack of long term studies, among other things related to practical challenges in long term classroom research and the technological readiness of robotic systems [6]. The paucity of long term studies is a real problem. It is not possible to say whether—and how exactly—insights about CRI transfer from the lab to the real world and to what extent lab studies have succeeded in capturing the richness of interactions as they would develop over time [15].

To advance the field of educational HRI, more studies are needed on how real users interact with robots and learning materials; unsupervised, in-the-wild, and over extended periods of time. Yet, with the current state of the art in HRI, we do not yet know well enough whether it is feasible to run long term HRI experiments in the classroom on the basis of a realistic behaviour and task repertoire.

In this paper we address the need for long term HRI studies. We conducted a four-month study in which 61 children worked unsupervised, in their classroom, on two learning tasks with an autonomous Robot-Extended Computer Assisted Learning (RECAL) system. Three different versions were deployed in three classes; one with a robot and two without. Embedded sensors in the learning materials allowed our system to automatically guide the learning process, while simultaneously recording children's interactions with the system. Through these logbooks we show how their learning and their sustained interactions with the system developed throughout the four months, as increasingly challenging learning content unfolded. We examine the extent to which the children could self-regulate this process. Finally, we also share our approach to addressing specific challenges that arose in practice during our study.

This study was part of the EU EASEL project. The same data collection was used in a study on the effects of effort-related feedback on children's mindset in learning, to be reported elsewhere. Additionally, we conducted semi-structured interviews with the children to capture their subjective experiences; here, we present illustrative remarks from these while details will be presented elsewhere.

2 BACKGROUND AND RELATED WORK

There can be significant differences in the way a user interacts with a robot for the first time, or after a habituation period [24]. Habituation effects are already seen in very young children when presented with novel stimuli [13]. Tanaka et al. [37–39] have shown that after a familiarisation period, children will have higher quality interactions with a robot as they become accustomed to it; a recognisable routine could play a role here [38]. In order to maintain a child's interest and engagement the robot should exhibit individual adaptability [1, 16, 17, 24]. Yet, others reported that 24 months after a disability unit at a school introduced robots in their curriculum, focus groups showed that people still reported generally positive about the Paro robot that has no adaptability whatsoever [34].

Different contexts require particular attention to the design of the robot. In a comprehensive survey of 24 papers on long term studies, Leite et al. [24] found that real-world long-term HRI commonly take place in work environments and public spaces, in health care and therapy, in users' homes, and in educational settings. Users generally accept robots and are willing to interact with them multiple times over a period of time. The number of sessions in studies varied considerably between 2-180 (m: 32, sd: 45); most studies reported between 5-20 sessions. Sessions were typically between 6-90 minutes long (m: 38, sd: 30) and took place over a period of roughly several weeks to several months (m: 11 weeks, sd: 8). Leite et al. [24] discuss several longitudinal effects such as strengthened social interactions and improved well-being of elderly; positive effects in autism therapies; and their potential role in behaviour change support systems. Over time, people may anthropomorphise the robot; they tend to form mental models regarding its behaviour and they may change their preferences regarding interaction and usage patterns. However, the robot did not always live up to users' initial expectations and during longer studies the importance of regular maintenance became increasingly evident. Finally, Leite et al. [24] identified guidelines for long-term studies regarding the robot's appearance, the continuity and incremental novelty of its behaviours, its empathy and affective interactions, and its memory and adaptation.

In long term educational settings, the pioneering work by Kanda et al. [16] investigated how Japanese children interacted with an English speaking robot for two weeks. Their results showed a sharp decrease in frequency of interactions after the first week. They suggest that the robot may not have lived up to children's initial expectations. Similar effects were found by Leite et al. [25] in an exploratory study on children playing chess with an iCat robot. As time passed, children gave the robot lower social presence ratings and looked at it less frequently. Building on such insights, researchers explored ways in which the interaction could be sustained beyond the novelty effects.

In a follow-up study Leite et al. [23] extended their chess-playing iCat robot with a memory of past interactions and a model for empathic responses. The empathic robot seemed to be better at maintaining the interaction over multiple sessions, being able to sustain high levels of perceived presence, engagement, helpfulness and children's self-validation throughout five interactions. Similarly, Kanda et al. [17] investigated whether they could retain children's interest by extending their robot with capabilities aimed at building social relationships. In particular, their robot incrementally expanded its

behaviour repertoire and confided personal matters to children who interacted with it more frequently. Their results suggest that children who saw the robot as a peer, interacted with it for the full two-month period, forming a friendly relationship, while children who saw it more as a tool, became bored after about five weeks.

More recent work has continued to explore how to keep children engaged. For instance, Coninx et al. [7] described a system that supports switching between multiple activities within single sessions, while maintaining a consistent personalised behaviour profile. Their results suggested that such a multi-activity approach may be a promising way to sustain longer interactions.

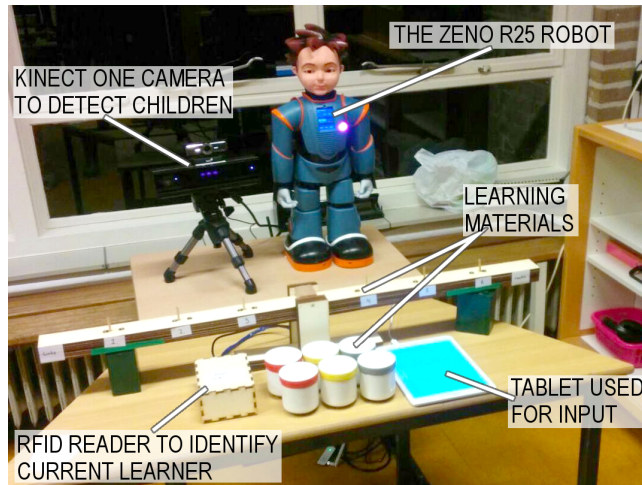
A different approach was taken by Jacq et al. [14], who explored improving children's handwriting through the use of a teachable robot that learned from demonstrations and corrections from the child. Children successfully took on this teaching role and built an affective bond with the robot, being engaged with this teaching process and motivated to focus on the task and pay attention throughout multiple lengthy sessions. Chandra et al. [5] found that children's perception of the robot's ability to learn was formed over time, and suggest that such perceptions may affect the child's own learning: children showed more improvements in their handwriting when tutoring a robot actually capable of learning.

Robots as tutors can also more directly instruct the child. Serholt and Barendregt [33] investigated how children's engagement with such a robot developed over time, looking at moments in the interaction where it initiated a social cue that was outside the scope of the current task. Overall, children seemed to be socially engaged with the robot during such cues. The robot consistently elicited gazes, facial expressions, and verbal responses when greeting or asking questions; indications were found that children were affected by the robot's praise. However, these patterns seemed to decrease over the course of several interactions, potentially also due to some disappointment in the robot's limited interaction capabilities. Gordon et al. [12] describe a platform aimed at personalising the robot's tutoring behaviours to the affective state of the child in a second language learning task. Overall, children learned new foreign words after several interactions; and they generally responded with more positive valence when working with the robot that offered personalised responses.

Some longer term studies with robots involved younger children. Tanaka et al. [37] looked at social interactions between toddlers and a humanoid robot. They show that, over time, children developed care-taking behaviours towards the robot, progressively treating it more as a peer than as a toy. Kozima et al. [22] found that preschool children displayed similar care-giving behaviours towards a robot with a more minimalistic design. Children were initially shy around the robot during the first sessions; in subsequent sessions many children continued interacting with the robot throughout the study and would often play with it together in groups.

These papers show beneficial effects of repeated interactions with educational robots, but most studies report that engagement declines considerably over time. To counter this, it is often suggested that the system should gradually show novel behaviours or increasingly varied interactions, but Kennedy et al. [18] showed that increased variations in social behaviours could be detrimental to the core learning process. We explored one aspect of this balancing act by investigating how children progress through two

Figure 1: The system with the balance scale task as it was deployed in one of the participating classrooms.



different consecutive learning tasks, while working with an interactive system that displays the same consistent set of behaviours throughout. We try to get a first feeling for the quantity and variation of learning content that is required for sustaining engagement for prolonged periods, and thus for the feasibility of performing long term HRI studies that may yet win past the novelty effect.

3 METHODS AND APPROACH

We aimed to analyse the emergent child-robot interaction in a long term unsupervised deployment of our RECAL system, and to share lessons learned from this deployment study. The study involved 61 children in three classrooms in different schools and took place over a duration of approximately four months.

Although a comparative study between robot and non-robot conditions was not our aim, we were still interested in first insights regarding whether the robot was helpful, damaging, or neither to the long term interaction. We know that robots in this kind of task setting can provide real value, transforming the way the children learn [42]. But we do not know whether in the longer run the robot is still helpful, or might in fact detract from the experience once the novelty of the robot wears off and disenchantment sets in.

Therefore, three variants were deployed in the three classrooms: 1) without robot, a baseline version that offered interactive task instructions and feedback, 2) without robot, a version that additionally offered personalised praise, and 3) a version where instructions, feedback, and personalised praise were delivered through a robot.

We consider the similarities and differences in usage patterns across the three versions. We also offer broader insights into how children’s interactions with the setup developed over the course of the study by analysing the aggregated data set.

3.1 Materials and System Overview

The technical setup was developed to offer a realistic interaction that was meaningful in the learning context of the particular schools and could be accessed without supervision. The setup was based on earlier work [8, 30] and consisted of several key components: two tangible learning tasks with embedded Arduinos and sensors, a

tablet interface, a robot, and a desktop machine running the RECAL system that orchestrated the fully autonomous interaction with the children in the learning task, as shown in Figure 1. Our robot was the Zeno R25 from Robokind¹, an approximately 50cm tall humanoid robot with expressive facial capabilities. Audio was played through headphones, to not disturb other children working in the classroom.

3.1.1 Learning Tasks. The learning tasks were constructed according to principles of inquiry learning to support a scientific process of discovery [19, 20]. In the first task, adapted from our earlier work [9, 42], children used a balance scale to explore the moment of force. They placed combinations of three differently weighted pots on three distances from a central fulcrum to discover how those variables affected the tilt of the balance. The second task, newly developed for this study, consisted of two sloped ramps from which children could race balls to explore potential energy and rolling resistance, while the angle of each slope could be adjusted to a low or high position. The children could select balls from a combination of materials, weights, and densities: small marbles; medium pingpong balls, wooden balls, and rubber balls; and large styrofoam balls and marbles. By racing them head-to-head, they discovered how their speed was influenced by these variables.

3.1.2 Feedback and Instructions: Navigating the Difficulty Levels. The tasks had assignments for 8 or 9 levels of difficulty, through which children could progress at their own pace. As they did so, the system offered verbal task-related instructions. In the lower levels, children were given precise preparatory instructions and were asked for their prediction regarding the outcome (e.g. “place a yellow pot on location 1 and a red pot on location 5, which way do you think it will tilt?”). Higher level assignments were more open-ended (e.g. “find all the balls that are faster than a pingpong ball”). The system offered personalised feedback and praise based on the child’s progress, whether or not they gave a correct answer, and at what level they were doing the task.

To offer appropriate feedback, and to gather more insights into the children’s self reported perception of the task and of their own performance, the system also regularly asked the children (through an app) to comment on their own effort and performance (how hard they worked and whether they thought they gave the right answers), the perceived difficulty of the task, and the desired next assignment difficulty, through in-task multiple-choice questions.

The difficulty levels, feedback, praise, and children’s subjective self-reporting were designed to facilitate children’s self-regulation of their progress through the difficulty levels, making the tasks build up nicely over time to facilitate longer term use.

3.1.3 Modules for Data Gathering. Data was gathered to carry out the task in an unsupervised way. To detect *physical manipulations of the learning materials*, they were enriched with embedded sensors. The sensors in the balance scale could detect the tilt of the scale, the placement of pots, and whether the supporting blocks had been removed. The sensors in the rolling ball ramp could detect the angle of each slope, when to release the balls after pressing a button, and the time that it took each ball to reach the finish after release. To get *direct input from the children*, the system used a tablet with Android 7 running our custom educational app. Besides displaying

¹Robokind: <https://www.robokind.com>

multi-modal task instructions, this app gathered children’s answers to exercises, as well as their responses about self reported difficulty and effort (see above). To *identify which child was interacting with the setup*, children could scan their RFID card. Additionally, for *sensing the presence and location of the child* (e.g., for robot gaze) we used a Microsoft Kinect One with SceneAnalyzer software [44].

Using these sources of data, the system automatically measured for every child how often, how long, and at which level each child worked on the task. Together with the various self reported subjective data regarding difficulty and effort, this delivered a lot of insight in how children navigated the difficulty levels of the tasks.

3.1.4 Robust Interaction Management. All devices were connected to a small desktop Windows 10 PC, hidden out of view, running the RECAL system which orchestrated the interaction based on the various inputs described above. The interaction manager (IM) was based on Flipper, a rule-based information state dialogue engine [27, 40]. The states were modelled around the progression of inquiry phases of each assignment, including error states (used when the child did not follow the proper steps), success states (when a phase was finished and the exercise could proceed to the next phase), and states for the subsequent steps within each phase (explanation by the system; actions by the children). A detailed explanation of this IM can be found in [42]. Once an exercise was finished and the child wanted to perform another one, the IM would select the (difficulty of the) next exercise based on the child’s preference.

We designed our IM for robust multi-modal interaction to enable our system to operate autonomously in a real classroom environment. Rather than capturing crucial user input through state-of-the-art perception and reasoning models, which are often susceptible to environmental and human factors, we designed our core interactions around the tablet interface. Whenever the IM expected input from the child it would display multiple-choice buttons or would display instructions for operating the learning materials. When no actions or responses were detected, the system used timers to repeat instructions or offer additional help. Additionally, the system dealt with various contingencies in the interaction (e.g. user walking away mid-task). Core interactions were extended with more elaborate behaviours to further enrich the learning experience. Through progressive enhancement and graceful degradation² techniques we designed additional dynamic behaviours, such as reactive and deictic gaze, depending on whether the system’s more advanced sensors were available and were returning reliable values. The core system would continue to operate as best as possible in the event of a malfunctioning module, which would then be automatically restarted at an appropriate point in the interaction.

3.1.5 Robot Behaviour Realisation. The IM requested spoken utterances, robot movement, and tablet interface updates, specified in the Behaviour Markup Language (BML) [3, 21]. The robot could display facial expressions (happiness, surprise), lip synchronisation to speech, interactive gaze (to the child, tablet, or learning materials), and life-like behaviours (blinking). The exact behaviour design of the robot was informed by design guidelines emerging from an extensive contextual analysis of inquiry learning tasks with our

²The terms *graceful degradation* and *progressive enhancement* have long been used in different fields, like web development [41].

	Age	Nr. of participants
	Mean (SD)	Total (girls/boys)
Freinet - Class 1	8.8 (0.73)	17 (12/5)
Montessori - Class 2	7.1 (0.83)	24 (10/14)
Montessori - Class 3	7 (0.82)	20 (8/12)

Table 1: Participant demographics

target user group [9]. The generated BML was sent to ASAPRealizer [31], which further orchestrated the scheduling, planning, and execution of the behaviours; the system’s speech was generated using the Fluency³ text-to-speech (TTS) engine.

3.2 Participants

The three versions of the system were deployed in parallel in two district locations of the same Montessori school and one classroom of a Freinet school, all located in similar suburbs of the same city. In the Montessori and Freinet educational systems children of mixed age groups learn together in the same classroom. There were 61 participating children between 6-10 years old, as described in Table 1.

Ethical approval was obtained from the ethical board of the University of Twente and parents signed an informed consent letter at the start of the study. Prior to this study it was agreed with the school directors that the learning tasks would fall under the school’s regular science education curriculum. Children who did not obtain consent from their parents still had the option to work with the learning tasks in order to benefit from the educational content. However, in such cases our system collected no data and we conducted no interviews with these children.

3.3 Procedure

The technological setups were installed in each classroom outside of school hours. The next morning, a researcher was briefly present to give an introduction of the system to highlight the various components and illustrate how to initiate a session. Children could work individually with the task on their own initiative and without supervision. All children started on level 3 of 8. The system greeted each child personally by name, after which the tablet displayed textual and visual instructions for completing the assignment. Children could press a “read aloud” button to have the system read the instructions out loud. After each completed assignment, the system asked them whether or not they got the correct answer and if they thought the task was “easy”, “ok”, or “hard”. The system then asked the child if they wanted to continue and whether they wanted to do an easier, equally difficult, or harder assignment. It then loaded a next assignment level according to their selected choice. The system would end the session after four completed assignments. The first task was removed by experimenters after approximately 6-7 weeks.

Two classes had a holiday just after the first task; after that, the experimenters installed and introduced the second task in each class. Now, all children started on level 1 of 9. Although the learning content of this task was different than the first, the interaction procedure and system behaviours were identical. After approximately 8-10 weeks the materials were removed and the study was ended.

³Fluency text-to-speech engine for Dutch speech: <https://www.fluency.nl/tts/>

3.4 Measurements

Measurements were collected automatically on all weekdays, as described in Section 3.1.3. This included the quantitative information on how our setups were used in practice as well as the children's self-reported task performance, subjective assignment difficulty, and their preferred next assignment difficulty. The data was analysed to gain insights on how their usage patterns evolved over time, how they progressed through the available assignment levels, and how well they were able to self-regulate this process.

4 LONGITUDINAL DEPLOYMENT IN PRACTICE: RESULTS

Overall usage statistics per task for each class are shown in Table 2. The first learning task, in which children explored the moment of force with a balance scale, was in classes for around 6 to 7 weeks. During that time the 61 children completed a total of 1131 assignments during 371 interaction sessions and worked with the system for an average of 6 minutes per session ($sd = 3:03$). The majority of children interacted 6 times or more, with several outliers of up to 17 interaction sessions ($m = 6$, $sd = 4$). To compare differences between classes we conducted pairwise Wilcoxon rank sum tests with Bonferroni correction ($\alpha = 0.0083$). All tests were two tailed. We found that children in class 3 (with robot and praise) initiated significantly more sessions ($m = 7.9$, $sd = 4.5$) than children in class 2 (without robot, with praise) ($m = 4.5$, $sd = 4.2$) ($W = 386$, $p = 0.0062$). No significant differences were found between class 1 (without robot and praise) ($m = 6$, $sd = 2.3$) and class 2 ($W = 137$, $p = 0.017$) nor between class 1 and class 3 ($W = 243$, $p = 0.24$). Figure 2 shows a breakdown of sessions and unique users per week for all classes combined. The system was used most frequently during the first weeks, with some individual children even interacting multiple times per week. Overall usage generally declined towards the end; yet we saw that a small number of unique children continued to interact with the system each week. For class 2 and class 3 this decline occurred after roughly 3-4 weeks. However, class 1 had a holiday in the fourth week, during which there were no interactions. The system was deployed in this class for one week longer as compensation; after the holiday we saw a somewhat sustained usage towards the end.

The second learning task, in which children used a ramp to discover potential energy and rolling resistance, was in classes for around 8-10 weeks. Children completed a total of 785 assignments during 274 sessions, with an average session length of 6:44 minutes per session ($sd = 3:02$ minutes). The majority of children initiated four or more sessions ($m = 4.5$, $sd = 2.9$). Similar to the first task, we found that children in class 3 initiated significantly more sessions ($m = 5.8$, $sd = 3.3$) than children in class 2 ($m = 3.1$, $sd = 2.6$) ($W = 346$, $p = 0.0047$). Again, no significant differences were found between class 1 ($m = 4.4$, $sd = 1.8$) and class 2 ($W = 93$, $p = 0.01$) nor between class 1 and class 3 ($W = 260$, $p = 0.17$). The second task was deployed in the schools longer than the first task because several disruptive events took place. Most notable was a two-week holiday, which took place in week 2-3 for class 1, week 5-6 for class 2, and week 6-7 for class 3. Furthermore, towards the end of the school year children were often busy studying and taking exams and finishing other school assignments. Additionally, class 3 was away practising for a musical in the week before their holiday. We found similar

weekly interaction patterns as during the first task; it was popular at the beginning and the number of interactions declined towards the end, although we saw no full abandonment during the study. However, it seems that the decline occurred sooner, after around 2-3 weeks. In week 6, 7, and 8 we saw a slight increase after children returned to school from their holidays. In both tasks we found that class 3 had more than double the number of stopped sessions.

The other measurements regarding self-reported measures and self-regulated pacing revealed no significant differences. We discuss these remaining results using the aggregated data from all children.

In their first interaction with task 1, all children started at level 3. Although children occasionally opted for lower levels, they typically progressed to higher levels as they repeatedly interacted with the system. The number of assignments that were completed were not uniformly distributed across the 8 available levels (mean = 141 assignments per level, $sd = 108$). This suggests that there may have been a ceiling effect. We found that the highest level was reached by 38 children and was, by far, used most frequently; around 20% of all assignments were completed at this level. Closer investigation reveals that this ceiling effect probably surfaced at around 4-6 interaction sessions, as shown in Figure 3.

In the second task children started at level 1, after which they progressed at their own pace in subsequent interactions. In contrast to the first task, the number of assignments completed at each level was more uniformly distributed (mean = 87 assignments per level, $sd = 16$). This suggests that the ceiling effect was less pronounced and seemed to occur later after around 6-7 sessions, as shown in Figure 3. This may be due to the fewer number of repeated interactions that took place. Many children didn't interact frequently enough to exhaust all available levels; only 16 children reached the highest level. It may also be due to the more open nature of the higher levels, where children could repeat the experiment many times in different combinations before arriving at a correct answer.

Regarding self-reported measures, we found no significant differences between the two tasks or the levels. Overall, children reported 73% of their answers to be correct. Similarly, we found no significant differences between the various levels regarding self-reported subjective difficulty, although the higher levels tended to be rated as "hard" slightly more often than the lower levels. However, we found that children seemed to rate assignments that they had correct as being easier. In those cases, they rated 66% as "easy", 22% as "ok", and 12% as "hard". Assignments they got wrong, on the other hand, were rated 22% as "easy", 48% as "ok", and 30% as "hard". Furthermore, after getting an assignment wrong, children seemed to be more inclined to select a lower level or stick to the same level for their next assignment (39%), compared to when they got the assignment right (26%). These results suggest that the consecutive levels of the tasks were appropriately challenging and that children were able to effectively regulate their own progression.

In semi-structured interviews we asked children to reflect on their experience with the system. Regarding the learning tasks, children enjoyed doing everything by themselves, like formulating and testing predictions and selecting their own difficulty level. When asked about the tablet, they said that they could press buttons to select their answers, that it displayed instructions, helped them by reading assignments out loud, and could recognise their actions and their answers. Additionally, children who worked with the

	System version		Total	Total	Total	Completed	Stopped	Abandoned
	Robot	Praise	weekdays	assignments	sessions	sessions	sessions	sessions
Task 1 - Balance scale								
Class 1	no	no	37	375	111	73	38	0
Class 2	no	yes	30	311	103	56	35	12
Class 3	yes	yes	29	445	157	60	87	10
Task 2 - Ramp								
Class 1	no	no	39	235	79	33	29	17
Class 2	no	yes	42	192	62	29	9	24
Class 3	yes	yes	47	358	133	33	63	37

Table 2: Number of weekdays and total number of assignments and sessions in each of the classes. A session was considered completed if the child had finished four assignments. A session was considered stopped when a child indicated to the system they did not wish to continue after completing an assignment. A session was considered abandoned if a child walked away in the middle of a assignment.

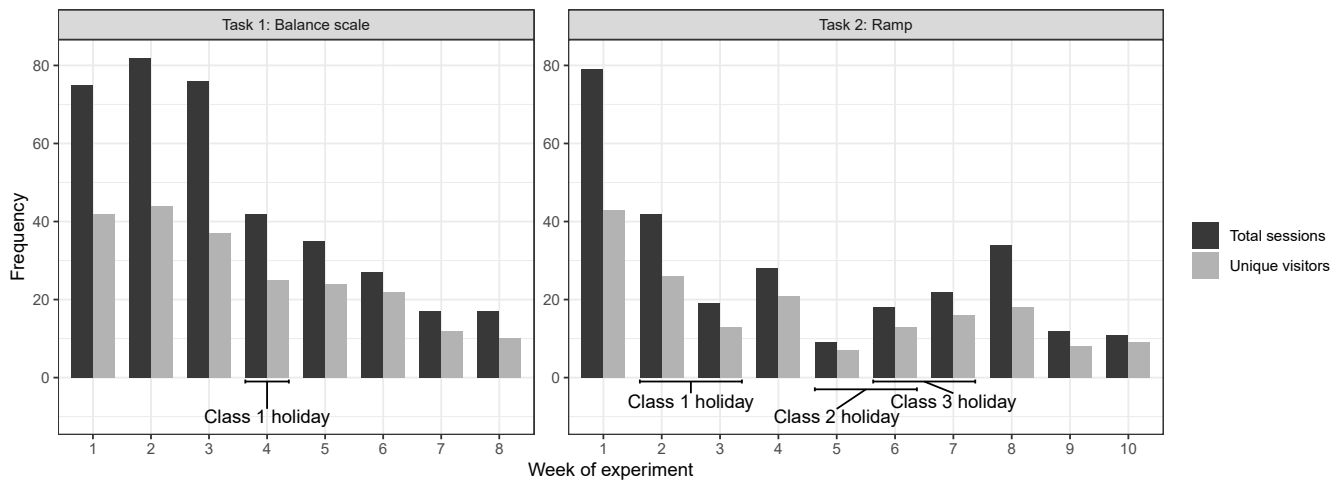


Figure 2: A breakdown of the total number of sessions and unique visitors that interacted with the system per week for each of the tasks.

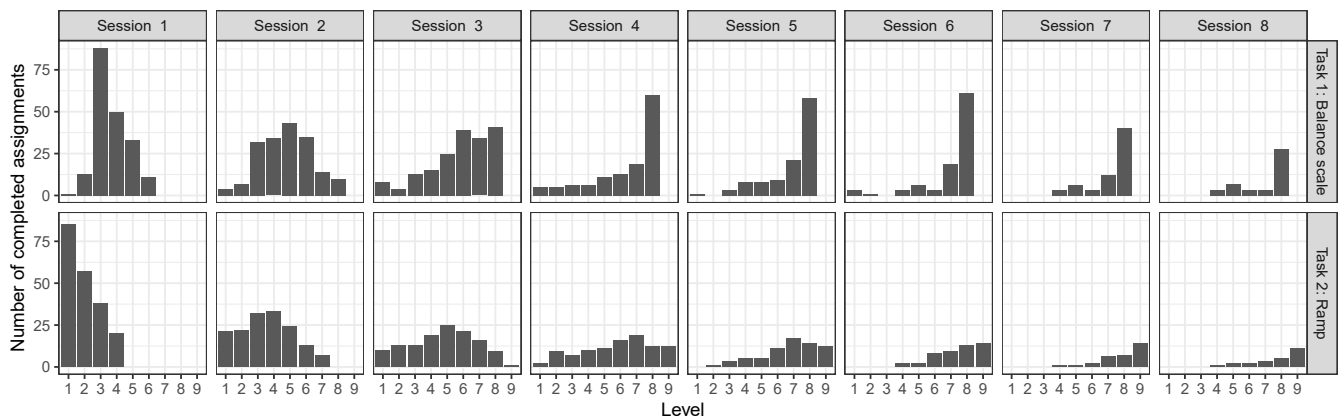


Figure 3: A breakdown of the number of assignments that were completed at a certain level during the first 8 sessions for each of the tasks. Note that in the first task children started at level 3 of 8. In the second task children started at level 1 of 9.

robot remarked about its appearance (e.g. “it’s a boy” or “he has funny hair”) and its behaviours (e.g. “he was blinking” or “he looked at me”). When talking with teachers afterwards, they stated that they enjoyed having the system in their class and that children had enjoyed working on the learning tasks. Although the teachers had a rough understanding of the learning tasks, they did not structurally incorporate them in their regular lessons and had not helped children with the tasks. Additionally, it appeared that none of the teachers had actually interacted with the system themselves.

5 DISCUSSION

5.1 From First Encounters to Sustained Use

During the first weeks we found indications of a novelty effect, characterised by a high frequency of use by many different children. This was corroborated during interviews; children mentioned that there was often a queue of classmates who wanted to work with the system. Such novelty effects have long been reported in similar studies, tapering off after a few weeks (e.g., Fernaeus et al. [11], Kanda et al. [16], Sung et al. [35]). De Graaf et al. [10] showed that during the novelty phase users rated their robot less favourably than after they gained some experience with it. They argue that this could be due to the complexity and unfamiliarity of such a novel interaction, which negatively influences the perceived joy and likeability of the interaction. Serholt [32] found similar indications that an overload of information during initial sessions may result in breakdowns in the interaction. To minimise complexity, our learning tasks were carefully designed to be familiar and accessible for children in our target group. When interviewing the children and analysing the results we found no indications that they encountered particular issues in understanding the system during initial or subsequent interactions.

Furthermore, we found indications for a second novelty effect directly after introducing the second learning task. These results are in line with related work showing how having multiple available activities (e.g. [7]) may help to keep children engaged. In our study, although the core interaction with the system, and especially the dialog with the robot remained the same as before, children likely wanted to explore the novel learning content. Other research has shown the merits of gradually introducing novel robot’s capabilities (e.g. [17]); yet in our study we saw sustained, recurring, voluntary interactions triggered by novel learning content without the introduction of novel social and behavioural content.

To ensure availability of a wide variety of learning content and accommodate children’s development over time we implemented several levels in each task. Our results show that children were able to successfully self-regulate the pace at which they progressed through the various levels. Surprisingly, in the first task many children continued interacting with the system for several more sessions after reaching that point, even though there was only a limited set of unique assignments available at that level. This may be an indication that these children were still engaged with the system even without new content. However, during the second task we found that many children didn’t interact with the system often enough to reach the highest level, and more frequently walked away and abandoned an interaction during a session. Still, we saw no complete abandonment even in the second task.

We found many similarities in usage patterns across our three versions; children progressed at a similar pace, rated the levels similarly, and reported similar performance. This suggests that such factors may not necessarily be greatly influenced (in either direction) by merely having a robot present.

5.2 Content and Variation: Was it Enough?

Deploying a robot in a classroom for an extended period requires sufficient variation in interactional content [24]. To maintain engagement, tasks may offer richness, depth, breadth, and variation. Although implementing such content is not trivial, it appeared manageable and worthwhile. Interest in the system naturally declined over time and resurged after introducing new learning content, but the system continued to attract unique recurring visitors until the end. With this unsupervised deployment study we have gained a first intuition about the extent of the variation that is needed to support a long-term interaction. Our two tasks with each 8-9 levels that adapted to the children’s progress, guided by an interactive system with a limited set of speech phrases and behaviour variations, were sufficient to keep them engaged for multiple interactions during roughly four months on a voluntary basis. We feel this is an order of magnitude that is a good starting point for studying longitudinal educational interactions that extend beyond initial novelty effects.

5.3 Practical Considerations

Our system was set in a familiar environment for the children: their own classroom during regular school hours. To make this possible we closely collaborated with the school management and the teachers involved. Although teachers were interested in having the learning task in their class, they expressed not having time to operate it, oversee and coordinate its use, nor help children interact with it. During the course of the study, however, teachers would occasionally get involved with the system. For example, they would make passing references to it during their lessons or they would regulate when children were allowed to use it (e.g. covering it up with a blanket when children were supposed to work on something else). Another important point raised by teachers was that any child interacting with the system should not interfere with regular lessons. In our system this was addressed by using headphones to play audio. However, especially in the first weeks we found that there were often several bystanders present. Additionally, we needed to be flexible when going to the school to accommodate the classes’ schedules. In some cases this meant that maintenance was delayed and that interviews were performed at moments when children were not busy with regular school tasks.

Differences between individual teacher’s lesson plans will likely have played a role in the frequency of use. When talking with teachers after the study, all three indicated that they had not included the system in regular lesson schedules. However, teachers had on occasion reminded children that they could play with it at certain times or had prevented them from using it at other times. One teacher had encouraged children to play with it just before the experiment ended, reminding them that it may be their last chance.

Communication with parents was very important. We organised an information evening to gather their input and discuss any concerns. Feedback was used to improve our information leaflet and consent form. Although most parents were enthusiastic, some

raised concerns regarding the educational value. This could be resolved in discussions with school management, who agreed that our learning tasks fit as parts within the regular curriculum.

Generally speaking, stakeholders often have similar requirements or constraints, or (unrealistic) expectations regarding the use and deployment of technology. Additionally, users have habits and schedules that must be taken into consideration when designing the interaction. Such issues can be uncovered and addressed at various moments, for example, by organising information events, focus group sessions, or co-design workshops for users and stakeholders.

5.4 Maintenance

Once a system is deployed for an extended amount of time, regular maintenance will likely be necessary [11, 36]. This may involve charging, cleaning and tidying, correcting (technical) failures, or replacing broken or stolen hardware. During our study, most time-consuming maintenance was related to hardware failures (e.g., replacing motors in the robot and sensors in our learning task). In some cases the children and teacher invented a story to explain away such events. For instance, at one point the robot's eyelids had become stuck. Later, children mentioned that the robot was squinting because it had gotten an eye infection. At other times, the robot's motor sounds would interrupt the teacher. Some children mentioned that it was looking for attention or said it was being rude and rebellious. Other common failures were related to disconnected headphones, USB cables, and power cables, caused by cleaning personnel who would occasionally unplug and move components of the system and forget to plug them back in. Unfortunately, we also encountered instances of missing equipment and theft: relatively harmless, like children taking marbles from the ramp task, or less so. Once, someone entered the school and stole one of the tablets, which made quite an impact on the children and was often mentioned during interviews (afterwards, the teacher took care of locking away valuable equipment at night).

In our study the researchers doing the maintenance were also involved in recruitment and interviews. Their involvement had therefore to remain concealed so as not to bias children's responses. To address this we scheduled setup and maintenance at moments after school time with no children present. As a consequence, however, unrecoverable technical errors remained unresolved till the next maintenance window. In general, we recommend that other researchers should think about such auxiliary planning issues and consider the trade offs that are acceptable in their specific study design.

6 CONCLUSION

This paper addressed the need for more long term studies in HRI. We identified several challenges, including the robustness of the technical system and interaction design, problems of logistics and organisation, novelty effects, the need for having a sufficient yet feasible amount of learning material to accommodate children's development over time, and the need for sufficiently varied, but not unnecessarily distractive dialog content in the child-robot interactions.

By developing a technical setup and deploying it long term, unsupervised, in the wild, we investigated the feasibility of such studies as well as explore some parameters pertaining to these. In our study,

children interacted with the robot and embodied educational materials in the familiar context of their classroom while the system guided them through two consecutive inquiry learning tasks; we analysed their usage patterns over time and compared between variants of the system (with a robot, and without).

Our platform of a robot with limited interaction variation, a tablet interface, and a modest selection of assignments with several sensorised learning materials, seems to be a good starting point for further research. We have shown that this setup is sufficient to sustain recurring, unsupervised and voluntary interactions over extended periods of time (four months). We also showed some limits to this, not completely getting rid of novelty effects, but clearly carrying our study far beyond the typical duration expected for initial novelty effects. There is a clear tapering off of the children's interest in staying involved, even if most kids stuck around till the end. This provides a starting point for future long term HRI studies.

Furthermore, it is feasible to automatically gather objective as well as self-report subjective data in an unsupervised way, at sufficient scale to follow individual children over time and run long term comparative experimental studies. Together with an occasional interview this fine-grained task-by-task data can offer rich insights into the development of children's interactions with an HRI system.

We showed that, supported by the RECAL system, children could effectively self-navigate the available difficulty levels, consistently progressing from easier to harder assignments. This offers a starting point for more nuanced and personalised guidance in such systems.

Finally, if we want to show the real potential of robots in daily life, we need to be able to do long term comparative studies which we compare the robot to a non-robot condition, in relatively comparable settings. This is not trivial, but we showed that it is feasible to pursue this: with our still fairly limited robot, the robot and non-robot variants performed somewhat comparably and both managed to "carry children along till the end of the study". Knowing that, we can now start working towards other long term studies where we attempt to tease out the actual benefits of (aspects of) robots. Many variables are potentially important to the effect of robots in class. Now we have shown the feasibility of carrying out long term studies, we should explore the details: how does the presentation of the robot, its physical embodiment [26, 29], its social gaze [43], its appearance, its facial expressions [4], its personality, its life-like behaviour, its background story [28], and so forth, affect children's involvement with the task and robot in longer term deployments? This need not necessarily be the same as in short term and/or lab studies – so we need to start carrying out more long term in the wild studies to start exploring this next level of interesting questions, as has been so well argued by Jung and Hinds [15].

ACKNOWLEDGMENTS

This research has been funded by the European Union 7th Framework Program (FP7-ICT-2013-10) under the grant agreement No 611971 (EASEL). We thank the participating schools and teachers for the pleasant collaboration, the children for their enthusiasm and motivation during the activities, and Andrea Papenmeier and Emiel Harmsen for developing the sensorised learning materials.

REFERENCES

- [1] P. Baxter, T. Belpaeme, L. Canamero, P. Cosi, Y. Demiris, and V. Enescu. 2011. Long-term human-robot interaction with young users. *IEEE/ACM Human-Robot Interaction 2011 Conference (Robots with Children Workshop)*, 1-4 (2011).
- [2] Tony Belpaeme, James Kennedy, Aditi Ramachandran, Brian Scassellati, and Fumihide Tanaka. 2018. Social robots for education: A review. *Science Robotics* 3, 21 (aug 2018). <https://doi.org/10.1126/scirobotics.aat5954>
- [3] BML. accessed 9 Dec. 2019. Behaviour Markup Language Specification v1.0. <http://www.mindmakers.org/projects/bml-1-0/wiki>
- [4] Cynthia Breazeal. 2003. Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies* 59, 1 (2003), 119–155.
- [5] Shruti Chandra, Raul Paradedda, Hang Yin, Pierre Dillenbourg, Rui Prada, and Ana Paiva. 2018. Do Children Perceive Whether a Robotic Peer is Learning or Not?. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18*. ACM Press, New York, New York, USA, 41–49. <https://doi.org/10.1145/3171221.3171274>
- [6] Vasiliki Charisi, Daniel Patrick Davison, Dennis Reidsma, and Vanessa Evers. 2016. Evaluation Methods for User-Centered Child-Robot Interaction. In *Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2016 (Proceedings IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN)*. IEEE ROBOTICS AND AUTOMATION SOCIETY, MoA3.4. <https://doi.org/10.1109/ROMAN.2016.7745171>
- [7] Alexandre Coninx, Paul Baxter, Elettra Oleari, Sara Bellini, Bert Bierman, Olivier Blanson Henkemans, Lola Canamero, Piero Cosi, Valentin Enescu, Raquel Ros Espinoza, Antoine Hiolle, Rémi Humbert, Bernd Kiefer, Ivana Kruijff-Korbayova, Rosmarijn Looije, Marco Mosconi, Mark Neerinx, Giulio Paci, Georgios Patsis, Clara Pozzi, Francesca Sacchitelli, Hichem Sahli, Alberto Sanna, Giacomo Somnavilla, Fabio Tesser, Yannick Demiris, and Tony Belpaeme. 2015. Towards Long-Term Social Child-Robot Interaction: Using Multi-Activity Switching to Engage Young Users. *Journal of Human-Robot Interaction* 5, 1 (aug 2015), 32. <https://doi.org/10.5898/JHRI.5.1.Coninx>
- [8] 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-MAN2016 Workshop on Long-term Child-robot Interaction*. IEEE ROBOTICS AND AUTOMATION SOCIETY.
- [9] Daniel Patrick Davison, Frances Martine 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. *International Journal of Social Robotics* 11, 3 (2019). <https://doi.org/10.1007/s12369-019-00555-6>
- [10] Maartje M.A. de Graaf, Somaya Ben Alouch, and Jan A.G.M. van Dijk. 2016. Long-term evaluation of a social robot in real homes. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems* 17, 3 (dec 2016), 461–490. <https://doi.org/10.1075/is.17.3.08deg>
- [11] Ylva Fernaeus, Maria Håkansson, Mattias Jacobsson, and Sara Ljungblad. 2010. How do you play with a robotic toy animal? A long-term study of Pleo. *Proceedings of the 9th International Conference on Interaction Design and Children - IDC '10* (2010), 39–48. <https://doi.org/10.1145/1810543.1810549>
- [12] Goren Gordon, Samuel Spaulding, Jacqueline Kory Westlund, Jin Joo Lee, Luke Plummer, Marayna Martinez, Madhurima Das, and Cynthia Breazeal. 2016. Affective Personalization of a Social Robot Tutor for Children's Second Language Skills. *Thirtieth AAAI Conference on Artificial Intelligence* (mar 2016), 3951–3967.
- [13] Carmel Houston-Price and Satsuki Nakai. 2004. Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development* 13, 4 (dec 2004), 341–348. <https://doi.org/10.1002/icd.364>
- [14] Alexis Jacq, Séverin Lemaignan, Fernando Garcia, Pierre Dillenbourg, and Ana Paiva. 2016. Building successful long child-robot interactions in a learning context. In *ACM/IEEE International Conference on Human-Robot Interaction*, Vol. 2016-April. IEEE, 239–246. <https://doi.org/10.1109/HRI.2016.7451758>
- [15] Malte Jung and Pamela Hinds. 2018. Robots in the Wild: A Time for More Robust Theories of Human-Robot Interaction. *ACM Trans. Hum.-Robot Interact.* 7, 1, Article 2 (May 2018), 5 pages. <https://doi.org/10.1145/3208975>
- [16] Takayuki Kanda, Takayuki Hirano, Daniel Eaton, and Hiroshi Ishiguro. 2004. Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial. *Human-Computer Interaction* 19, 1 (jun 2004), 61–84.
- [17] T. Kanda, R. Sato, N. Saiwaki, and H. Ishiguro. 2007. A Two-Month Field Trial in an Elementary School for Long-Term Human-Robot Interaction. *IEEE Transactions on Robotics* 23, 5 (oct 2007), 962–971. <https://doi.org/10.1109/TRO.2007.904904>
- [18] James Kennedy, Paul Baxter, and Tony Belpaeme. 2015. The Robot Who Tried Too Hard. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '15*. ACM Press, New York, New York, USA, 67–74. <https://doi.org/10.1145/2696454.2696457>
- [19] David Klahr. 2000. *Exploring Science: The Cognition and Development of Discovery Processes*. The MIT Press, Cambridge.
- [20] David Klahr and Kevin Dunbar. 1988. Dual Space Search During Scientific Reasoning. *Cognitive Science* 12, 1 (jan 1988), 1–48.
- [21] Stefan Kopp, Brigitte Krenn, Stacy Marsella, Andrew N. Marshall, Catherine Pelachaud, Hannes Pirker, Kristinn R. Thórisson, and Hannes Vilhjálmsson. 2006. Towards a Common Framework for Multimodal Generation: The Behavior Markup Language. In *International Conference on Intelligent Virtual Agents, IVA 2006*, Vol. 4133. Springer, Berlin, Heidelberg, 205–217. <https://doi.org/10.1007/11821830>
- [22] Hideki Kozima, Marek P. Michalowski, and Cocoro Nakagawa. 2009. Keepon: A playful robot for research, therapy, and entertainment. *International Journal of Social Robotics* 1, 1 (jan 2009), 3–18. <https://doi.org/10.1007/s12369-008-0009-8>
- [23] Iolanda Leite, Ginevra Castellano, André Pereira, Carlos Martinho, and Ana Paiva. 2014. Empathic Robots for Long-term Interaction: Evaluating Social Presence, Engagement and Perceived Support in Children. *International Journal of Social Robotics* 6, 3 (aug 2014), 329–341. <https://doi.org/10.1007/s12369-014-0227-1>
- [24] Iolanda Leite, Carlos Martinho, and Ana Paiva. 2013. Social Robots for Long-Term Interaction: A Survey. *International Journal of Social Robotics* 5, 2 (jan 2013), 291–308. <https://doi.org/10.1007/s12369-013-0178-y>
- [25] Iolanda Leite, Carlos Martinho, Andre Pereira, and Ana Paiva. 2009. As Time goes by: Long-term evaluation of social presence in robotic companions. In *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 669–674. <https://doi.org/10.1109/ROMAN.2009.5326256>
- [26] 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.
- [27] Ter Mark Maat and Dirk Heylen. 2011. Flipper: An Information State Component for Spoken Dialogue Systems. In *Intelligent Virtual Agents*. Springer Verlag, Reykjavik, 470–472.
- [28] 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 - IEEE International Workshop on Robot and Human Interactive Communication*, Vol. 2005. IEEE, 158–163. <https://doi.org/10.1109/ROMAN.2005.1513773>
- [29] 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 - HRI '18*. ACM Press, New York, New York, USA, 59–68. <https://doi.org/10.1145/3171221.3171250>
- [30] Dennis Reidsma, Vicky Charisi, Daniel Patrick Davison, Frances Martine Wijnen, Jan van der Meij, Vanessa Evers, David Cameron, Samuel Fernando, Roger Moore, Tony Prescott, Daniele Mazzei, Michael Pieroni, Lorenzo Cominelli, Roberto Garofalo, Danilo de Rossi, Vasiliki Vouloutsi, Riccardo Zucca, Klaudia Grechuta, Maria Blancas, and Paul Verschure. 2016. The EASEL project: Towards educational human-robot symbiotic interaction. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 9793. Springer, Cham, 297–306. https://doi.org/10.1007/978-3-319-29793-1_17
- [31] Dennis Reidsma and Herwin van Welbergen. 2013. AsapRealizer in practice - A modular and extensible architecture for a BML Realizer. *Entertainment Computing* 4, 3 (aug 2013), 157–169. <https://doi.org/10.1016/j.entcom.2013.05.001>
- [32] Sofia Serholt. 2018. Breakdowns in children's interactions with a robotic tutor: A longitudinal study. *Computers in Human Behavior* 81 (apr 2018), 250–264. <https://doi.org/10.1016/j.chb.2017.12.030>
- [33] Sofia Serholt and Wolmet Barendregt. 2016. Robots Tutoring Children: Longitudinal Evaluation of Social Engagement in Child-Robot Interaction. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction - NordiCHI '16*. ACM Press, New York, New York, USA, 1–10. <https://doi.org/10.1145/2971485.2971536>
- [34] D. Silvera-Tawil and C. R. Yates. 2018. Socially-Assistive Robots to Enhance Learning for Secondary Students with Intellectual Disabilities and Autism. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 838–843. <https://doi.org/10.1109/ROMAN.2018.8525743>
- [35] JaYoung Sung, Henrik I. Christensen, and Rebecca E. Grinter. 2009. Robots in the wild: understanding long-term use. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction - HRI '09*. ACM Press, New York, New York, USA, 45–52. <https://doi.org/10.1145/1514095.1514106>
- [36] Ja Young Sung, Rebecca E. Grinter, and Henrik I. Christensen. 2010. Domestic robot ecology: An initial framework to unpack long-term acceptance of robots at home. *International Journal of Social Robotics* 2, 4 (dec 2010), 417–429. <https://doi.org/10.1007/s12369-010-0065-8>
- [37] Fumihide Tanaka, Aaron Cicourel, and Javier R Movellan. 2007. Socialization between toddlers and robots at an early childhood education center. *Proceedings of the National Academy of Sciences of the United States of America* 104, 46 (nov 2007), 17954–8. <https://doi.org/10.1073/pnas.0707769104>
- [38] F. Tanaka and J.R. Movellan. 2006. Behavior Analysis of Children's Touch on a Small Humanoid Robot: Long-term Observation at a Daily Classroom over Three Months. In *ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 753–756. <https://doi.org/10.1109/ROMAN.2006.314491>

- [39] Fumihide Tanaka, Javier R. Movellan, Bret Fortenberry, and Kazuki Aisaka. 2006. Daily HRI evaluation at a classroom environment. In *Proceeding of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction - HRI '06*. ACM Press, New York, New York, USA, 3–9. <https://doi.org/10.1145/1121241.1121245>
- [40] Jelte van Waterschoot, Merijn Bruijnes, Jan Flokstra, Dennis Reidsma, Daniel Davison, Mariët Theune, and Dirk Heylen. 2018. Flipper 2.0: A pragmatic dialogue engine for embodied conversational agents. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. ACM Press, Sydney, Australia, 43–50.
- [41] W3C. accessed 9 Dec. 2019. W3C Wiki Entry Concerning Graceful Degradation and Progressive Enhancement. https://www.w3.org/wiki/Graceful_degradation_versus_progressive_enhancement
- [42] Frances Martine Wijnen, Daniel Patrick Davison, Dennis Reidsma, Jan van der Meij, Vicky Charisi, and Vanessa Evers. In press. Now we're talking: Learning by explaining your reasoning to a social robot. *ACM Transactions on Human Robot Interaction* (In press).
- [43] Cristina Zaga, Roelof AJ de Vries, Jamy 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.
- [44] Abolfazl Zarak, Daniele Mazzei, Manuel Giuliani, and Danilo De Rossi. 2014. Designing and Evaluating a Social Gaze-Control System for a Humanoid Robot. *IEEE Transactions on Human-Machine Systems* 44, 2 (apr 2014), 157–168. <https://doi.org/10.1109/THMS.2014.2303083>