Project No. 688014

L2TOR

Second Language Tutoring using Social Robots

Grant Agreement Type: Collaborative Project

Grant Agreement Number: 688014

D2.1: Specifications of interactions for number domain

Due Date: 31/04/2017
Submission Date: 02/05/2017

Start date of project: **01/01/2016** Duration: **36 months**

Organisation name of lead contractor for this deliverable: **TiU**

Responsible Person: **Paul Vogt** Revision: **1.0**

Project co-funded by the European Commission within the H2020 Framework Programme		
Dissemination Level		
PU	Public	**PU**
PP	Restricted to other programme participants (including the Commission Service)	
RE	Restricted to a group specified by the consortium (including the Commission Service)	
CO	Confidential, only for members of the consortium (including the Commission Service)	

Table of Contents

Executive summary

The aim of this deliverable is to discuss the design of the child-robot interactions in the number domain. We explain the design consideration we made during the project, the setup between robot and child and the technical hardware needed to perform the evaluation study. Lastly, we provide the storyboard that we created for the first two lessons.

Principle contributors

Tilburg	Paul Vogt, Mirjam de Haas, Emiel Krahmer, Bram Willemsen, Jan de Wit
Bielefeld	Kirsten Bergmann, Thorsten Schodde
Plymouth	Fotios Papadopoulos
Koç	Junko Kanero
Softbank Robotics	Jean-Marc Montanier
Utrecht	Ora Oudgenoeg-Paz, Rianne Vlaar

Revision History

Version 1.0 (02-05-2017)
First version

Version 1.1 (30-10-2017)
First version with amendments requested by the reviewers

Introduction

According to the project proposal, deliverable D2.1 is defined as follows:

> **D2.1: Specifications of interactions for number domain.** (TIU, R, M15). Report on the specifications on the storyboard, interactions and common ground as completed in tasks T2.1, T2.3 and T2.4 relevant to the number domain.

In brief, the deliverable provides a formal specification to translate the lessons defined in deliverable D1.1 regarding the number domain to provide guidelines for the technical implementation of these lessons to be carried out in work packages WP3 to WP6. More specifically, D2.1 provides a clear formulation of the (expected) input and output that the learning environment needs to handle. This is documented in the storyboards for each lesson. In addition, this document provides specifications of the hardware used in the evaluation study (and related experiments), as well as the experimental setup.

All specifications were developed based on a set of design features that were obtained through a literature review and the results of some preliminary experiments, as reported in Vogt et al. (2017) and added to this report as appendix I. In constructing the design features, we reviewed the literature on child language to list important interactional features that have a positive effect on the children's (second) language development, and evaluated to what extent these can be achieved given the technical (in)capabilities of the NAO robot and relevant state-of-the-art technologies in AI.

The design features reported in Vogt et al. (2017) are summarized as follows:

1. The robot is framed as a peer, acts as a peer, but scaffolds target language as human adults do. This way the child perceives the robot as a friend (which helps build trust and rapport), but the interactions are designed such that the tutoring is as effective as possible.
2. The robot should be introduced to children in a group classroom setting prior to the one-on-one tutoring sessions. This not only helps build trust and rapport, especially for young and shy children, but also helps developing a common ground between children and the robot (particularly from the children's perspective), and in addition will reduce the novelty effect of the robot during the first sessions.
3. The lessons, also for the number domain (contrary to the original proposal), are presented on a tablet rather than with real physical objects. We experimentally found that children learned equally well using a tablet or with real physical objects (Vlaar et al, 2017).
4. The child responds to the questions and tasks through interaction with the tablet. Since automatic speech recognition for children is not sufficiently reliable (Kennedy et al., 2017), we decided to focus on comprehension rather than production.
5. The robot and child establish joint attention on the scenes displayed on the tablet using eye-gaze, gestures (mainly pointing) and verbal instructions. (Additional iconic gestures may be added to provide social cues that could help identify the referents of objects, depending on the results obtained in the experiment designed by De Wit et al., 2017, that are currently being analyzed.)
6. The robot can provide three forms of feedback: positive feedback ("Well done!"), implicit corrective feedback through reformulation ("*Three* means three. You should move three

elephants"), and explicit corrective feedback ("No, that's wrong! *Three* means three."). We are currently conducting an extra experiment to test the effectiveness of the different types of feedback more and we will elaborate on the results of this experiment and the different types of feedback in Deliverable 1.3.

The original proposal mentioned that we intend to make the lesson series adaptive to keep within the child's Zone of Proximal development (cf. Vygotsky, 1978). However, we decided in the consortium as part of the evaluation work package (WP7) that we keep the adaptivity of the robot to a minimum, so that results between children can easily be compared in an experimental setting.

In this deliverable, the design features are translated to the experimental setup (below) and the storyboards (Appendix II). The storyboards were constructed by adapting the lessons from deliverable D1.1 into interaction specifications based on the observations from deliverable D1.2. In particular, we adopted interaction features, such as feedback or gesture use, from those that we observed in the interactions that human tutors displayed as reported in D1.2.

The technical considerations

Hardware

In order to build an autonomous learning environment that can perceive and understand the surrounding environment, the following hardware is used:

1. **NAO robot v5:** SoftBank Robotics' NAO is used as our robot platform.
2. **Microsoft Kinect v2:** The Kinect is used to extract head gaze direction and facial features, to infer the child's behavior and engagement. With that information, the robot can adapt its strategy and guide or motivate the child when distracted. In addition, the Kinect will provide the possibility to detect other non-verbal behaviors such as gestures that may be used in future setups.
3. **Camera tripod:** Used to mount the Kinect.
4. **Microsoft Surface 4 Tablet:** The tablet is used to display an interactive context of the lessons. It displays scenes that the child can manipulate in response to the robots' instructions, thus allowing the system to detect children's interactions. The reason for using a tablet to detect children's responses is that speech recognition for children is at present not working reliably (Kennedy et al., 2017), although voice activity detection may be used to recognize whether the child produces a sound. In addition, the tablet acts as a teacher providing limited feedback and additional recorded verbal input. Many children are used to working with tablets so interacting with this device would not give any problems (Geist, 2014). Besides running the tablet application of the lesson, the tablet can also be used to process input of the child and the Kinect data, detect if there is voice activity, and log children's activity to keep track of their performance for post-hoc analyses and provide data for the adaptive system.
5. **WiFi router:** Used to communicate between robot, tablet and other devices if necessary.
6. **Plastic box:** (Or something similar) Used as a low table to put the tablet on, and to hide cables and other accessories.
7. **Cushion or a hoop:** Used to indicate the child's position.

Experimental setup

The child and robot will sit next to each other in a 90-degree angle. In front of the child, there is a plastic box with on top the Surface tablet that is faced towards the child. All the power supplies, cables, router and extensions are hidden inside the plastic box. The child and robot will sit on the

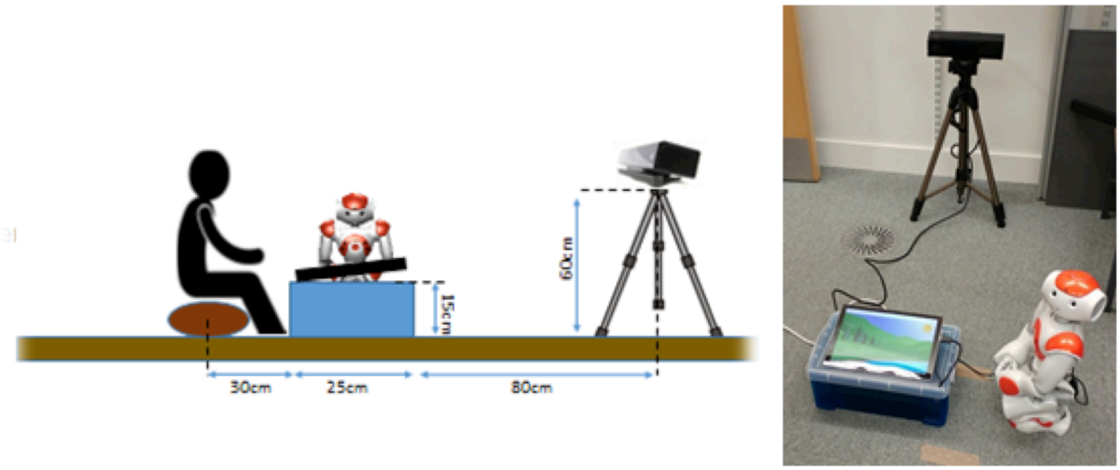

Figure 1. The setup of our child-robot interaction.

ground and a cushion will be provided to the children to sit on. The robot will be crouched during the interaction because this will create room to move its upper body and will still be smaller than the child. The Kinect will be positioned on a tripod in front of and facing the child. The interaction will take place on the floor, as this is safer and more comfortable for the child (see Figures 1 and 2).

Common Ground

Common ground between the robot and the child is achieved at different levels. Common ground can improve the social bond between robot and child in long-term but also in short-term and has a positive effect on language learning (Kanda, 2014). We make a distinction between the initial common ground (prior to the interaction, for example the expectations of the child towards the robot), short-term common ground (during the interaction, for example regaining a child's attention) and long-term common ground (the knowledge that is established over time and sessions, for example how the child responds towards certain feedback) (cf. Clark, 1996). Below, we briefly specify how these levels are attained.

Initial common ground

As mentioned, initial common ground is achieved during a group introduction of the robot in a classroom setting with the presence of the regular teacher. In the first experiment carried out at TIU (De Haas et al., 2017), we gave the introduction prior to the start of the data collection for all children in the school and repeated this at the start of the day for those children that would interact with the robot that day. In a later study (De Wit et al., 2017), we only provided the introduction the week before and felt this was sufficient. We will use this introduction, outlined below, prior to the evaluation study.

A few weeks prior to the introduction, we will send coloring pages of the robot to the preschools (e.g., during recruitment) and ask the preschool and kindergarten teachers to talk a little bit about the robots with the children. About one week before the experimental trials, the experimenters will introduce the robot in the classroom during 'circle time', as this provides a safe and familiar environment with the whole group in which the pedagogical assistants introduce new topics or new activities. One experimenter first introduces the robot by telling a story about Robin, the name we gave to our robot, using a makeshift picture book. In this story the experimenter explains the similarities and dissimilarities between the robot and children to construct the type of common ground that Kanda et al. (2004) considered to have a positive effect on the learning outcome. For example, we tell that Robin enjoys dancing and wants to meet new friends, and even though he does not have a mouth and because of that cannot smile, he expresses enjoyment using his eyes.

After this story, another experimenter enters the room with the robot while it actively looks at faces (using NAO's autonomous life) to provide an animate appearance. The robot introduces itself by saying his name and saying that he is looking for new friends, thus framing itself as a peer. All children are then invited to shake the robot's hand as a greeting. Here we rely on the peer leaders in the class to take the lead and thus stimulate the other children to feel comfortable to shake the robot's hand. We do not force the children to participate; if they do not want to, they can also watch from a distance. The robot tells the children that it likes to dance and initiates a familiar dance routine that the children are encouraged to partake in. At the end of the dance, the robot says it is very tired and lies down to sleep to mark the end of the interaction. The end of the introduction consists of getting a blanket for the robot and laying it in its 'bed'. Finally, the experimenter asks the children if they have any questions and if they would like to see the robot again in the future.

Short-term common ground

Short-term common ground refers to the situation and the relevant things therein that the robot and child share in the 'here and now' during the interaction. This, of course, includes the experimental setting, but also the scene that is displayed on the tablet, the things that are being said and done, as well as the task at hand. To ensure that this common ground is constructed, the robot's task is to draw the child's attention to the relevant aspects of the task, such as the tablet, an activity the child needs to perform on the tablet, or the robot itself. This is achieved, by using gaze, pointing gestures or through speech and other attention grabbers (see Vogt et al., 2017, and the specifications in the storyboard, Appendix II for more details).

The system also needs to monitor what the child is doing, which is carried out using the tablet computer and the Kinect sensor. The tablet will keep track of whether and what the child is pressing (e.g. a picture) or dragging (e.g., an elephant to its cage). The camera of the robot will help monitor the affective state of the child and a Kinect can determine the child's gaze direction. This information is used to sense when the child appears bored and disinterested in the task, or when it is fully engaged with the robot. This system is designed in the input manager (WP4) and the interaction manager (WP5).

Long-term common ground

Long-term common ground relates to what the child and robot acquire as common knowledge over time. For the L2TOR robot this means that the robot needs to keep track of its interactions with the child, and thus should store the affective state of the child during the interactions as recorded with the robot's camera, the child's answers with the tablet and the child's gaze direction with the Kinect. The objective for constructing long-term common ground is that this would allow the robot to adapt to the child's development during the lesson series. For instance, to monitor the timeline of the child's development, to keep track of the child's engagement or its gesture use, and to -in the future- train a model to detect the child's voice activity, recognize speech or gesture use (see WP5). During our evaluation study (see deliverable 7.1), we will only focus on adapting the feedback and attention grabbers for each child over time; however, we will not personalize the task for each child.

Input measurements

While the plan for L2TOR was to use automatic speech recognition (ASR) and object recognition during the interactions, feasibility studies carried out in WP4 indicated that the performance of these technologies is at present insufficient for use in the L2TOR project. The consequence is that the L2TOR system cannot monitor children's speech productions or object manipulations. For this reason it was decided to focus the tutoring on comprehension and mediate the interactions using a tablet in all domains, thus including the number domain.

Robot

While ASR is not used, the robot takes input from the microphone for automatic voice activity detection, which is used to speech up the robots' responses (if no voice activity is detected when children are requested to respond verbally, the robot will proceed to the next stage after a few seconds).

The robot's camera is used to detect the location of the child's face to be able to control the robot's gaze to the child's face. For this behavior, NAO's standard procedure to track faces will be used.

Kinect

Microsoft Kinect is used to track children's gaze orientation, which is used as an indicator of children's engagement. If children turn out to be unengaged, the robot will use this information to attract the child's attention.

Tablet

The tablet is not only used to display context to the child, but also detects responses from the child after the robot asked the child to perform some action on the tablet, e.g., to click on a certain object, or to relocate an object from part of the screen to another part of the screen.

Storyboard functions

The storyboard provided in Appendix II specifies how the lesson series defined in D1.1 will be implemented in the robot and tablet. The format of the storyboard is an Excel table, with columns specifying the input to the system, the robot's output in L1, L2, gaze and gestures, the tablet's output, variables used by the system, comments, type of instruction and function that specifies some generic behavior. While most of the storyboard should speak for itself, the functions are specified below.

General behavior specifications

The robot will display some general behaviors for every child. First, the robot will start the lesson with greeting the child using the child's name. Second, during the interaction, the robot will use breathing animations to create the impression that it is alive. These animations are based on the autonomous life function provided by SoftBank Robotics, but adapted for our project to be performed by a crouched robot instead of a standing robot. Third, if the child is distracted or unengaged the robot will use an attention grabber strategy to regain the child's attention. As mentioned in D1.2 we will call the child's name as the main attention grabber within a session, and occasionally accompany this with a non-verbal gesture (e.g. waving arms). A probability determines whether a gesture has to be added to the child's name and whether the robot says the child's name, and uses for example its arms to redirect the child's attention (see Function 1).

```
AttentionGrabber(probability, name)
      IF child is unengaged
            IF probability > threshold
                  AddGesture
            Say Child's name
```

Function 1. The attention grabber

Input to

Gesture use

Gestures can contribute to language learning (see e.g. D1.2 for the use of gestures during class room activity and deWit et al. (2017) for our experiment with iconic gestures and the robot). However, the use of gesture may distract the child (Kennedy et al., 2017) and using gestures increases the duration of the experiment due to the time it takes for the robot to execute a gesture. The model determines whether a gesture has to be added to the robot's utterance based on a probability function. The robot will, for example, provide instructions accompanied by a gesture, if the associated probability of gesture use crosses a certain threshold.

```
useGesture ( probabilityGesture )
    update probabilityGesture
    IF probabilityGesture > threshold
        AddGesture
```

Function 2. This function determines whether the robot will use a gesture.

Feedback

Every time the child has to perform a task, the robot provides a response. These responses differ for each action and use the activity of the child (objectTouched, VoiceActivity) and the target word as input. Depending of the child's action, the robot will provide either positive or negative feedback, might repeat the question or show the task itself. As children respond differently to different types of feedback (as shown in De Haas et al. (2017)), the feedback is dependent on the child's affective state and is adapted for each child. In this deliverable, we only show when the robot will provide negative or positive feedback, however, how these feedback utterances will be adapted for each child will be further explained in deliverable 1.3 and 5.1. The system decides, depending on the child's response, a different feedback type. Function 3 shows the rules for the robot's responses to the children's actions, the child's action can result in negative or positive feedback. In the case that the child does not respond within a certain time frame (the timer), the robot will perform the task for the child or provide hints as another form of feedback to scaffold the language learning. The duration of timer still needs to be determined.

```
giveResponseToSelectObject (objectTouched, VoiceActivity, target)
      Start Timer 5 seconds
      IF objectTouched is target
            giveFeedback(target, correct)
      ELSE IF objectTouched is not target
            giveFeedback(target, incorrect)
      ELSE IF voiceActivity is TRUE AND no objectTouched
            repeatQuestion(target)
      ELSE
            requestAnswer(target)
      After timer:
            giveHelp(target)


giveResponseToMoveObject (objectTouched, VoiceActivity, target)
      Start Timer 5 seconds
      IF objectTouched is target
            rephraseQuestion(target)
      ELSE IF objectTouched is moving
            giveFeedback(target, correct)
      ELSE IF objectTouched is not target
            giveFeedback(target, incorrect)
      ELSE IF voiceActivity is TRUE AND no objectTouched
            repeatQuestion(target)
      ELSE
            requestAnswer(target)
      After timer:
            giveHelp(target)

giveResponseOnSpeech (touch, VoiceActivity, target)
      Start Timer 5 seconds
      IF voiceActivity is True AND no objectTouched
            giveFeedback(target, correct)
      ELSE
            requestAnswer(target)
      After timer:
            giveHelp(target)
```

Function 3. The different response functions. For each type of expected input of the child, different functions are created.

Gaze of the robot

The robot will follow the child's face during the interaction and the robot will gaze towards the tablet if the screen displays a scene that requires an action (for example giving the child an instruction to select an object, or showing how to perform a certain task).

Tablet output

The tablet output is specified in the storyboard, the tablet will display a 3D environment in which the child can manipulate objects. For example, Figure 2 shows one part of the zoo environment. The child can drag the animals to the tree, behind the fence or can feed the elephants with grass. The screen of the tablet will be turned off during the speech of the robot to distract the child less so the child can completely focus on the robot's output.

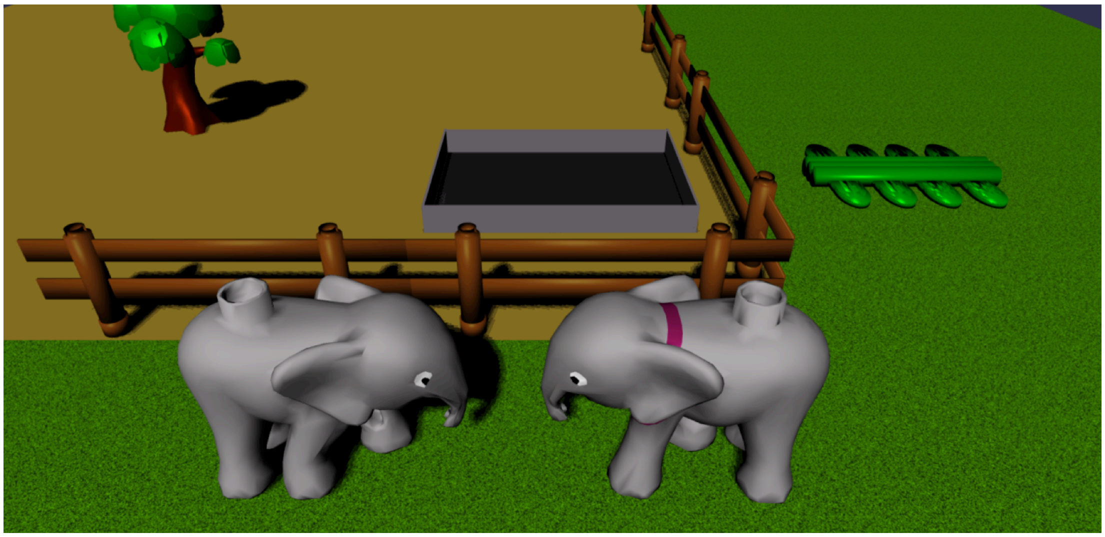

Figure 2. The 3D environment of the tablet in which animals can be moved around.

References

Clark, H. H. (1996). Communities, commonalities, and communication. *Rethinking linguistic relativity*, *17*, 324-355.

De Haas, M., Baxter, P., de Jong, C., Krahmer, E., & Vogt, P. (2017, March). Exploring different types of feedback in preschooler and robot interaction. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, (pp. 127-128). Vienna, NY: ACM

De Wit., J., Schodde, T., Willemsen, B., Bergmann, K., de Haas, M., Kopp, S., Krahmer, E. & Vogt, P. (2017). Exploring the Effect of Gestures and Adaptive Tutoring on Childrens Comprehension of L2 Vocabularies. , in *Proceedings of the Workshop R4L at ACM/IEEE HRI 2017*

Geist, E. (2014). Using tablet computers with toddlers and young preschoolers. *YC Young Children*, *69*(1), 58.

Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. *Human-Computer Interaction*, *19*(1), 61-84.

Kennedy, J., Baxter, P., & Belpaeme, T. (2017). Nonverbal Immediacy as a Characterisation of Social Behaviour for Human–Robot Interaction. *International Journal of Social Robotics*, 1-20.

Kennedy, J., Lemaignan, S., Montassier, C., Lavalade, P., Irfan, B., Papadopoulos, F., ... & Belpaeme, T. (2017, March). Child speech recognition in human-robot interaction: evaluations and recommendations. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, (pp.* 82-90). Vienna, NY: ACM.

Vlaar, R., Verhagen. J., Oudgenoeg-Paz, O. &Leseman. P (2017). Comparing L2 Word Learning through a Tablet or Real Objects: What Benefits Learning Most? , in *Proceedings of the Workshop R4L at ACM/IEEE HRI 2017*

Vogt, P., De Haas, M., De Jong, C., Baxter, P., & Krahmer, E. (2017). Child-Robot Interactions for Second Language Tutoring to Preschool Children. *Frontiers in Human Neuroscience*, *11*.

Vygotsky, L. (1978). Interaction between learning and development. *Readings on the development of children*, *23*(3), 34-41.

Appendix I. Paper Vogt et al. 2017

frontiers in Human Neuroscience

PERSPECTIVE
published: 02 March 2017
doi: 10.3389/fnhum.2017.00073

Child-Robot Interactions for Second Language Tutoring to Preschool Children

Paul Vogt *, Mirjam de Haas, Chiara de Jong, Peta Baxter and Emiel Krahmer

Tilburg Center for Cognition and Communication, Tilburg University, Tilburg, Netherlands

In this digital age social robots will increasingly be used for educational purposes, such as second language tutoring. In this perspective article, we propose a number of design features to develop a child-friendly social robot that can effectively support children in second language learning, and we discuss some technical challenges for developing these. The features we propose include choices to develop the robot such that it can act as a peer to motivate the child during second language learning and build trust at the same time, while still being more knowledgeable than the child and scaffolding that knowledge in adult-like manner. We also believe that the first impressions children have about robots are crucial for them to build trust and common ground, which would support child-robot interactions in the long term. We therefore propose a strategy to introduce the robot in a safe way to toddlers. Other features relate to the ability to adapt to individual children's language proficiency, respond contingently, both temporally and semantically, establish joint attention, use meaningful gestures, provide effective feedback and monitor children's learning progress. Technical challenges we observe include automatic speech recognition (ASR) for children, reliable object recognition to facilitate semantic contingency and establishing joint attention, and developing human-like gestures with a robot that does not have the same morphology humans have. We briefly discuss an experiment in which we investigate how children respond to different forms of feedback the robot can give.

Keywords: social robots, second language tutoring, education, child-robot interaction, robot assisted language learning

OPEN ACCESS

Edited by:
Mila Vulchanova, Norwegian University of Science and Technology, Norway

Reviewed by:
Ramesh Kumar Mishra, University of Hyderabad, India
Vera Kempe, Abertay University, UK

****Correspondence:***
Paul Vogt
p.a.vogt@uvt.nl

Received: *26 October 2016*
Accepted: *06 February 2017*
Published: *02 March 2017*

Citation:
Vogt P, de Haas M, de Jong C, Baxter P and Krahmer E (2017) Child-Robot Interactions for Second Language Tutoring to Preschool Children. Front. Hum. Neurosci. 11:73. doi: 10.3389/fnhum.2017.00073

SOCIAL ROBOTS FOR SECOND LANGUAGE TUTORING

Given the globalization of our society, it is becoming increasingly important for people to speak multiple languages. For instance, the ability to speak foreign languages fosters people's mobility and increases their chances for employment. Moreover, immigrants to a country need to learn the official host language. Since young children are most flexible at learning languages, starting second language (L2) learning in preschool would provide them a good opportunity to acquire the second language more fluently at a later age (Hoff, 2013).

One trend in the digital age of the 21st century is that technologies are being developed for educational purposes, including technologies to support L2 tutoring. There exist many forms of digital technologies for PCs, laptops or tablet computers that support second language learning, although there is little evidence about their efficacy (Golonka et al., 2014; Hsin et al., 2014). While children can benefit from playing with such technologies, these systems lack the situated and embodied interactions that young children naturally engage in and learn from (Glenberg, 2010; Leyzberg et al., 2012). Social robots represent an emerging technology that provides situatedness and embodiment, and thus have potential benefits for educational purposes. In essence, social robots are autonomous physical agents, often with human-like feature, that can interact socially with humans in a semi-natural way for prolonged periods of time (Dautenhahn, 2007). The use of social robots, in comparison to more traditional digital technologies, allows for the development of tutoring systems more akin to human tutors, especially with respect to the situated and embodied social interactions between child and robot. Thus, this offers the opportunity to design robots such that they interact in a way that optimizes the child's language learning.

Recently, an increasing interest has emerged to develop social robots to support children with learning a second language (Kanda et al., 2004; Belpaeme et al., 2015; Kennedy et al., 2016). While a social robot cannot provide tutoring to the level humans can, recent studies suggest that using social robots can result in an increased learning gain compared to digital learning environments for tablets or computers (Han et al., 2008; Leyzberg et al., 2012). It is, however, unclear why this is the case. Perhaps the physical presence of the robot draws the attention of children for longer periods of time, but the embodiment and situatedness of the learning environment perhaps also helps the children to ground the language more strongly than interactions with virtual objects do.

While there is a fair body of research on robot tutors, a comprehensive description of the design features for a second language robot tutor based on what is known about children's language acquisition is lacking. What are the design features of child-robot interactions that would support second language learning? And, to what extent can these interactions be implemented in today's social robot technologies? In this perspective article, we try to answer these questions based on theoretical accounts from the literature on children's language acquisition in combination with our own experiences in designing a tutor robot.

DESIGNING CHILD-ROBOT INTERACTIONS

In our project, we aim to design a digital learning environment in which preschool children interact one-on-one with a social robot that supports either their learning of English as a foreign language, or the school language for those children who have a different native language (Belpaeme et al., 2015). In particular, the project aims to develop a series of tutoring sessions revolving around three increasingly complex domains (numbers, spatial relations and mental vocabulary). In each session, the child will engage with the robot (a Softbank Robotics NAO robot) in a game-like scenario focusing on learning a small number of target words. The contextual setting is generally displayed on a tablet computer that occasionally also provides some verbal support, however, the robot acts as the interactive tutor. Below we discuss the design features and considerations that we believe are crucial to design a successful tutoring system.

Peer-Like Tutoring

One of the first questions that comes up when designing a robot tutor is whether the robot should take the role of a teacher or a peer. Research on children's language acquisition has demonstrated that children learn more effectively from an adult who can use well-defined pedagogical methods for teaching children using clear directions, explanations and positive feedback methods (Matthews et al., 2007). However, designing and framing the robot as an adult tutor has the disadvantage that children will form expectations about the robot's behavior and proficiency that cannot be met with current technology (Kennedy et al., 2015). Due to technological limitations of the robot and underlying software, communication breakdowns are more likely to occur than with a human. For a peer robot introduced as a fellow language learner, breakdowns in communication are more acceptable. Moreover, interacting with robots acting as peers is conceived as more fun (Kanda et al., 2004), allows for learning-by-teaching (Tanaka and Matsuzoe, 2012) and has a proven to be efficient in teaching children how to write (Hood et al., 2015). Furthermore, there is some evidence that children's learning can benefit from interacting with peers (Mashburn et al., 2009). Given these considerations, we believe it is desirable to frame or introduce the robot as a peer and friend, yet design its interactions insofar possible based on pedagogically well-established strategies to scaffold language learning.

First Impressions

To implement effective tutoring, the robot needs to interact with children in multiple sessions, so they have to be motivated to engage in long-term interactions with the robot. Establishing common ground between child and robot can contribute to this (Kanda et al., 2004), but first impressions to establish trust and rapport are also crucial (Hancock et al., 2011).

Despite the wealth of studies regarding the introduction of entertainment robots as toys to children (e.g., Lund, 2003), surprisingly little research has been conducted on designing protocols on how to introduce a robot tutor to a group of preschool children. Fridin (2014) presents one exception, and found that introducing a robot tutor to children in group sessions improved subsequent interactions compared to introducing the robot to children in individual sessions. Another study by Westlund et al. (2016) found that the way a robot is framed, either as a machine or a social entity, affected the way children later engaged with the robot. They concluded that introducing the robot as a machine could create a more distant relation between child and robot, thus reducing acceptance. We therefore decided to frame the robot in our project as a social playmate for the children and introduced the robot in a group session. However, the NAO robot is slightly taller and more rigid than the fluffy huggable Tega robot, which Westlund et al. (2016) used, and we observed that some 3-year-old children were somewhat intimidated by the NAO robot on their first encounter. Such a first impression of the robot could reduce the trust that the child had for the robot, which could negatively affect their willingness to interact with the robot in the short-term, but also in the long-term. To develop a successful first encounter and to build trust between the child and robot, we designed the following strategy for introducing the robot to 3-year-old children at their preschool.

Pilot studies revealed that some children got anxious when the robot was introduced and then suddenly started to move. To familiarize children prior to their first encounter with the robot, it is therefore advisable to prepare them well. For our study, we sent coloring pages of the robot to the preschools during recruitment and asked the pedagogical assistants to talk a little bit about the robots to the children. About 1 week before the experimental trials, the experimenters introduced the robot in class during their daily "circle time", as this provided a safe and familiar environment with the whole group in which the pedagogical assistants usually introduce new topics or new activities. One experimenter first introduced the robot by telling a story about Robin, the name of our robot, using a makeshift picture book. In this story we explained the similarities and dissimilarities between the robot and children to construct the type of common ground considered to have a positive effect on the learning outcome (Kanda et al., 2004). For example, we told that Robin enjoys dancing and wants to meet new friends, and even though he does not have a mouth and because of that cannot smile, he can smile using his eye LEDs.

After this story, another experimenter entered the room with the robot while it was actively looking at faces to provide an animate feeling. The robot introduced itself with a small story about itself and by performing a dance in which the children were encouraged to participate. The end of the circle time consisted of getting a blanket for the robot so it could "sleep". This introduction was repeated later on the days we conducted the experiment in one-on-one sessions. While by then most children were comfortable interacting with the robot, some were still timid and anxious. To encourage these children to feel comfortable, one of the experiment leaders would sit next to the child during the warm-up phase of the experiment and motivate the child to respond to the robot when necessary until the child was sufficiently comfortable to interact with the robot by herself/himself. We found that the younger 3-year olds required more support from the experimenters than the older 3-year olds (Baxter et al., 2017). Although we are still analyzing the experiments, preliminary findings suggest that our introduction helped children to build trust and common ground with the robot effectively.

Temporal Contingency

Research has shown that it is crucial for children's language development that their communication bids are responded to in a temporally contingent manner (Bornstein et al., 2008; McGillion et al., 2013). This, however, faces a technological challenge. While adults tend to take over turns very rapidly, robots require relatively long processing time to produce a response. Nevertheless, in our first experiment (de Haas et al., 2016), we observed that children were at first surprised by the delayed responses, but quickly adapted to the robot and waited patiently for a response. Perhaps this is because children also require longer than adults to take turns (Garvey and Berninger, 1981) and having framed the robot as a peer children made the delays more plausible or expected. Nevertheless, while a lag in temporal contingency may not harm the interaction with children, it may harm learning. One way to remedy this may be to have the robot start responding by providing a backchannel signal, such as "uhm" to indicate the robot is (still) taking his turn, but requires more time to process (Clark, 1996).

Semantic Contingency

Robots should not only respond to children in a timely fashion, but also in a semantically contingent fashion (i.e., consistent with the child's focus of attention), as this too has a positive effect on children's language acquisition (Bornstein et al., 2008; McGillion et al., 2013). For instance, research has shown that by responding in a semantically contingent manner, either verbally or by following children's gaze, (joint) attention is sustained for a longer duration (Yu and Smith, 2016), allowing children to learn more about a situation. To achieve semantically contingent responses, the robot should be able to understand the child's communication bids, construct joint attention with the child, or at least identify what the child is attending to. Monitoring children's behavior and establishing joint attention are therefore considered crucial for designing a successful robot tutor.

Monitoring Children's Behavior

To understand children's communication bids, as well as to test their pronunciation of the L2, it is important that the robot be equipped with well-functioning automatic speech recognition (ASR). However, the performance of state-of-the-art ASR for children is still suboptimal, especially for preschool-aged children (Fringi et al., 2015; Kennedy et al., 2017). Reasons for this include that children's pronunciation is often flawed and that their speech has a different pitch than adults. Moreover, relatively little research has been carried out in this domain and not much data exist to train ASR on. While it can be expected that the performance of ASR for children will improve in the not too distant future (Liao et al., 2015), until then alternative strategies need to be developed that do not (exclusively) rely on ASR.

In our project, we explore various strategies to achieve this, both based on monitoring non-verbal behaviors of the children and focusing on comprehending rather than producing L2. The first strategy relies on providing children tasks they have to perform in the learning environment, such as placing "a toy cow behind a tree" when teaching spatial language. This, however, requires the visual object recognition on the robot to work well, which is only the case when the scene contains a limited set of distinctively recognizable objects, such as distinctly colored objects (Nguyen et al., 2015). A potential solution explored in our project is to use objects with build-in RFID sensors that can be tracked automatically. The second solution we explore is to use a touch screen tablet that displays scenes the child can manipulate, which not only has the advantage of avoiding the problem of object recognition, but also allows us to control the robot's responses and vary the scenes in real time. A downside, however, is that it takes away the 3-dimensional physical aspect of embodied cognition that would help the children to better entrench what they learn (Glenberg, 2010). Currently, experiments are underway to investigate the effect of using real vs. virtual objects. These solutions not only aid in understanding the child's communication bids, it also helps in identifying their attention and can thus contribute to establishing joint attention.

Joint Attention and Gestures

Joint attention, where interlocutors attend on the same referent, is a form of social interaction that has been shown to support children's language learning (Tomasello and Farrar, 1986). One way to establish joint attention with a child is to guide their attention to a referent using gestures, such as pointing or iconic gestures. The ability to produce gestures in the real world is potentially one of the main advantages of using physical robots as opposed to virtual agents, who may have a harder time to establish joint attention. However, many robots' physical morphologies do not correspond one-to-one to the human body. Hence, many human gestures cannot be translated directly to robot gestures. For instance, the NAO robot that we use in our research has a hand with three fingers that cannot be controlled independently, so index finger pointing cannot be achieved (see **Figure 1**). Will children still recognize NAO's arm extension as a pointing gesture? And if so, will they be able to identify the object the robot refers to? We are currently running an experiment to investigate how NAO's pointing gestures are perceived, and preliminary findings show that participants have difficulty identifying the referred object on a small tablet screen. Similar issues arise when developing other gestures. One of the other non-verbal behaviors we are using is the coloring of NAO's eye LEDSs to indicate the robot's happiness as a form of positive feedback, since the robot cannot smile with its mouth.

Feedback

Feedback, too, is an interactional feature known to help language learning (Matthews et al., 2007; Ateş -Şen and Küntay, 2015). The question is how should the robot provide feedback, such that it is both pleasant and effective for learning? While adults provide positive feedback explicitly, they usually provide negative feedback implicitly by reformulating children's errors in the correct form. In child-child interactions, however, Long (2006) found that there was a clear advantage in learning from explicit negative feedback (e.g., by saying "no, that's wrong, you need to say 'he ran'") when compared to reformulating feedback (the learner says "he runned" and the teacher reacts with "he ran").

FIGURE 1 | NAO pointing to a block with three fingers. (Note that written, informed consent was obtained from the parents of the child for the publication of this image).

To investigate how children experience feedback from a peer robot, we carried out an experiment among 85 3-year-old Dutch-speaking children at preschools in Netherlands (de Haas et al., 2016, 2017). In this experiment, the children interacted with a NAO robot during which they received a short lesson on how to count from 1 to 4 in English. After a short training phase, in which the children were presented with the four counting words twice in relation to body parts and wooden blocks, they were given instructions by the robot to pick up a given number of blocks. While the instructions were given in their native language, the numbers were uttered in English. In response to the child's ability to achieve the task, the robot provided feedback. The experiment followed a between-subjects design with three conditions: adult-like feedback (explicit positive and implicit negative), peer-like feedback (no positive and explicit negative) and no feedback. We did not find significant differences in learning gain between the conditions, probably because the target words were insufficiently often repeated. However, we explored the way in which the children engaged with the robot after they received feedback and we found that children looked less often at the experimenter in the feedback conditions than in the no feedback condition. Further analyses are carried out to evaluate how the children responded to the various forms of feedback to find out what type of feedback would be most effective for achieving both acceptable and effective tutoring interactions.

Zone of Proximity and Adaptivity

Finally, from a pedagogical point of view it is desirable that the interactions between child and robot be sufficiently challenging and varied so that the child has a target to learn from, but at the same time interactions should not be too difficult, because that may frustrate the child causing it to lose interest in the robot (Charisi et al., 2016). In other words, the robot should remain in Vygotsky's Zone of Proximity that supports an effective learning environment (Vygotsky, 1978). In order to achieve this, the robot should be able to keep track of the children's advancements in language learning and perhaps their emotional states during the tutoring sessions, and adapt to these. While the former can be monitored as discussed previously, it may be possible to detect emotional states known to influence learning (e.g., concentration, confusion, frustration and boredom) using methods from affective computing (D'Mello and Graesser, 2012). Using this type of information, it is possible to adapt the tutoring sessions by either reducing or increasing the number of repetitions, and/or change the subject (Schodde et al., 2017).

CONCLUSION

This perspective article presented some design features that we consider crucial for developing a social robot as an effective second language tutor. We believe the robot is most effective when it is framed as a peer, i.e., as a fellow language learner and playmate, but that is designed to use adult-like interaction strategies to optimize learning efficacy. In order to establish common ground and trust to facilitate long-term interactions, we consider it essential that the robot be introduced with appropriate care on the first encounter. As an example, we outlined our strategy for introducing a robot to preschool children. Interactions between child and robot should be contingent and multimodal, and provide appropriate forms of feedback. We argued that the robot should remain within Vygotsky (1978) Zone of Proximal Development and thus should adapt to the individual level of the child.

We also discussed some technical challenges that need to be solved in order to implement contingent interactions; the most important of which we believe is ASR, which presently does not work well for children's speech. While various technical challenges still remain, we expect that social robots will provide effective digital technologies to support second language development in the years to come.

The present list of design features covers many aspects that need to be considered when developing a tutor robot, but it is not yet comprehensive. One aspect that has not been covered, for instance, concerns the design of robots for children from different cultures, which could require different design choices (Shahid et al., 2014). For example, in some cultures education is more teaching-centered (Hofstede, 1986) and thus designing the tutor as a peer robot may be less effective or acceptable (Tazhigaliyeva et al., 2016). Concluding, this perspective article offers only a first step towards a comprehensive list of design features for tutor robots and additional research is needed to complete and optimize the list.

ETHICS STATEMENT

The Research Ethics Committee of Tilburg School of Humanities approved this study, and the parents of all participating children gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

PV, MH and EK designed the conceptual aspects of the article; PV, MH, CJ and PB carried out the literature review; PV, EK and MH designed the feedback study; MH, CJ and PB designed the introduction study; MH, CJ and PB carried out the studies; PV and MH wrote the article; CJ, PB and EK revised the article critically.

FUNDING

This work has been supported by the EU H2020 L2TOR project (grant 688014). CJ and PB thank the research trainee program of the Tilburg School of Humanities for their support.

ACKNOWLEDGMENTS

The authors wish to thank all members of the L2TOR project for their support and advice regarding this research. We also thank Kinderopvanggroep Tilburg and all participating daycare centers and preschools for their assistance in this research. Finally, a big thank you to all the children and their parents for participating in our research.

REFERENCES

Ateş-Şen, B. A., and Küntay, A. C. (2015). Children's sensitivity to caregiver cues and the role of adult feedback in the development of referential communication. *The Acquisition of Reference*, eds L. Serratrice and S. E. M. Allen (Amsterdam: John Benjamins), 241–262.

Baxter, P., De Jong, C., Aarts, A., de Haas, M., and Vogt, P. (2017). "The effect of age on engagement in preschoolers' child-robot interactions," in *Companion proceedings of the 12th Annual ACM International Conference on Human-Robot Interaction* (HRI'17), (Vienna, Austria).

Belpaeme, T., Kennedy, J., Baxter, P., Vogt, P., Krahmer, E. E. J., Kopp, S., et al. (2015). "L2TOR-second language tutoring using social robots," in *First Workshop on Educational Robots* (WONDER), (Paris, France).

Bornstein, M. H., Tamis-LeMonda, C. S., Hahn, C. S., and Haynes, O. M. (2008). Maternal responsiveness to young children at three ages: longitudinal analysis of a multidimensional, modular and specific parenting construct. *Dev. Psychol.* 44, 867–874. doi: 10.1037/0012-1649.44.3.867

Charisi, V., Davison, D., Reidsma, D., and Evers, V. (2016). "Children and robots: a preliminary review of methodological approaches in learning settings," in *2nd Workshop on Evaluating Child Robot Interaction - HRI*, (Christchurch, New Zealand).

Clark, H. H. (1996). *Using Language.* Cambridge: Cambridge University Press.

Dautenhahn, K. (2007). Socially intelligent robots: dimensions of human-robot interaction. *Philos. Trans. R. Soc. B Biol. Sci.* 1480, 679–704. doi: 10.1098/rstb.2006.2004

de Haas, M., Baxter, P., de Jong, C., Vogt, P., and Krahmer, E. (2017). "Exploring different types of feedback in preschooler and robot interaction," in *Companion proceedings of the 12th Annual ACM International Conference on Human-Robot Interaction* (HRI'17), (Vienna, Austria).

de Haas, M., Vogt, P., and Krahmer, E. J. (2016). "Enhancing child-robot tutoring interactions with appropriate feedback," in *Proceedings of First Workshop on Long-Term Child-Robot Interaction. IEEE Ro-Man*, (New York, NY).

D'Mello, S., and Graesser, A. (2012). Dynamics of affective states during complex learning. *Learn. Instr.* 22, 145–157. doi: 10.1016/j.learninstruc.2011.10.001

Fridin, M. (2014). Kindergarten social assistive robot: first meeting and ethical issues. *Comput. Hum. Behav.* 30, 262–272. doi: 10.1016/j.chb.2013.09.005

Fringi, E., Lehman, J., and Russell, M. J. (2015). "Evidence of phonological processes in automatic recognition of children's speech," in *16th Annual Conference of the International Speech Communication Association*, (Dresden, Germany), 1621–1624.

Garvey, C., and Berninger, G. (1981). Timing and turn taking in children's conversations. *Discourse Process.* 4, 27–57. doi: 10.1080/01638538109544505

Glenberg, A. M. (2010). Embodiment as a unifying perspective for psychology. *Wiley Interdiscip. Rev. Cogn. Sci.* 4, 586–596. doi: 10.1002/wcs.55

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., and Freynik, S. (2014). Technologies for foreign language learning: a review of technology types and their effectiveness. *Comput. Assist. Lang. Learn.* 27, 70–105. doi: 10.1080/09588221.2012.700315

Han, J. H., Jo, M. H., Jones, V., and Jo, J. H. (2008). Comparative study on the educational use of home robots for children. *J. Inf. Process. Syst.* 4, 159–168. doi: 10.3745/jips.2008.4.4.159

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J., and Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Hum. Factors* 53, 517–527. doi: 10.1177/0018720811417254

Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: implications for closing achievement gaps. *Dev. Psychol.* 49, 4–14. doi: 10.1037/a0027238

Hofstede, G. (1986). Cultural differences in teaching and learning. *Int. J. Intercult. Relat.* 10, 301–320. doi: 10.1016/0147-1767(86)90015-5

Hood, D., Lemaignan, S., and Dillenbourg, P. (2015). "When children teach a robot to write: an autonomous teachable humanoid which uses simulated handwriting," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, (New York, NY), 83–90.

Hsin, C.-T., Li, M.-C., and Tsai, C.-C. (2014). The influence of young children's use of technology on their learning: A. *Educ. Technol. Soc.* 17, 85–99. Available online at: https://eric.ed.gov/?id=EJ1045554

Kanda, T., Hirano, T., Eaton, D., and Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: a field trial. *Hum. Comput. Interact.* 19, 61–84. doi: 10.1207/s15327051hci1901&2_4

Kennedy, J., Baxter, P., Senft, E., and Belpaeme, T. (2015). "Higher nonverbal immediacy leads to greater learning gains in child-robot tutoring interactions," in *International Conference on Social Robotics*, eds A. Tapus, E. André, J.-C. Martin, F. Ferland and M. Ammi (New York, NY: Springer International Publishing), 327–336.

Kennedy, J., Baxter, P., Senft, E., and Belpaeme, T. (2016). "Social robot tutoring for child second language learning," in *Proceedings of the 11th Annual ACM/IEEE International Conference on Human-Robot Interaction* (HRI'16), (Christchurch, New Zealand), 231–238.

Kennedy, J., Lemaignan, S., Montassier, C., Lavalade, P., Irfan, B., Papadopoulos, F., et al. (2017). "Child speech recognition in human-robot interaction: evaluations and recommendations," in *Proceedings of the 12th Annual ACM International Conference on Human-Robot Interaction* (HRI'17), (Vienna, Austria).

Leyzberg, D., Spaulding, S., Toneva, M., and Scassellati, B. (2012). "The physical presence of a robot tutor increases cognitive learning gains," in *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, (Sapporo, Japan), 1882–1887.

Liao, H., Pundak, G., Siohan, O., Carroll, M. K., Coccaro, N., Jiang, Q. M., et al. (2015). "Large vocabulary automatic speech recognition for children," in *Interspeech*, (Dresden, Germany), 1611–1615.

Long, M. H. Ed. (2006). "Recasts in SLA: the story so far," in *Problems in SLA. Second Language Acquisition Research Series*, (Mahwah, NJ: Lawrence Erlbaum Associates), 75–116.

Lund, H. H. (2003). "Adaptive robotics in the entertainment industry," in *Proceedings of IEEE International Symposium on Computational Intelligence in Robotics and Automation (cira2003)*, (Vol. 2), (Kobe, Japan), 595–602.

Mashburn, A. J., Justice, L. M., Downer, J. T., and Pianta, R. C. (2009). Peer effects on children's language achievement during pre-kindergarten,. *Child Dev.* 80, 686–702. doi: 10.1111/j.1467-8624.2009.01291.x

Matthews, D., Lieven, E., and Tomasello, M. (2007). How toddlers and preschoolers learn to uniquely identify referents for others: a training study. *Child Dev.* 6, 1744–1759. doi: 10.1111/j.1467-8624.2007.01098.x

McGillion, M., Herbert, J., Pine, J., Keren-Portnoy, T., Vihman, M., and Matthews, D. (2013). Supporting early vocabulary development: what sort of responsiveness matters? *IEEE Trans. Auton. Ment. Dev.* 5, 240–248. doi: 10.1109/tamd.2013.2275949

Nguyen, T. L., Boukezzoula, R., Coquin, D., Benoit, E., and Perrin, S. (2015). "Interaction between humans, NAO robot and multiple cameras for colored objects recognition using information fusion," in *8th International Conference on Human System Interaction* (HSI), (Warsaw, Poland), 322–328.

Schodde, T., Bergmann, K., and Kopp, S. (2017). "Adaptive robot language tutoring based on bayesian knowledge tracing and predictive decision-making," in *Proceedings of the 12th ACM/IEEE International Conference on Human-Robot Interaction* (HRI 2017), (Vienna, Austria).

Shahid, S., Krahmer, E., and Swerts, M. (2014). "Child–robot interaction across cultures: how does playing a game with a social robot compare to playing a game alone or with a friend? *Comput. Human Behav.* 40, 86–100. doi: 10.1016/j.chb.2014.07.043

Tanaka, F., and Matsuzoe, S. (2012). Children teach a care-receiving robot to promote their learning: field experiments in a classroom for vocabulary learning. *J. Hum. Robot Interact.* 1, 78–95. doi: 10.5898/jhri.1.1.tanaka

Tazhigaliyeva, N., Diyas, Y., Brakk, D., Aimambetov, Y., and Sandygulova, A. (2016). "Learning with or from the robot: exploring robot roles in educational context with children," in *International Conference on Social Robotics*, eds A. Agah, J.-J. Cabibihan, A. M. Howard, M. A. Salichs and H. He (New York, NY: Springer International Publishing), 327–336.

Tomasello, M., and Farrar, M. J. (1986). Joint attention and early language. *Child Dev.* 57, 1454–1463. doi: 10.2307/1130423

Vygotsky, L. (1978). *Mind in Society*. Harvard: Harvard University Press.

Westlund, J. M. K., Martinez, M., Archie, M., Das, M., and Breazeal, C. (2016). "Effects of framing a robot as a social agent or as a machine on children's social behavior," in *Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, eds S. Y. Okita, T. Shibata and B. Mutlu (Washington, DC: IEEE), 688–693.

Yu, C., and Smith, L. B. (2016). The social origins of sustained attention in one-year-old human infants. *Curr. Biol.* 26, 1235–1240. doi: 10.1016/j.cub.2016.03.026

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix II. The storyboards of lesson 1 and 2.

		Robot		Tablet					
#	Input (touch and speech)	Text L1	Text L2	Scene	Objects	Say	Comment	Instructional methodology	Function
		[Introduction: 2 minutes]							
1		<tablet(off)><Gaze(child)><Face(Neutral)>Hello <name>. Let's play together! Do you like games? <wait(2000)>					probably yes, answer doesn't really matter	introduction	
2	voiceActivity, timer(2second)	I really like games! <Face(Happy)> <tablet(on)>Look <Gaze(tablet)> we will visit a new place today. <Face(Neutral)> Look where we're going today! <Gaze(child)> <Gesture(Pretends to touch tablet)>		town			The robot and child avatars are walking towards the zoo.	introduction	
3		<Gaze(child)>Cool <Face(Happy)>, today we'll visit the zoo! I really like animals. Do you also like animals? <wait(2000)>			display zoo		probably yes, answer doesn't really matter	introduction	
4	probabilityGesture	<Face(Neutral)>Touch <pointAt(tablet)> <Gaze(tablet)> the zoo and we will enter it! <Gaze(child)>						instructions	useGesture(probabilityGesture)
5	objectTouched, voiceActivity, support	<giveResponseToSelectObject(zoo)>					did the child select the correct object?		<giveResponseToSelectObject>
		[Modelling of words: about 10 minutes]							
6		Cool <Face(Happy)>, a monkey! Touch it <pointAt(tablet)> <Gaze(tablet)> and we'll <Gaze(child)> hear the English word for monkey.		zoo_1					useGesture(probabilityGesture)
7	objectTouched, voiceActivity, support	<giveResponseToSelectObject (monkey_1)>				monkey	did the child select the correct object?		<giveResponseToSelectObject>
8		<Gaze(child)>Ah, <tablet(off)>an monkey is in English a Can you also say	monkey monkey						
9	objectTouched, voiceActivity, support	<giveResponseOnSpeech(monkey)>					did the child say something?		giveResponseOnSpeech
10		Now, I think there's a very important task for us! <tablet(on)>The monkey is loose and we have to put them in their cage! Put the in its cage <pointAt(tablet)> <Gaze(tablet)> .	monkey		display arrow(between)				useGesture(probabilityGesture)
11	objectTouched, voiceActivity, support	<giveResponseToMoveObject(monkey_1, in, cage_1)>			remove arrow and display stars	<happy_sound>	did the child move the correct object? Elephants go in their cage elephant and the elephand makes a happy elephant sound,		<giveResponseToMoveObject>
12		Now we have one . Let's hear what one is in English. Can you touch the in the cage	monkey monkey						useGesture(probabilityGesture)
13	objectTouched, voiceActivity, support	<giveResponseToSelectObject (monkey_1, cage_1)>				one monkey	did the child select the correct object?		<giveResponseToSelectObject>
14		<tablet(off)>so one is say	one one						

15	objectTouched, voiceActivity, support	<giveResponseOnSpeech(one)>				did the child say something?		giveResponseOnSpeech
				display elephant_1 and elephant_2				
16		<tablet(on)><Gaze(tablet)> Cool, elephants! Touch them <pointAt(tablet)> and we'll <Gaze(child)> hear the English word for elephant.						useGesture(probabilityGesture)
17	objectTouched, voiceActivity, support	<giveResponseToSelectObject(elephant)>			elephant	did the child select the correct object?		<giveResponseToSelectObject>
18		<tablet(off)>Ah, an elephant is in English an	elephant					
		Can you also say	elephant					
19	objectTouched, voiceActivity, support	<giveResponseOnSpeech(elephant)>				did the child say something?		giveResponseOnSpeech
20		<tablet(on)>let's see <Gaze(tablet)> what we have to do now.						
		<Gaze(child)> The elephants are loose and we have to put them in their cage! Put	one elephant					useGesture(probabilityGesture)
	probabilityGesture	in its cage. <pointAt(tablet)> <Gaze(tablet)> .		display arrow between elephant and cage				useGesture(probabilityGesture)
21	objectTouched, voiceActivity, support	<giveResponseToMoveObject(elephant, in, cage_2)>		remove arrow and display stars	<happy_sound>	did the child move the correct object? Elephants go in their cage elephant and the elephand makes a happy elephant sound, child and robot receive a star		<giveResponseToMoveObject>
22		<Gaze(child)> there is still	one elephant	remove stars				
	probabilityGesture	outside of the cage. Add it to the cage <Gaze(tablet)> and we <Gaze(child)> will hear what 'add' is						useGesture(probabilityGesture)
23	objectTouched, voiceActivity, support, target	<giveResponseToMoveObject (elephant, in, cage_2)>			add	did the child select the correct object?		<giveResponseToSelectObject>
26	probabilityGesture	<Gaze(child)> Now there are two elephants in the cage! Touch <pointAt(tablet)><Gaze(tablet)> them, then we'll <Gaze(child)> hear what two is.		remove stars				useGesture(probabilityGesture)
27	objectTouched, voiceActivity, support, target	<giveResponseToSelectObject(cage_2, elephant)>			two	did the child select the correct object?		<giveResponseToSelectObject>
28		<tablet(off)>	two					
		say	two					
29	objectTouched, voiceActivity, support, target	<giveResponseOnSpeech(two)>				did the child say something?		giveResponseOnSpeech
30		<tablet(on)><Gaze(child)> Great, we have	two elephants					

		in the cage. One of the cages has more animals. Touch <pointAt(tablet)><Gaze(tablet)> the cage with more animals than the other, and we'll <Gaze(child)> hear the English word for more.						useGesture(probabilityGesture)
31	objectTouched, voiceActivity, support, target	<giveResponseToSelectObject(cage_2)>			more	did the child select the correct object?		<giveResponseToSelectObject>
		<Gaze(child)> Now, lets go to the next cage						useGesture(probabilityGesture)
32	probabilityGesture			display giraffe_1, giraffe_2 and giraffe_3				
		Cool	more					
		animals! Now we have	one monkey, two elephants					
	probabilityGesture	and three giraffes. Let's see what that is in English. Touch <pointAt(tablet)> the giraffe <Gaze(tablet)> and let's find out.						useGesture(probabilityGesture)
33	objectTouched, voiceActivity, support, target	<giveResponseToSelectObject(giraffe)>			giraffe	did the child select the correct object?		<giveResponseToSelectObject>
34		<Gaze(child)><tablet(off)>	giraffe					
		say	giraffe					
35		<giveResponseOnSpeech(giraffe)>						
36		<tablet(on)>Nice, there are	more giraffes					
		than other elephants and monkeys. Look at this, we have another important task. Let's put the	*giraffe*					
		in its cage!						useGesture(probabilityGesture)
37	objectTouched, voiceActivity, support, target	<giveResponseToMoveObject(giraffe, in, cage_3)>		remove arrow		While dragging the tablet says 'add'		<giveResponseToMoveObject>
38		<Gaze(child)>There are still	two giraffes					
		outside of the cage. There are	more giraffes					
		outside of the cage than inside of the cage. Can you	add one giraffe					useGesture(probabilityGesture)
		to the cage?						
39		<giveResponseToMoveObject(giraffe,in,cage_3)>						
40		<Gaze(child)> now there are	two giraffes	remove arrow, display stars	<happy_sound>	elephant makes a happy elephant sound, and the child and the robot receive a star		
		Now there are	*more giraffes*					
		inside the cage than outside the cage. Can you	add one giraffe?		add			useGesture(probabilityGesture)
41		<giveResponseToMoveObject(giraffe,in,cage_3)>						
42		Now <Gaze(tablet)>there are three giraffes! This cage has	more					
		animals in it than the other cages. Touch them <pointAt(tablet)>, then we'll hear what three is?						useGesture(probabilityGesture)

						three giraffes			
43		<giveResponseToSelectObject(cage_3, giraffe)>							
44		<tablet(off)>	three giraffes						
		say	three						
45		<giveResponseOnSpeech(three)>							
46		<Gaze(child)> You did an awesome job, <tablet(on)>	*one monkey, two elephants and three giraffes*						
		are back in their cage! One cage has most animals <Gaze(tablet)>. Touch <pointAt(tablet)> the cage with most animals, and we'll hear the English word for most.							useGesture(probabilityGesture)
47		<giveResponseToSelectObject(cage_3, giraffe)>				most			
48		<tablet(off)>	most						
		say	most						
49		<giveResponseOnSpeech(most)>							
50		<Gaze(child)> Out of all the animals, I like giraffes	most						
		Which animal do you like	most						
		Do you like elephants, monkeys or giraffes	most						
51		<wait(2000)>							
		Cool! <tablet(on)>The last thing we need to do is put food in the cage with the giraffes. <Gaze(tablet)>This cage has							
52		the	most						
		so they need the	most						
		food. Put these trees in the cage so the giraffes can eat from							useGesture(probabilityGesture)
		them. We <Gaze(child)> have	three						
		giraffes so we need	three						
		trees. Can you put <pointAt(tablet)><Gaze(child)> the trees in the cage? Count them while dragging			display arrow between trees and cage				useGesture(probabilityGesture)
53		<giveResponseToMoveObject(tree,in,cage_2)>	one						
54		<giveResponseToMoveObject(tree,in,cage_2)>	two						
55		<giveResponseToMoveObject(tree,in,cage_2)>	three						
		Great! Now <Gaze(tablet)>each giraffe has their own tree,							
56		because there are	three		remove arrow				
		giraffes and	three						
		trees! <tablet(off)>You <Gaze(child)> did great, now let's play another game!							
		[Task: about 5 minutes]							
57		<tablet(on)> Look! <Gaze(tablet)><pointAt(tablet)> There are some	giraffes	zoo_2	giraffe_1, giraffe_2,giraffe_3,giraffe_4,giraffe_5,giraffe_6, cage_1, foodtray_1, lake_1		Tablet displays a new screen with 6 giraffes in their cage, a food tray next to the cage, and a lake next to the cage.)	instructions	useGesture(probabilityGesture)
	probabilityGesture	in the cage, <Gaze(child)> but today the weather is very nice and they're going to swim in the lake! Put in the lake.	one giraffe						
58	objectTouched, voiceActivity, support, target	<giveResponseToMoveObject(giraffe,in,lake_1)>					did the child move the correct object?		<giveResponseToMoveObject>

59		<Gaze(child)> Can you put <pointAt(tablet)><Gaze(tablet)> next to the food tray?	two giraffes		display confetti	<happy_sound>	the giraffe makes a happy sound (maybe also like confetti or something else festive)		useGesture(probabilityGesture)
60		<giveResponseToMoveObject(giraffe, next to, foodtray_1)>							
61		<giveResponseToMoveObject(giraffe, next to, foodtray_1)>							
62		<Gaze(child)> Can you to the lake?	add two giraffes		remove confetti			instructions	useGesture(probabilityGesture)
63	objectTouched, voiceActivity, support, target	<giveResponseToMoveObject(giraffe,in,lake_1)>					did the child move the correct object?		<giveResponseToMoveObject>
64		<giveResponseToMoveObject(giraffe,in,lake_1)>							
65		<Gaze(child)> Where are Touch <Gaze(tablet)> <pointAt(tablet)>the area on the tablet	three giraffes?		display confetti remove confetti	<happy_sound>	the giraffe makes a happy sound (maybe also like confetti or something else festive)	instructions	useGesture(probabilityGesture)
66	objectTouched, voiceActivity, support, target	<giveResponseToSelectObject(lake_1)>					did the child select the correct object?		<giveResponseToSelectObject>
67	probabilityGesture	<Gaze(child)> Where are in the cage or next to the food tray? Touch <Gaze(tablet)> <pointAt(tablet)>the area on the tablet.	more giraffes					instructions	useGesture(probabilityGesture)
68		<giveResponseToSelectObject(foodtray)>							
69		<Gaze(child)> where are the Touch <Gaze(tablet)><pointAt(tablet)> the area on the tablet	most giraffes?						useGesture(probabilityGesture)
70	objectTouched, voiceActivity, support, target	<giveResponseToSelectObject(lake_1)>					did the child select the correct object?		<giveResponseToSelectObject>
		[End of lesson]							
71		<Gaze(child)>Yay we got a star! It was very nice to play with you! Bye!		town	display robot and child and big_star		(On the tablet, the 'home screen' with the town and the avatars pops back up again. The child and the robot receive a big star for their work, in a festive animation.)		

#	Robot: Text L1	Robot: Text L2	Tablet: Scene	Tablet: Objects	Tablet: Say	Comment	Instructional methodology	Function
	Introduction: 2 minutes							
1	<tablet(off)><Gaze(child)><Face(Neutral)>Hello <name> I really liked <Face(Happy)>playing with you last time! Did you also like to play last time?<Face(Neutral)>					probably yes, answer doesn't really matter	introduction	
2	<wait(2000)> Today we'll play another game, because we're going to a different part of the town! <Face(Happy)> <tablet(on)>Look <Gaze(tablet)> Let's see where we're going today! Touch <pointAt(tablet)> <Gaze(tablet)>the tablet to begin <Gaze(child)>		town					
3	<giveResponseToSelectObject(town)>			display flowershop				
4	<Face(Neutral)>Cool, today we're going to help in the flower shop! Touch <pointAt(tablet)> <Gaze(tablet)> the tablet to begin! <Gaze(child)>		town			The robot and child avatars are walking towards the zoo.	introduction	
5	<giveResponseToSelectObject(flowershop)>			display flowershop				
	Modelling of words: about 12 minutes					probably yes, answer doesn't really matter	introduction	
6	Wow <Face(Happy)>, there are so many flowers! Touch <pointAt(tablet)> <Gaze(tablet)>the flowers to hear <Gaze(child)> the English word for flowers.		flowershop_1					
7	<giveResponseToSelectObject(flower)>				flower		instructions	useGesture(probabilityGesture)
8	<Gaze(child)>Ah,<tablet(off)> in English flowers are	flowers.				did the child select the correct object?		<giveResponseToSelectObject>
	Can you also say	flowers?						
9	<giveResponseOnSpeech(flowers)>							useGesture(probabilityGesture)
10	<tablet(on)>Look <Gaze(tablet)>, there are a lot of empty vases. I think we <Gaze(child)> have to fill them with flowers! Touch <pointAt(tablet)> <Gaze(tablet)> the yellow vase and we'll <Gaze(child)>hear what to do.					did the child select the correct object?		<giveResponseToSelectObject>
11	<giveResponseToSelectObject(vase_1)>				add two flowers			
12	Cool! Can you	add two flowers,						
	so add <pointAt(tablet)> <Gaze(tablet)> two flowers?<gaze(child)>			display arrow(between)		did the child say something?		<giveResponseOnSpeech()>
13	<giveResponseToMoveObject(flower,in,vase_1)>				one			useGesture(probabilityGesture)
14	<giveResponseToMoveObject(flower, in, vase_1)>				two			

15	<gaze(child)> Now, let’s fill the orange vase! Touch <pointAt(tablet)> <Gaze(tablet)> the orange vase and we’ll <gaze(child)> hear what to do.		remove arrow				
16	<giveResponseToSelectObject(vase_2)>			add three flowers	did the child move the correct object? Elephants go in their cage elephant and the elephand makes a happy elephant sound, child and robot receive a star		<giveResponseToMoveObject>
17	Cool! Can you so add <pointAt(tablet)> <Gaze(tablet)> three flowers?<gaze(child)>	add three flowers	display arrow(between)				
							useGesture(probabilityGesture)
18	<giveResponseToMoveObject(flower,in,vase_2)>			one			
19	<giveResponseToMoveObject(flower, in, vase_2)>			two			
20	<giveResponseToMoveObject(flower, in, vase_2)>		remove arrow	three			
21	<gaze(child)> There’s still room for another flower! Let’s <pointAt(tablet)> <Gaze(tablet)>	add one flower	display arrow(between)				
22	<giveResponseToMoveObject(flower, in, vase_2)>		remove arrow and display stars	<happy_sound>			
23	<gaze(child)>Cool, now we have four flowers in the vase! Touch <pointAt(tablet)> <Gaze(tablet)>the vase to hear the English word for four.						
24	<giveResponseToSelectObject(vase_2)>			four			useGesture(probabilityGesture)
25		<tablet(off)>four					
	say	four		one monkey	did the child select the correct object?		<giveResponseToSelectObject>
26	<giveResponseOnSpeech(four)>						
27	<tablet(on)>Great! Now, let’s fill the other vase. Touch <pointAt(tablet)> <Gaze(tablet)>the green vase and we’ll <gaze(child)>hear what to do.						
28	<giveResponseToSelectObject(vase_3)>			add four flowers	did the child say something?		<giveResponseOnSpeech()>
29	Cool! Can you	add four flowers?	display arrow				
30	<giveResponseToMoveObject(flower, in, vase_3)>			one			useGesture(probabilityGesture)
31	<giveResponseToMoveObject(flower, in, vase_3)>			two			useGesture(probabilityGesture)

32	<giveResponseToMoveObject(flower, in, vase_3)>			three		useGesture(probabilityGesture)
						useGesture(probabilityGesture)
33	<giveResponseToMoveObject(flower, in, vase_3)>		remove arrow	four		
34	There’s still room for another flower! Let’s	add one flower.	display arrow			
35	<giveResponseToMoveObject(flower, in, vase_3)>		remove arrow and display stars	<happy_sound>		
36	<gaze(child)>Cool, now we have five flowers in the vase! Touch <pointAt(tablet)> <Gaze(tablet)> the vase to hear <gaze(child)> the English word for five.				did the child say something?	<giveResponseOnSpeech()>
37	<giveResponseToSelectObject(vase_3)>			five		
38		<tablet(off)><gaze(child)>five				useGesture(probabilityGesture)
	say	five				useGesture(probabilityGesture)
39	<giveResponseOnSpeech(five)>				did the child move the correct object? Elephants go in their cage elephant and the elephand makes a happy elephant sound, child and robot receive a star	<giveResponseToMoveObject>
40	Great! <tablet(on)><gaze(tablet)>We’ve filled all <gaze(child)>the vases with flowers: the yellow one has the orange one has and the green one has Which vase has fewer flowers than the orange vase? Touch <pointAt(tablet)><gaze(tablet)> the tablet and we’ll <gaze(child)> hear the English word for fewer.	*two flowers* *four flowers* *five flowers*				
40	<giveResponseToSelectObject(vase_1)>			fewer		
41		<tablet(off)><gaze(child)>fewer				useGesture(probabilityGesture)
	say	fewer				useGesture(probabilityGesture)
42	<giveResponseOnSpeech(fewer)>					
43	<tablet(on)>Which vase has the fewest flowers? Touch <pointAt(tablet)> <gaze(tablet)> the tablet and we’ll <gaze(child)>hear the English word for fewest					
44	<giveResponseToSelectObject(vase_1)>			fewest		

45		<tablet(off)><gaze(child)>fewer	fewest					useGesture(probabilityGesture)
	say	fewest						useGesture(probabilityGesture)
46	<giveResponseOnSpeech(fewest)>							
47	<tablet(on)> <gaze(child)>Hm I think the green vase is a bit too full now with The should be in it. Let's take away <pointAt(tablet)><gaze(tablet)> and put in the vase with the then we'll <gaze(child)> hear what the English word for take away is.	*five flowers* *fewer flowers* *one flower* *fewest flowers*		display arrow				
48	<giveResponseToMoveObject(flower, from, vase_3,in,vase_1)>				take away			
49		<tablet(off)><gaze(child)>take away		remove arrow				useGesture(probabilityGesture)
	say	take away						useGesture(probabilityGesture)
50	<giveResponseOnSpeech(take away)>							
51	The tablet said to from the green vase, so you Now it has than before. It now only has left, and that is <tablet(on)><gaze(tablet)>Let's see what we have to do now <Gesture(Pretends to touch tablet)>	*take away one flower* *took away one flower* *fewer flowers* *four flowers* *fewer than five*						
52				display shelve_1, shelve_2, shelve_3, bucket_1,butcket_2,bucket_3		Now not only the tablet with vases, but also the shelves in the background are visible. There are three shelves. There are two buckets on two shelves and one bucket on one shelf.)		
53	<gaze(child)>Okay, let's put away our nice vases to make some room! We should put them on the shelves, but they're filled with buckets. Touch <pointat(tablet)> < gaze(tablet)>a bucket to hear the English word for bucket<gaze(child)>							
54	<giveResponseToSelectObject(bucket)>				bucket			useGesture(probabilityGesture)
55		bucket						
	say	bucket						

56	<giveResponseOnSpeech(bucket)>						
57	Which shelve has the Touch <pointAt(tablet)> <Gaze(tablet)> the shelf with the	*fewest buckets?* *fewest buckets* *<gaze(child)>*		good job, that was the shelf with the *fewest buckets'*			
58	<giveResponseToSelectObject(shelf_1)>						
59	Okay, there are on the shelves! There should be if we want to store the vases. I guess we need to a lot of buckets to make space for the vases. to make room for the vases. Start with the shelf with the	*five buckets* *fewer buckets* to *take away* *Take away five buckets* *fewest buckets* .	arrow				
60	<giveResponseToMoveObject(bucket_1,on,ground)>						
61	<giveResponseToMoveObject(bucket_2,on,ground)>						
62	<giveResponseToMoveObject(bucket_3,on,ground)>						
63	<giveResponseToMoveObject(bucket_4,on,ground)>						
64	<giveResponseToMoveObject(bucket_5,on,ground)>						
65	The tablet said to Now there are and there is enough room on the shelves to put the vases there. Put the vases on the shelves.	*take away five buckets* *fewer buckets*	arrow				
66	<giveResponseToMoveObject(vase,on,shelf)>						
67	<giveResponseToMoveObject(vase,on,shelf)>						
68	<giveResponseToMoveObject(vase,on,shelf)>						
71	Wow look how good we helped in the flower shop! I think the owner will be very pleased with us! Let's play another game!				Tablet displays a new screen with 6 giraffes in their cage, a food tray next to the cage, and a lake next to the cage.)	instructions	useGesture(probabilityGesture)
72	TASK: about 5 minutes Wow look at that bucket with the huge amount of flowers, it's overflowing! We also have 2 cool vases, a red one and a blue one. Let's take flowers from the bucket, and let's create two nice bouquets with it. from the bucket and put them in the red vase.	Take away five flowers					
73	<giveResponseToMoveObject()>				did the child move the correct object?		<giveResponseToMoveObject>
74	Now from the bucket and put them in the blue vase.	take away six flowers	display confetti	<happy_sound>	the giraffe makes a happy sound (maybe also like confetti or something else festive)		useGesture(probabilityGesture)
75	<giveResponseToMoveObject()>						
76	Which bucket has	fewer flowers?	remove confetti			instructions	useGesture(probabilityGesture)
77	Touch <pointAt(tablet)> <Gaze(tablet)> that bucket.						

78	<giveResponseToSelectObject()>				did the child move the correct object?		<giveResponseToMoveObject>
79	Now	take away two flowers					
	from the big bucket and put them in the red vase.		display confetti	<happy_sound>	the giraffe makes a happy sound (maybe also like confetti or something else festive)		
80	<giveResponseToMoveObject()>		remove confetti			instructions	
81	Which bucket has	seven flowers			did the child select the correct object?		<giveResponseToSelectObject>
82	Touch <pointAt(tablet)> <Gaze(tablet)> that bucket.					instructions	
83	<giveResponseToSelectObject()>						useGesture(probabilityGesture)
	Ending of lesson						
84	Yay we got another big star! I really liked playing with you! Bye!						