The effect of a robot's gestures on engagement and learning gain in second language learning

Tilburg University

Master Communication and Information Sciences

Cognitive Sciences and Artificial intelligence

Master thesis

Supervisors: P. Vogt & M. Alimardani

Reinjet Oostdijk

ANR: 481926

SNR: 2008556

21 January 2019

Abstract

Technology is integrated into our lives and therefore also used in educational environments. In most elementary schools children use laptops or tablets as E-learning devices as part of their educational program. Lately, research is done with robots in educational settings. A long-term study, named L2TOR, investigates how a robot can be used to teach a second language to five year old children. Preschool children followed a series of seven lessons with a robot where they learned several English words. In this thesis, data of the L2TOR project is used. Two conditions are compared with each other, in one condition the robot used iconic and deictic gestures, while in the other condition the robot used only deictic gestures. For both conditions, children's task engagement and robot engagement is studied. Furthermore, children's learning gain is evaluated. Results show that there was no positive effect of conditions on robot engagement and neither on task engagement. Moreover, the level of engagement (robot and task) dropped over time. In both conditions, learning gain was found, but there was no difference between the conditions. Finally, there was a positive relation between engagement and learning gain, students who were more engaged had a greater learning gain.

Keywords: child-robot interaction, engagement, gestures, learning gain, education

Table of contents

Introduction

Due to globalization, it is nowadays important to speak more than one language. A second language can be learned anytime, however, children learn a second language faster and easier than adults do (Johnson & Newport, 1989). Therefore children often start learning a second language in elementary school. For some children, the language used at school is already the second language, since they speak a different language at home. When children start elementary school, there is already a difference between their linguistic knowledge. Some children only speak one language, while others know two or more languages. And even if two children speak just one and the same language, their knowledge can differ. There are multiple factors that influence the learning ability of the child during second language learning, for example: motivation (Dörnei, 1990), aptitude and attitude (Schumann, 1975). This makes it harder for some children to learn a second language at school than for others, which results in different needs per student. The biggest challenge of teaching a second language is the interaction between teacher and students, mostly in that there are too many students to provide each of them in their individual needs. Some children need a more challenging program than others and also the speed of the lessons can differ per child. It is therefore preferred to teach children in a smaller class or ideally one-on-one (Brühwiler & Blatchford, 2011; Kim, Curby & Winsler, 2014).

The use of technology in the classroom changed completely over the last two decades. The computer entered the classroom setting in the early '90s and children played on it once in a while (Clements, 1998). Currently, almost every child has their own tablet and uses it daily as part of their educational program. As said before, children all have specific needs when it comes to learning, but meeting those needs is mostly not possible with human teachers due to lack of money and time. The use of technology in education is a solution and can contribute to the specific needs of the students. E-learning environments are created in such a way that they are responding to the performance of the student.

Recent developments in artificial intelligence and robotics made it possible to bring robots to the classroom. Researchers are experimenting with the use of robots in educational settings in various ways. Mostly, robots are used as a tutor who helps the child in their learning process (Mubin, Stevens, Shahid, Mahmud & Dong, 2013). One of the advantages of a robot compared with a tablet or laptop is that it can show expressions such as gestures. Gestures can be very helpful in a learning process, especially when it comes to language learning (Tellier, 2005). Gestures can help in the understanding of unfamiliar words and using the right amount of gestures can help the student with their attention in the lesson. However, too fuzzy gestures can lead to distraction (Tai, 2014).

In summary, every individual student has different needs when it comes to learning. Technology is a great outcome to create a personalized learning environment where the students needs are taken into account. Additionally, when it comes to language learning, gestures can play an important role in the learning process. Therefore, a robot who makes gestures could be a great outcome in language learning. De Wit et al. (2018) presents a study where both are integrated (gestures and technology) and examined the effect of social robots with iconic gestures in second language learning. The results of this study show the importance of gestures, children in the iconic gesture condition had a larger learning gain (measured after one week) and were more engaged in the task than children in the no gestures condition. The study by De Wit et al. (2018) consisted of only one training session with the children and therefore raises the question if the same effects will be found on long-term. In addition, De Wit et al. (2018) focussed only on iconic gestures compared with no gestures. Since there are different kinds of gestures (McNeill, 1992), this thesis will study the differences between only deictic gestures and deictic gestures in combination with iconic gestures. To investigate this, the following research question has been formulated:

RQ:

Does the engagement of the students in the lessons influence their learning gain and do those results differ between children who are tutored by a robot performing only deictic gestures and a robot performing deictic and iconic gestures?

Theoretical framework

Robots in Education

Recent developments in artificial intelligence and robotics have allowed the automation of many activities and work processes. It is used for transportation, healthcare, legal services and also for education. A study by Ivanov (2016) showed the advantages and disadvantages of robot teachers. One of the advantages is that a robot can work 24 hours a day, 7 days per week. Thereby, robots can be used for various subjects, they are not specialised in a specific field, they provide constant or improving quality of their work and they can repeat the explanations numerous times. But, on the other hand, robots need a human to prepare the materials, they are not creative and not capable at social communication (Yang, Bellingham, Dupont, Fischer, et all., 2018). However, the benefits outweigh the disadvantages and therefore a lot of researchers are studying the possibilities of using robots for education. A robot can take different roles in the learning process, depending on the learning goal. In robotics education, the robot can be used as a learning tool / teaching aid, where students would be building and programming robots (Mubin, et al., 2013). In more general education, the robot can take the role of tutor or peer. As tutor, the robot gives hints or instructions to the child, like a teacher will do. As peer (or co-learner), the robot participates actively and spontaneously in the task and adopts empathetic behaviours towards the child (Mubin et al., 2013; Charisi, Davison, Reidsma & Evers, 2016). Moreover, children feel supported by a peer robot, in the same way they feel supported by their fellow students (Leite, 2014).

Most educational systems involve learning at school, but also at home, therefore children get daily or weekly homework. Especially at a young age, parents play a role in helping their children with their homework, but this is not always easy in their hectic lives. Therefore, technology can be a solution. Nowadays, homework can also be provided on a computer in an e-learning environment. Han, Jo, Jones and Jo (2008) studied the interaction between children and a home robot. They examined children's interest, concentration and achievements in English learning and compared those effects with other instructional media: books in combination with audio tape and a computer. The results show that social robots have a distinct advantage over the book with audio tape and the computer. When the social robot was used, children performed better and were more focused than when the same content was delivered via the book with audio tape or the computer. Those results were found since a robot is able to provide different types of interactions with the child. Additionally, it can show expressions such as gestures, motions and facial expressions. However, those results are in contrast with a study by Westlund, Dickens, Jeong, Harris, DeSteno & Breazeal (2015), where no difference was found in performance of the children. Here, children followed two sessions with a human teacher, tablet or robot where they learned animal names. Results show the three different sources of information (human, tablet and robot) did not influence the performance of the child. But, children showed a clear preference for learning with the robot, since their enthusiasm and engagement was higher than other children. Those results are comparable with the results of a study by Serholt (2014) where children had to perform a task instructed by a humanoid robot or a human tutor. There was also no difference in the performance of the children. However, in this study the engagement of the children did not differ between both conditions.

In summary, the use of technology in classrooms is rapidly increasing. Where twenty years ago the computer made its entrée to the classroom, nowadays children learn via their own tablets and researchers are experimenting with the use of robots in the classroom. It seems that robots have several advantages compared to other learning tools or even humans. On the other hand, children do not necessarily perform better when their tutor is a robot compared to a human tutor.

Gestures in language learning

Various research has been done in the field of language learning methods and theories. Depending on someone's personality, ability to learn and experiences in language learning, a specific technique can be used to learn a second language (Gu & Johnson, 1996; Dörnyei, 2003). If a language is taught by a teacher, different tools can support learning, such as gestures. In a classroom, students' comprehension is often challenged by instructional discourse that presents new concepts and uses unfamiliar terms. Moreover, classrooms can be very noisy, with multiple people talking at the same time. So under such circumstances, gestures can play an important role in comprehension (Alibali & Nathan, 2007; Corts & Pollio, 1999). There are four types of gestures a person can use: metaphoric, beat, deictic and iconic gestures. Metaphoric gestures present an image of an abstract concept, such as knowledge or the genre of the narrative. Beats are movements that do not relate to anything that is said, they are mostly small, low energy, rapid flicks. Deictic gestures are pointing movements that may refer to real objects, like a pointing finger (McNeill, 1992). Iconic gestures have a visually similar relationship to the action, object or attribute they portray (Eisenstein & Davis, 2006). Studies have shown that the use of iconic gestures can help the student to reiterate teacher speech, which eventually helps their learning (Roth, 2001).

Research has repeatedly demonstrated that iconic gestures, when performed by a human teacher, has a beneficial effect on second language learning performance (Kelly, McDevitt & Esch, 2009; Tellier, 2008; Macedonia, Müller & Friederici, 2011). Kelly et al. (2009) experimented with congruent gestures against incongruent gestures, both in combination with speech. Results show that participants learned more words in the congruent gestures condition. This suggests that the representational meaning of hand gestures that simultaneously accompany speech may index a newly learned word to an established concept. This connection may help those words endure longer in memory. In a study by Tellier (2005) children in the age of 5 were presented with a video of a person who only pronounced the words or additionally presented them with iconic gestures congruent with the meaning of the words they had to learn. Children who were exposed to the video with gestures, performed better on a free recall task compared to children who learned words without gestures. Not only effect is found in showing gestures to the learners, also imitating gestures helps second language learners in their learning process. Stevanoni and Salmon (2005) found that children who were instructed to use gestures during an interview were able to report (two weeks later) more correct information than children who were prevented from gesturing. Similar effects were found in a study by De Nooijer, Van Gog, Paas & Zwaan (2012), gesture imitation led to better recall than gesture observation, but an effect of imitation was only present for the object-manipulation verbs (e.g., to chisel) and not for the abstract and locomotion verbs. Showing gestures do not only help to acquire knowledge, but also to retain it overtime (Cook, Mitchell & Goldin-Meadow, 2007).

Since a positive effect of using gestures in second language learning is found with a human teacher, similar research is done with virtual agents (Bergmann & Macedonia, 2013) and robots (Van Dijk, Torta & Cuijpers, 2013; De Wit et al., 2018). Bergmann & Macedonia (2013) studied the effect of iconic gestures performed by a virtual agent compared to iconic gestures performed by a human. The results showed improved memory scores when participants learned with a virtual agent. Van Dijk et al. (2013) studied the differences between the presence and absence of gestures performed by a robot in a human-robot interaction. Results show that the use of iconic gestures aids retention. De Wit et al. (2018) examined the effect of social robots with iconic gestures in second language learning. In this experiment, Dutch children in the age of four till six played a game where they learned the words for six different animals in English. In one experimental condition, the robot did not use any gestures. In the second condition the robot used iconic gestures, for example by moving his arms up and down to show 'bird'. The results of this study show that there is a larger learning gain (measured after one week) when the robot shows iconic gestures with the target words, than when the robot does not show any gestures. So, not only a positive effect of using gestures is found for human teachers, similar results are found for virtual agents and robots.

Engagement and learning

Sider, Lee, Kidd, Lesh & Rich (2005) explains engagement as a process by which individuals in an interaction start, maintain and end their perceived connection to one another. Mastin & Vogt (2016) described engagement as the complex ways individuals interact with and within their environment. They distinguish two different kinds of engagement: solitary engagement and joint engagement. Solitary engagement occurs when an individual does not interact with others, the individual may watch others or interact with only objects. Joint engagement occurs when an individual interacts with other individuals or a group. Nakano & Ishii (2010) state that engagement signals are bodily expressions, such as facial expressions and gestures and nonverbal signals, such as eye gaze and head nods.

Engagement is an important aspect in learning. Webber (2004) created a social cognitive model which shows the dynamic nature of the engagement between learner and environment. This approach emphasises the relationship between the perception and experience of the student and their engagement with education. If a student is engaged in the lesson, and therefore is more attentive, he or she will receive and process more information than when the student is less engaged. Linnebrink & Pintrich (2010) distinguished three types of engagement in educational settings, all needed for learning and achievement. The first type is behavioural engagement, which involves observable behaviour such as effort (e.g. working hard on a task), persistence and help-seeking. The second type of engagement is cognitive engagement, paying attention (behavioural engagement) may not be enough for learning. When a student is cognitively engaged, he or she thinks deeply about the content to be learned, uses different strategies for learning and thinks critically and creatively about the material to be learned. The last type of engagement is motivational engagement, meaning that the student is interested in the task and having a positive emotional or affective experience while they are learning.

In the study by De Wit et al. (2018) not only a positive effect of gestures was found on language learning, gestures also had a positive effect on engagement. Children were more engaged in the lesson when their tutoring robot used gestures compared to children which tutoring robot did not use any gestures. Belpaeme et al. (2012) focused on the design of long-term adaptive social interaction between robots and children in real-word settings. When it comes to engagement of a child towards the robot, the robot's behaviour becomes central in building and maintaining engagement, aided by the context that the task provides. A richer repertoire of social behaviour of the robot results in a deeper level of user engagement. Thereby, the lack of fixed structure to the interaction provides more scope for flexibility in the robot's behaviour, which has a positive effect on the engagement of the child. This study also found that children who followed sessions with a personalized adaptive robot were more engaged than children who followed sessions with a non-personalized robot.

L2TOR project

L2TOR (pronounced 'el tutor') is a scientific research project, called Second Language Tutoring using Social Robots. The project is funded by the Horizon 2020 programme of the European Commission and aims to design a child-friendly tutor robot that can be used to support teaching preschool children a second language (L2) by interacting with children in their social and referential world (Belpaeme et al., 2015). In this project, multiple universities around Europe perform different studies to contribute to the main research project. In the Netherlands, a long-term study was conducted in a collaboration between Utrecht University and Tilburg University.

L2TOR Large-scale study

A total of 194 children from 9 different primary schools in The Netherlands participated in the experiment and were pseudo-randomly assigned to one of the four conditions: robot with iconic gestures, robot without iconic gestures, tablet only and control condition. In the control condition the children were not exposed to any of the lessons, but played a game with the robot. In the tablet condition, the children followed English lessons on the tablet, without a robot. In both robot conditions the children followed the same English lessons on the tablet in combination with a robot. All conditions had an equal balance of gender and there were fewer children in the control condition. In this thesis, only the robot conditions will be analysed.

The experiment started with a pre-test, to test the level of English for each child. The main part of the experiment consisted of seven different lessons where the children learnt English, depending on their condition with a tablet or a tablet and a Nao Robot. The robot had the role of a tutor, guiding the children through the lesson. A tablet was used for instructions and displayed the learning context. Essentially, the children played educational games on the tablet in all conditions. The child was sitting, on the floor, in front of the tablet. If a robot was present, the robot was placed in an angle of 90 degrees to the right of the child, the set-up is shown in Figure 1. Additionally two cameras were placed in the experimental setting. One in front of the child, to record the child's movements and facial expression. The second camera was placed in a corner, to record an overview of the whole setting.

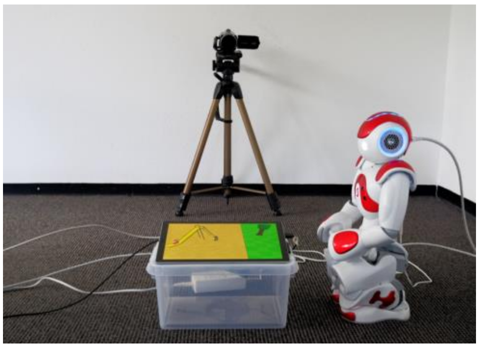

Fig. 1. Set up for the lessons

After the lessons, the two post-tests were administered by a trained researcher in a one-on-one session. The first post-test took place after a maximum of 2 days after the last lesson. The delayed post-test took place between 2 and 5 weeks after the last lesson. Both post-test consisted of the same tasks, but the immediate post-test also included a questionnaire regarding the child's perception of the robot. Therefore the immediate post-test took about 40 minutes, while the delayed post-test took 30 minutes.

The aim of this study was to investigate to what extent social robots can be effective when used in one-on-one tutoring sessions and if a robot would be more effective than a tablet application. The results demonstrated that the educational system used in the lessons was effective, there was learning gain between the pre-test and post-tests, but did not show an added value of the robot compared to using only a tablet application (Vogt, van den Berghe, de Haas, Hoffman, Kanero, et al., 2019).

The effect of gestures on engagement and learning gain

Earlier mentioned work by De Wit et al. (2018), studied the effect of robots with iconic gestures in second language learning. The results showed a larger learning gain, measured after one week, for the condition where robot showed iconic gestures compared to the condition where the robot does not use any gestures. The children also showed more engagement in the lesson when their robot did use iconic gestures. Linnebrink & Pintrich (2010) stated that engagement is needed for learning, therefore the following question has raised: ‘do children who are more engaged, learn more?’. If that is true, the condition where children are more engaged would be more useful to use in education. So, this paper will focus on the engagement of the children per condition and study if this influence their learning gain.

Taking the aforementioned into consideration, this study hypothesizes that:

H1a: Children are more engaged with the robot when their robot tutor uses deictic and iconic gestures, than with one that only uses deictic gestures.

H1b: Children are more engaged with the task when their robot tutor uses deictic and iconic gestures, than with one that only uses deictic gestures.

H2: Children had a larger learning gain when their robot tutor uses deictic and iconic gestures, than with one that only uses deictic gestures.

H3a: Children who are more engaged with the robot, have a larger learning gain.

H3b: Children who are more engaged with the task, have a larger learning gain.

Method

Design

In the L2TOR large-scale study, participants were divided into four different conditions, but only the two robot conditions will be used in this paper, namely: robots with iconic gestures and robot without iconic gestures.

Robot with iconic gestures: participants in this condition interacted with the robot and the tablet, where the robot gave instructions and tasks were performed on the tablet. In this condition, the robot showed an iconic gesture each time it said a target word in English. For example, the word ‘two’ was gestured by showing the child two fingers; the word ‘running’ was shown by moving the arms of the robot back and forth while holding them in an angle of 90 degrees. Figure 2 shows some examples. Furthermore, the robot also showed deictic gestures, which were timed at certain specific points in the lessons. The iconic gestures used in the lessons were tested before in another experiment where one group of adults performed the gestures which then were tested by another group of adults on their clarity (Kanero, Demir-Lira, Koskulu, Oraniç, Franko, Küntay & Göksun, 2018). The lessons in this condition took a bit longer than the other conditions, since it took time for the robot to perform the gesture.

Robot without iconic gestures: In this condition the robot was not making iconic gestures, but only deictic gestures. The other components were the same as the robot with iconic gestures condition.

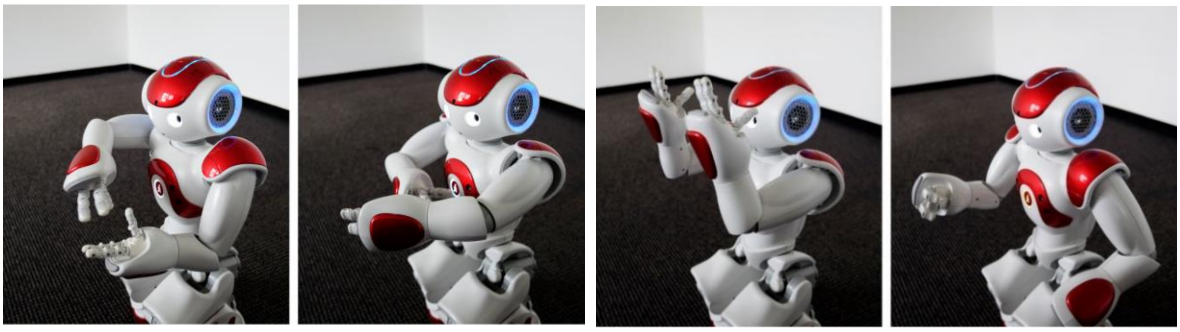

Fig. 2 Iconic gestures.

Participants

A total of 119 children (58 boys, 61 girls) participated in the robot conditions. The average age of the children participating was 5 years and 8 months (*SD* = 5 months) and all of them were native Dutch speakers. There were 62 children in the condition with iconic gestures, and 57 in the condition without iconic gestures. At the time of coding engagement, not all video recordings were available for analysis yet. A total of 49 participants were analysed, 23 participants in the condition with iconic gestures and 26 participants in the condition without iconic gestures.

Materials

The experiment started with a pre-test. In this one-on-one session with a trained researcher, children were tested on their current level of English. The test consisted of an English – Dutch translation task of all the target words of the lessons (34 English words). Children heard the English words, spoken by a female native English speaker, on a computer and were asked what they meant in Dutch. If a child had a higher score than 17 on the pre-test, he or she was excluded from the experiment, since its level of English was too high.

During the seven different lessons, participants learned words from two domains: mathematics (numbers, counting words and basic measurement) and space (spatial relations, prepositions and action verbs). In addition, other English words such as animal names (monkey, giraffe) and other nouns (ball, boy, girl) were used to embed the target words in English phrases. In total, 34 target words were selected. Each lesson had its own theme, which was displayed on the tablet. The theme was connected to the target words. An overview of all lessons, with their theme and target words are shown in table 1.

TABLE 1

Overview of the lessons

Lesson	**Theme**	**Target words**
1	Zoo	One, two, three, add, more, most
2	Bakery	Four, five, take away, fewer, fewest
3	Zoo	Big, small, heavy, light, high, low
4	Fruit shop	On, above, below, next to, falling
5	Forest	In front of, behind, walking, running, jumping, flying
6	Playground	Left, right, catching, throwing, sliding, climbing
7	Picture book	*All target words (recap lesson)*

After the lessons with the robot, two post-tests were conducted. The first post-test immediately after the last lesson (maximally 2 days) and a delayed post-test between 2 and 5 weeks after the last lesson. Both post-tests contained three parts:

- Translation from English to Dutch,
- Translation from Dutch to English,
- Comprehension test of English target words.

For both translations tasks, all 34 target words were tested in the same way as in the pre-test. In the comprehension task, participants saw three pictures or videos and were asked to choose the picture or video corresponding to the target word they heard. In the comprehension task, only half of the target words were tested. Those words were asked 3 times, so the comprehension task involved 54 rounds.

Procedure

All participating children started with a group session where they were introduced to the robot by one of the experimenters. The robot was introduced as 'Robin the robot' and was framed as a peer who would join the children to learn English. The experimenter also explained to the children how to interact with the robot. After the introduction, the pre-test consisting of an English – Dutch translation task took place, to establish the current English level of the children.

For each lesson with the robot, participants were asked during their 'normal' school program to participate in the robot lesson. Children were guided from their classroom to a room dedicated to the experiment. From the moment they walked into the experimental room they were video recorded. The child was instructed to sit in front of the tablet and every lesson the experimenter asked the child if he or she still knew how to interact with the robot. During the lesson, the interaction between the participant and the researcher was kept to a minimum. Each lesson took up about 20 to 30 minutes, depending on the child. After the lesson, participants received a digital star, and were brought back to their classroom by the researcher. Then, the researcher started over again with the next participant

After all seven lessons were completed, the two post-test were administered by a trained researcher. As with the pre-test, the post-tests were administered in one-on-one sessions. Both post-test consisted of the same tasks, namely an English – Dutch translation task, a Dutch – English translation task and a comprehension test.

Analysis

Learning gain

The learning gain of the children was measured using the scores of the English – Dutch translation task of the pre-test, post-test and delayed post-test, since this task was the only one who occurred in all three tests. Learning gain is calculated by subtracting the pre-test core of the immediate post-test score, the same is done for the delayed post-test core. So, in the end there were two learning gain scores. Also the learning outcome is evaluated by comparing the comprehension task results between conditions, this task was only done in both post-test.

Engagement

For this paper, the engagement during 3 lessons was analysed: lesson 2, 4 and 6. Criteria for analysing a participant was that at least 2 of the 3 videos must be available. From every video of the analysed participants, 3 fragments were taken, where every fragment per lesson starts at the same point with a duration of 2 minutes. For this study, only the first fragment is used (all three clips are used in the long-term study, but due to lack of time that was not possible for this paper). This first fragment involves several different activities for the child, for example repeating a word and interacting with the tablet. This first fragment is at the beginning of the lesson, so the children are 'fresh' and not bored or tired yet. In the sixth lesson the child also had to perform a gesture (the robot asked the child if he or she can show 'climbing'). Table 2 shows an overview of the tasks in the first fragment per lesson. The fragment is always two minutes, so depending on correct / incorrect answers and condition, not all tasks will come up in those two minutes, but in most cases they did.

TABLE 2

Overview of tasks in first fragments per lesson

Lesson	Tasks in first fragment
2	o Robot asks the student to touch something on the tablet o Robot asks if the child can say the English word 'five'. o Robot asks what 'less' is on the tablet, the child has to touch the product with less items. o Robot asks if the child can repeat the word 'fewer'.
4	o Robot asks to say the English word 'on'. o Robot asks if the child can say 'on' again. o Robot asks if the child can touch any apple on the tablet.

	o Robot asks if the child can say the English word ‘above’. o Robot asks if the child can say ‘above’ one more time. o Robot asks if the child can touch another apple on the tablet.
6	o Robot asks if the child can say the English word ‘right’. o Robot asks if the child can say the whole English sentence: ‘The girl is to the right of the boy’. o Robot asks if the child can say the English word ‘climbing’. o Robot asks if the child can say the whole sentence: ‘The girl is climbing’. o Robot asks if the child can show the gesture for climbing.

Engagement of the children is analysed on two different levels, namely task engagement and robot engagement. Task engagement involves the engagement of the child with the task he or she had to perform. This could be touching something on the tablet, repeating a word or performing a gesture. Naturally, the child was not active the whole lesson, there were also moments he or she had to listen to an explanation or an example was shown at the tablet. During those moments, task engagement means that the child was paying attention by listening or watching the tablet. The other type of engagement was robot engagement, which is determined by the extent to which the child talks to the robot and looks towards the robot. Also mimicking the gestures of the robot is a sign of robot engagement. If a child only watches the tablet and ignores the robot, he or she will score low on robot engagement. A child can be engaged with the robot, while he or she is completely ignoring the task.

The level of engagement is rated with the ZiKo evaluation scheme (Laevers, 2005). This coding scheme involves 5 levels of engagement, where 1 is extremely low and 5 is extremely high. For the L2TOR project, the scheme was extended to include particular situations in the lessons and differences between the two types of engagement (task or robot), see appendix A for both schemes. For the robot engagement, the score depended on how engaged the child was with the robot. If a child ignored the robot completely, the child scored 1 on robot engagement. If the child was looking and responding to the robot, but was also distracted easily he or she scored in the middle, so 3 points. The child got 5 points when he or she was highly engaged, which means that the child was deeply interacting with the robot, for example: the child reacted on the robot even when the robot was not asking to repeat him. Or, in the iconic gesture condition, the child mimicked the robot's gestures. For the task engagement, the child was rated on how engaged he or she was with the task. So, if a child was not doing the task at all, he or she received 1 point for task engagement. If the child was doing the task, but also got distracted easily (looking away, not responding directly etc.), 3 points were given. If the child was continuously working on the task and he child enjoyed doing the task and was fully concentrated, the child was rated as 'highly engaged' and got 5 points.

All videos were rated with a tool specially created for this research. The tool asked for a participant number, lesson, clip number (fragment) and type of engagement (task or robot). Then, in a new screen, the tool showed the front-view and side-view of the video (if both were available), the coding scheme, a place to type comments and the rating scale (Figure 3). The tool was made in such a way that someone who wanted to rate a video, had to watch the whole video before he or she could rate it. After rating a video, the rating could be saved and a new video could be downloaded.

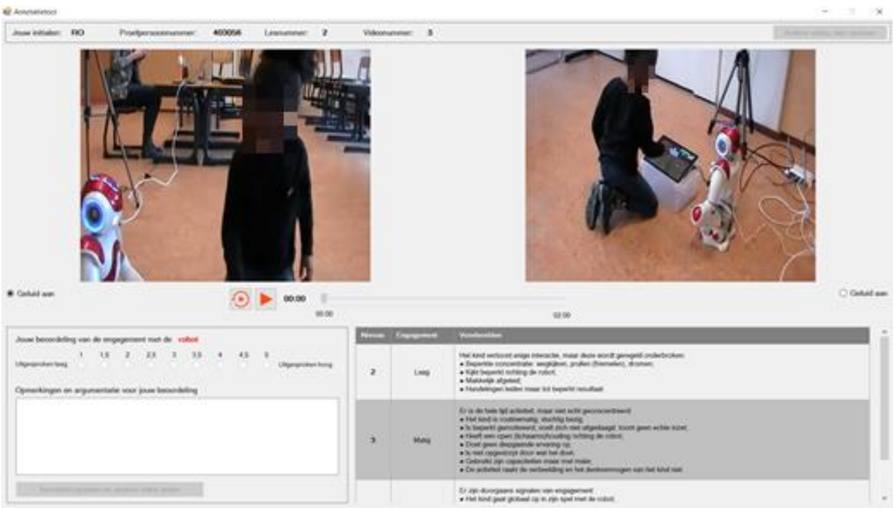

Fig. 3 Coding Tool

A group of 8 coders worked together to analyse the videos for the L2TOR project, but only the videos of the children in the robot conditions were used for this thesis. Before the coders could start coding, they had to participate in a training day. All coders received a document witch coding instructions and the coding scheme (appendix A). During the training the coding scheme was explained and examples were given. They also watched some example videos that were rated and discussed by the whole group. The purpose of this training was to make the coders familiar with the coding scheme, show them how it works and to discuss some videos so everyone had the same idea how specific behaviour was rated.

After the coding training, all coders received a scheme with the videos they had to code. The amount of videos depended on the time they were available. During a weekly skype session, all coders could ask questions to each other about the problems they had during coding in the previous week. By sharing difficulties and giving examples to each other, the coders could help each other out. During this session, the coders were able to discuss how to rate certain videos, this also enabled the coders to check whether they rated certain behaviour in the same way other coders did. Besides the weekly skype meeting, all coders were part of a ‘Slack channel’ where they could ask questions and share videos which they found difficult to rate.

Results

Robot Engagement

To test if children were more engaged with the robot in the condition with iconic gestures than children in the condition without iconic gestures, a mixed ANOVA was performed. Mauchly's test of Sphericity was not significant (p = 0.119), so no correction was needed. The ANOVA showed a significant main effect for lessons $F(2, 48) = 3.85$, $p = 0.028$, $\eta_{partial}^2 = 0.138$. The overall robot engagement for lesson 2 (M = 3.27, SD = 0.62) was higher than for lesson 4 (M = 2.79, SD = 0.62), the overall engagement of lesson 6 (M = 2.90, SD = 0.98) was lower than lesson 2 but higher than lesson 4. There was no significant main effect for condition $F(1, 24) = 0.67$, $p = 0.416$, $\eta_{partial}^2 = 0.028$. The mean score for robot engagement for children in the condition with iconic gestures (M = 3.1, SD = 0.64) was higher than the mean score for robot engagement for children in the condition without iconic gestures (M = 2.9, SD = 0.8). All mean scores per condition per lesson are displayed in Figure 4. Finally, there was no significant interaction effect, $F(2, 48) = 0.391$, $p = 0.679$, $\eta_{partial}^2 = 0.016$, as can be seen in Figure 5.

Post Hoc test using the Bonferroni correction revealed that there was a significant difference between lesson 2 and lesson 4 (p = .009), but there was no significant difference between lesson 2 and lesson 6 (p = 0.224), nor between lesson 4 and lesson 6 (p = 1.0).

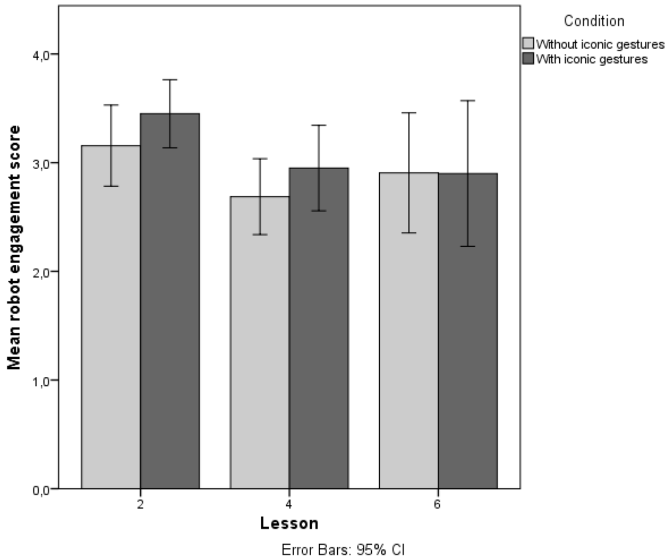

Fig. 4, Mean robot engagement score per lesson per condition

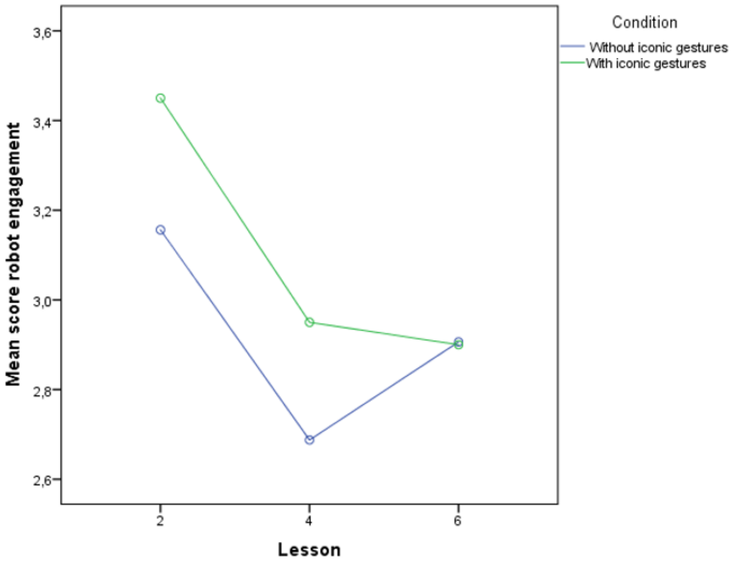

Fig. 5, Robot engagement per lesson per condition

Task engagement

To test if children were more engaged with the task in the condition with iconic gestures than children in the condition without iconic gestures, a mixed ANOVA was performed. Mauchly's test of Sphericity was significant (p = 0.039), therefore the Greenhouse-Geisser correction is used. The ANOVA showed a significant main effect for lessons $F(1.61, 38.52) = 6.64$, p = 0.006, $\eta_{partial}^2 = 0.217$. The overall task engagement for lesson 2 (M = 3.5, SD = 0.53) was higher than for lesson 4 (M = 3.0, SD = 0.66), and for lesson 6 (M = 3.0, SD = 0.81). Lesson 4 and lesson 6 had the same mean task engagement score over all three lessons. There was no significant main effect for condition $F(1, 24) = 1.71$, $p = 0.203$, $\eta_{partial}^2 = 0.066$. The mean score for task engagement for children in the with iconic gestures condition (M = 3.02, SD = 0.75) was lower than the mean score for task engagement for children in the without iconic gestures condition (M = 3.26, SD = 0.61). All mean scores per condition per lesson are displayed in Figure 6. Finally, there was no significant interaction effect, $F(1.61, 38.52) = 0.783$, $p = 0.438$, $\eta_{partial}^2 = 0.032$, as can be seen in Figure 7.

Post Hoc test using the Bonferroni correction revealed that there was a significant difference between lesson 2 and lesson 4 (p = .001), there was also significant difference between lesson 2 and lesson 6 (p = 0.013), but not between lessen 4 and lesson 6 (p = 1.0).

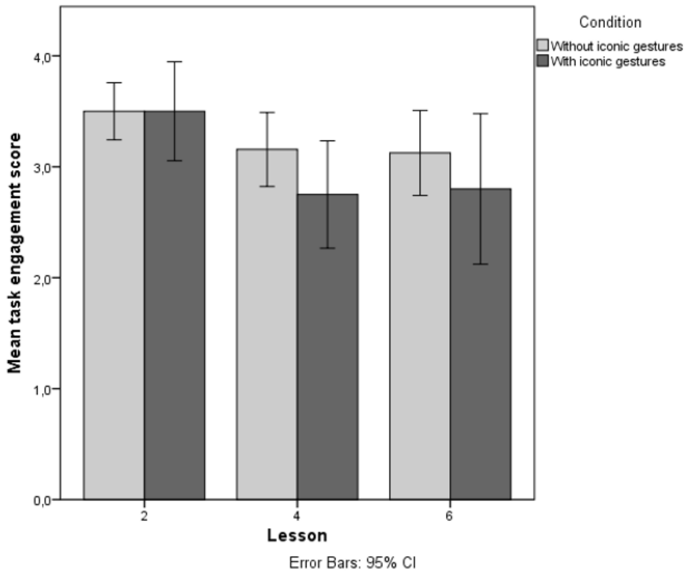

Fig. 6, Mean task engagement score per lesson per condition.

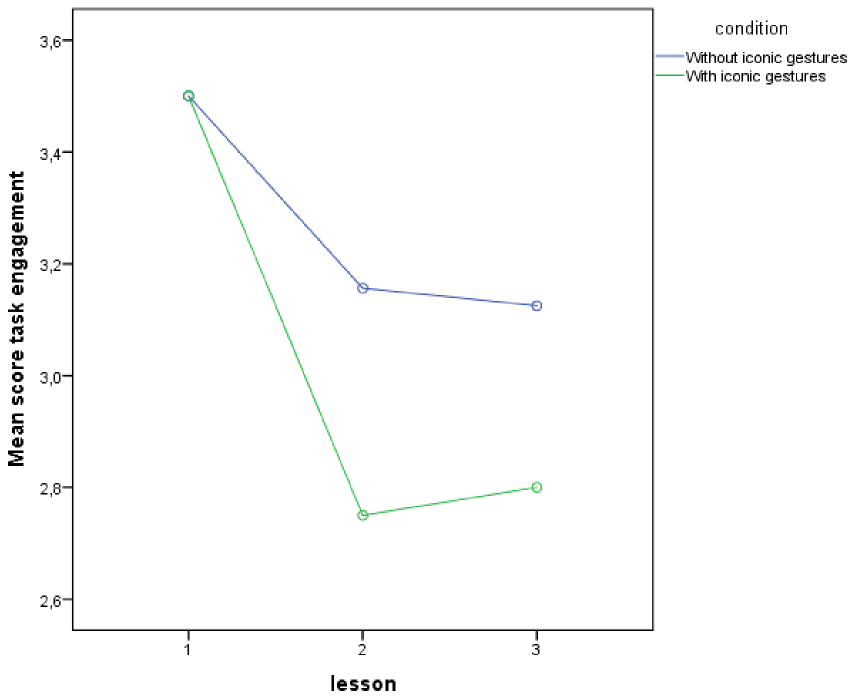

Fig. 7, Task engagement per lesson per condition.

Learning gain

To test if children had a larger learning gain when their robot tutor uses deictic and iconic gestures, than when it only uses deictic gestures, a mixed ANOVA is performed. Mauchly's test of Sphericity was significant ($p < .001$), therefore the Greenhouse-Geisser correction is used. The ANOVA showed a significant main effect for test moments (pre-test, immediate post-test and delayed post-test) $F(1.48, 69.45) = 65.62$, $p < .001$, $\eta_{partial}^2 = 0.583$. The overall score for the pre-test ($M = 3.47$, $SD = 2.68$) was lower than for the immediate post-test ($M = 7.57$, $SD = 5.16$). Children scored the highest at the delayed post-test ($M = 7.90$, $SD = 5.03$). There was no significant main effect for condition $F(1, 47) = 0.47$, $p = 0.828$, $\eta_{partial}^2 = 0.001$. The mean score for all test results for children in the with iconic gestures condition ($M = 6.45$, $SD = 4.60$) was higher than the mean score for all test results for children in the without iconic gestures condition ($M = 6.19$, $SD = 4.06$). All mean scores per condition per test are displayed in Figure 8. Finally, there was no significant interaction effect, $F(1.48, 69.45) = 0.778$, $p = 0.428$, $\eta_{partial}^2 = 0.016$, as can be seen in Figure 9.

Post Hoc test using the Bonferroni correction revealed that there was a significant difference between the pre-test and immediate post-test ($p < .001$), and also between the pre-test and delayed post-test ($p < .001$), but there was no significant difference between both post-test ($p = .723$)

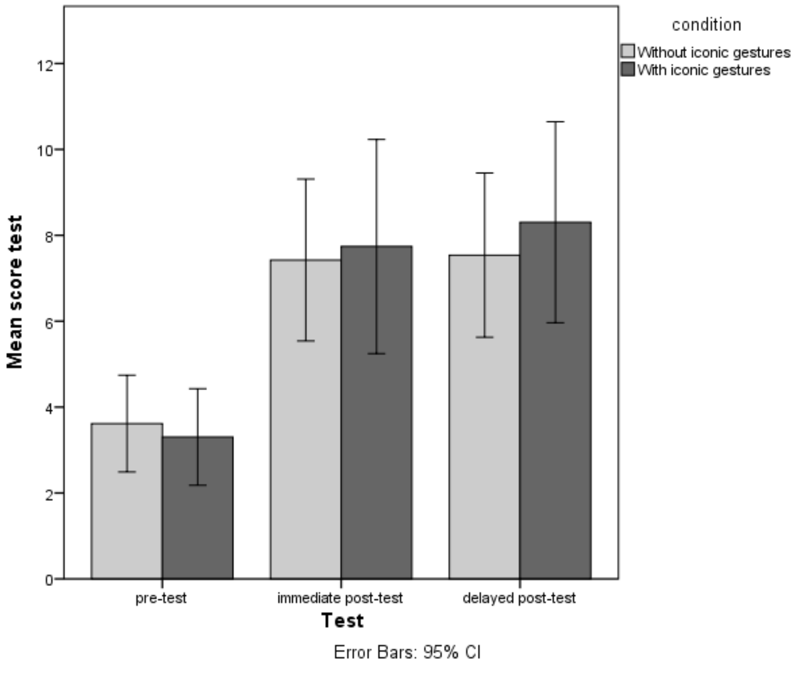

Fig. 8, Mean score per test, per condition.

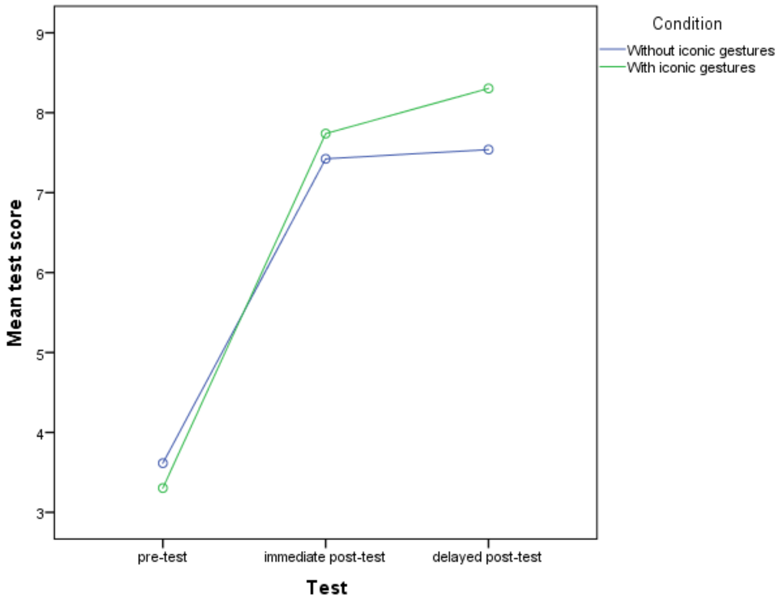

Fig. 9, Mean score per test, per condition.

Learning outcome

The post-tests consisted of a comprehension, where 18 target words were asked 3 times. The results of the immediate post-test were not normally distributed and were therefore analyzed with a Mann-Whitney U test, to indicate if there was any difference between both conditions. The results showed that the test results for children in the without iconic gestures condition (*Mdn* = 29) and children in the with iconic gesture condition (*Mdn* = 30) did not differ significantly ($U = 269$, $z = -.602$, $p = .547$).

Correlation between engagement and learning gain

Multiple correlation tests are performed to see if there was any correlation between engagement and learning gain, measured from the English – Dutch translation task. First, the correlation between both engagements (robot and task) and learning gain is measured from the immediate post-test scores (immediate post-test score minus pre-test score). Thereafter, the same test is done for learning gain measured from the delayed post-test scores (delayed post-test score minus pre-test score).

Correlation robot engagement and immediate post-test

The average robot engagement score was 2.99 (*SD* = 0.56), and the average learning gain was 4.10 (*SD* = 3.62). The variables were not normally distributed (learning gain z-$\text{score}_{\text{skewness}}$ = 2.88), therefore a Spearman's correlation test is used to test the relation between robot engagement and learning gain, based on the immediate post-test results. The analysis showed a significant correlation between robot engagement and learning gain $\rho = 0.307$, $p = .032$.

Correlation task engagement and immediate post-test

The average task engagement score was 3.11 (*SD* = 0.53), and the average learning gain was 4.10 (*SD* = 3.62). The variables were not normally distributed (learning gain $\text{z-score}_{\text{skewness}} = 2.88$), therefore a Spearman's correlation test is used to test the relation between task engagement and learning gain, based on the immediate post-test results. The analysis showed a significant correlation between task engagement and learning gain $\rho = 0.299$, $p = .037$.

Correlation robot engagement and delayed post-test

The average robot engagement score was 2.99 (*SD* = 0.56), and the average learning gain was 4.43 (*SD* = 3.22). The variables were not normally distributed (learning gain $\text{z-score}_{\text{skewness}} = 2.26$), therefore a Spearman's correlation test is used to test the relation between robot engagement and learning gain, based on the delayed post-test results. The analysis showed a significant correlation between robot engagement and learning gain $\rho = 0.325$, $p = .023$.

Correlation task engagement and delayed post-test

The average task engagement score was 3.11 (*SD* = 0.53), and the average learning gain was 4.43 (*SD* = 3.22). The variables were not normally distributed (learning gain $\text{z-score}_{\text{skewness}} = 2.26$), therefore a Spearman's correlation test is used to test the relation between task engagement and learning gain, based on the delayed post-test results. The analysis showed a significant correlation between task engagement and learning gain $\rho = 0.285$, $p = .047$.

Conclusion and Discussion

Discussion

In this paper, the effect of robots' gestures on engagement and learning gain is studied. Additionally, the relationship between engagement and learning gain is evaluated. To investigate this, the following research question was formulated: " Does the engagement of the students in the lessons influence their learning gain and do those results differ between children who are tutored by a robot performing only deictic gestures and a robot performing deictic and iconic gestures?". To answer this research question, the engagement of the children with the robot as well as with the task is analysed. Additionally, the learning gain of the children is tested by a pre-test and two post-tests to see if they made any improvements.

Robot engagement

Engagement is tested in two different ways: robot engagement and task engagement. Robot engagement is determined by the extent to which the child looks and talks to the robot and also mimicking the robot is a sign of robot engagement. The results show that children in the condition with iconic gestures ($M = 3.1$) were over all lessons more engaged with the robot than children in the without iconic gestures condition ($M = 2.9$). According to Beattie (2010) gestures attract visual attention, so therefore it is logical that children in the iconic gestures condition were more attracted to the robot than children in the condition without iconic gestures, since the robot moved more when performing iconic gestures. However, the difference between both conditions is 0.2 points, so the mean robot engagement of both conditions does not differ that much. Statistical analysis show that there is no significant main effect between both conditions on robot engagement and therefore H1a is rejected. It can be concluded that the presence of iconic gestures did not influence the robot engagement of the children.

On the other hand, Figure 4 shows that in lesson 2 and lesson 4 children in the with iconic gestures condition were more engaged with the robot than children in the without iconic gestures condition. It could be that the iconic gestures designed on the robot in those lessons provoke more attention than the deictic gestures did. If this is the case, the iconic gestures and deictic gestures in lesson 6 created the same amount of attention. Furthermore, the target words consisted of some abstract words, which were harder to portray. Therefore they were maybe not recognized by the child and seem to be a deictic gesture which did not catch the attention of the child. However, this is not tested in this study and therefore further research could be done on this subject.

Taking a look at the robot engagement per lesson, per condition, the robot engagement for the condition with iconic gestures became lower every lesson. In the condition without iconic gestures, the engagement was the highest in the second lesson, then dropped in the fourth lesson, but increased a bit in the sixth lesson. For both conditions it is logical that in the second lesson (the first one analysed) the robot engagement is highest, since the second session is still a new experience for the children and the robot is a new and very interesting object. It could be that children in the first lessons are trying to discover what the robot can and cannot do. Therefore the robot engagement drops over time, since the children know what the robot does and it became less interesting and exciting for them.

Task engagement

The other kind of engagement which was tested was task engagement, which involves playing the games on the tablet, repeating words and an active attitude. Results show that children in the condition with iconic gestures overall scored lower on task engagement (M = 3.02) than children in the condition without iconic gestures (M = 3.26). Those results seem logical compared to the robot engagement outcome, where children in the with iconic gesture condition were more engaged with the robot. When the child is more focused on the robot, it therefore has less attention for the task. So, when children are not so attracted to the robot (when there are no iconic gestures) their focus will be more on the task. However, also here no significant main effect was found for condition and therefore H1b is not supported by the data.

Figure 6 shows that task engagement in lesson 2 was the same for both conditions. As mentioned before, during the first lesson everything was new and probably interesting for the child. So regardless of the condition, the children had the same level of interest in the game and may have had the same difficulties to understand how the tablet worked. The task engagement of lesson 4 and lesson 6 were for both conditions for both lessons more or less the same. It is interesting to see that the task engagement did not increase nor decrease between those lessons. It could be that the tasks of both lessons were designed similarly and therefore children paid the same amount of attention to the task.

Looking at the task engagement per lesson per condition, the task engagement for children in the condition without iconic gestures dropped between lesson 2 and lesson 4, but the difference between the lessons is not very big, and stayed the same between lesson 4 and lesson 6. Also the task engagement for children in the iconic gestures condition was the highest in lesson 2, but decreased more in lesson 4 and also stayed the same between lesson 4 and lesson 6. This drop may be a result of less interest in the lessons during the experiment, where the children were more enthusiastic and interested in the beginning and after a few lessons they were less interested.

Learning gain

Results show that there was an overall learning gain between the pre-test and immediate post-test and between pre-test and delayed post-test. So therefore it could be said that the lessons with the robot, no matter which condition a child was in, had an effect on their knowledge of English words. There was no main effect for condition found, so the learning gain did not depend on the robot's use of gestures and therefore H2 is not supported by the data.

Nevertheless, the results show (fig. 8) that the pre-test score of children in the with iconic gestures condition were lower than those of children in the without iconic gesture condition, but increased more in the immediate post-test and also in the delayed post-test. This means that children in the with iconic gesture condition learned on average more words than children in the without iconic gestures condition. Although the without iconic gestures conditions increased less, they also learned more English words during the lessons. All lessons in both robot conditions were mediated by the tablet, which displayed the learning context, while the robot took the role of supporting tutor. Since there is no big difference between the robot conditions, it is likely that the education program on the tablet causes the learning gain and the robot was an additional tool which did not influence the learning gain. This does not mean the robot was useless. It may have effected the confidence or comfort of the child during the lessons. Those results are in line with the long-term study of L2TOR, where not only two robot conditions were tested, but also a control condition (no lessons) and a tablet only condition. The results of this study show that there was no difference in learning gain between robot conditions and the tablet condition (Vogt et al., 2019).

The results of this study are in conflict with previous research showing a positive effect of iconic gestures in second language learning. Tellier (2015) proved that children in the age of 5 learned more words when their tutor used iconic gestures while explaining the words they had to learn, than children with a tutor who did not show iconic gestures. In contrast to this paper, the study by Tellier (2015) was done with human tutors and not with robots. Although, Van Dijk (2018) did a similar study in the use of gestures in a human-robot interaction. The results of that study was that gestures helped in the learning process, specifically remembering the words on long term. In both of those studies, the tutor (human or robot) was the only information source for the students, which could be the reason why gestures matters for the learning gain of the children. In this paper, the main source of information was the tablet, so that may effected the learning gain of the children more than the robot. The experiment by De Wit et al. (2018) is comparable with experiment in this thesis. De Wit et al. (2018) also used a tablet as main information source, while the robot operated as tutor. The results showed a larger learning gain, measured after one week, for the condition with iconic gestures. One of the differences between the study by De Wit et al. (2018) and the L2TOR experiment is that the long-term study involved words that were harder to portray, for example ‘more’, ‘fewer’, ‘on’ and ‘next to’. Those words are not easy to recognize and are easy to get confused with, since those words are more abstract. This may be one of the reasons why there are no results found for iconic gestures in this thesis. Another reason why there was no effect found for gestures could be due to the clarity of the gestures. The physical limitations of the robot causes sometimes clumsy movements. For example, the Nao robot has only three fingers and could stretch them only simultaneously and not separate from each other, which makes it harder to show numbers. Another reason why no effect was found for iconic gestures is that they were shown every time the robot said a target word. This could have been an overkill of gestures and it also slowed the robot down in its speech flow, which may have distracted the child from the learning task. Furthermore, in this experiment the children observed the gestures. Some children also mimicked the gestures of the robot, but this was not asked from them. De Nooijer et al. (2012) found a positive effect of imitating gestures on the learning process. Therefore, it is questionable if different results were found if children had to imitate the gestures of the robot in this experiment.

Relation between engagement and learning gain

The relation between engagement and learning gain is analysed with a correlation test. Results show that there is a relationship between engagement, with both robot and task, and learning gain. Higher levels of engagement, no matter if it was task engagement or robot engagement, resulted in greater learning gain. Webber (2004) stated that engagement leads to more attention which leads to receiving and processing more information and eventually improves learning.

This is in line with the results of this paper and means that engagement is an important aspect of the learning process. This is a logical result, disengagement normally results in distraction which means that the student is not paying attention to the task anymore. Therefore he or she is not receiving all the information, which makes it harder to process the information he or she gets.

Interestingly, robot engagement as well as task engagement influence learning gain positively. On one hand, it is logical that both engagements results in learning gain, since in both situation students were engaged and therefore paying attention. In addition, when a student was highly engaged with the robot, he or she could also be highly engaged with the task. So, it could be that students who scored high on robot engagement naturally scored high on task engagement as well and students who scored low on robot engagement also scored low on task engagement.

In summary, H3 is supported by the data. Children who are more engaged with the robot and/or with the task had a larger learning gain. Therefore it could be said that engagement is important in the learning process and educational systems should try to keep the student engaged.

Conclusion

It was assumed that iconic gestures had a positive effect on the robot and task engagement of the child, since this was also the case for a similar, but smaller, study by De Wit et al. (2018). However, results show that this was not the case for this long-term study. Learning gain was expected and also found in this research, but the use of iconic gestures did not influence the learning gain. Both conditions, with and without iconic gestures, had the same amount of learning gain. Therefore it is questionable if the use of iconic gestures is needed for teacher robots, given that there is no different effect in learning gain. Additionally, there was a relationship between engagement and learning gain. This means that an educational system should focus on maintaining a high level of engagement of the student to stimulate the learning process.

Limitations and future research

Engagement of the children was divided into robot and task engagement and was coded by a team of trained coders with the ZiKo evaluation scheme. The original scheme was used as baseline and extended to include particular situations in the lessons and differences between the two types of engagement (task and robot), but some situations were not described in the scheme and therefore harder to analyse. Evidence was found for the relationship between engagement and learning gain, but it did not matter which engagement a child had (robot or task). It could be that the evaluation scheme used by the coders for both engagements was too similar and therefore the difference between robot engagement and task engagement was hard to distinguish. Research has shown that attention leads to learning (Webber, 2004), so it may not be needed to distinguish two kinds of engagement. If in the future a similar experiment will be designed, the decision could be made to just evaluate overall engagement or attention. The other option is to distinguish robot engagement and task engagement more from each other.

Due to lack of time, most videos were not double coded yet, so there was no interrater agreement. This could result in less reliable ratings. The L2TOR project is still running and all available videos will be coded. The results of the L2TOR long-term study will show if there is indeed a significant higher robot engagement for children in the condition with iconic gestures than for children in the condition without iconic gestures, and the other way around for task engagement.

No effect was found for the use of iconic gestures in combination with deictic gestures compared to only deictic gestures. Tellier (2007) states that the use of iconic gestures, in this case from a human teacher, can have an effect on memorisation. But, a prerequisite to make it more effective is that the children should reproduce the gestures while repeating the words. Thus, they will be more active in their repetition and reinforce its trace in memory. In the experiment, iconic gestures were used by the robot, but only in lesson 6 the child was asked to reproduce the gesture. So watching the iconic gesture may not have affected the child, but performing it may have. Therefore it could be interesting to observe cases where the child is more active in the session with the robot and reproduces the gestures of the robot. Since an effect was found with a human teacher and this teaching style, it is interesting to see of the same effects appear with a robot teacher.

References

Alibali, M. W., & Nathan, M. J. (2007). Teachers' gestures as a means of scaffolding students' understanding: Evidence from an early algebra lesson. *Video research in the learning sciences*, 349-365.

Beattie, G., Webster, K., & Ross, J. (2010). The fixation and processing of the iconic gestures that accompany talk. *Journal of Language and Social Psychology*, *29*(2), 194-213.

Belpaeme, T., Baxter, P., Read, R., Wood, R., Cuayáhuitl, H., Kiefer, B. & Looije, R. (2013). Multimodal child-robot interaction: Building social bonds. *Journal of Human-Robot Interaction*, *1*(2), 33-53.

Belpaeme, T., Kennedy, J., Baxter, P., Vogt, P., Krahmer, E. E., Kopp, S. & Pandey, A. K. (2015). L2TOR-second language tutoring using social robots. In *Proceedings of the ICSR 2015 WONDER Workshop*.

Bergmann, K., & Macedonia, M. (2013, August). A virtual agent as vocabulary trainer: iconic gestures help to improve learners' memory performance. In *International workshop on intelligent virtual agents* (pp. 139-148). Springer, Berlin, Heidelberg.

Brühwiler, C., & Blatchford, P. (2011). Effects of class size and adaptive teaching competency on classroom processes and academic outcome. *Learning and Instruction*, *21*(1), 95-108.

Charisi, V., Davison, D., Reidsma, D., & 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)*.

Chiat, S., "Non-word repetition," in *Methods for assessing multilingual children: Disentangling bilingualism from language impairment*, . N. M. E. S. Armon-Lotem, J. de Jong, Ed. Bristol: Multilingualism Matters, 2015, pp. 227–250.

Clements, D. (1998). Young Children and Technology.

Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition*, *106*(2), 1047-1058.

Corts, D. P., & Pollio, H. R. (1999). Spontaneous production of figurative language and gesture in college lectures. *Metaphor and Symbol*, *14*(2), 81-100.

Van Dijk, E. T., Torta, E., & Cuijpers, R. H. (2013). Effects of eye contact and iconic gestures on message retention in human-robot interaction. *International Journal of Social Robotics*, *5*(4), 491-501.

Dunn, L. M. & Schlichting, L. (2005). *Peabody picture vocabulary test-III-NL*. Amsterdam: Pearson, 2005.

Dörnyei, Z. (1990). Conceptualizing motivation in foreign-language learning. *Language learning*, *40*(1), 45-78.

Dörnyei, Z. (2003). Attitudes, orientations, and motivations in language learning: Advances in theory, research, and applications. *Language learning*, *53*(S1), 3-32.

Eisenstein, J., & Davis, R. (2006, July). Natural gesture in descriptive monologues. In *ACM SIGGRAPH 2006 Courses* (p. 26). ACM.

Gu, Y., & Johnson, R. K. (1996). Vocabulary learning strategies and language learning outcomes. *Language learning*, *46*(4), 643-679.

Han, J. H., Jo, M. H., Jones, V., & Jo, J. H. (2008). Comparative study on the educational use of home robots for children. *Journal of Information Processing Systems*, *4*(4), 159-168.

Ivanov, S. (2016) Will robots substitute teachers? Paper presented at the 12th International Conference “Modern science, business and education”, 27-29 June 2016, Varna University of Management, Bulgaria. Yearbook of Varna University of Management, Vol. 9, pp. 42-47.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive psychology*, *21*(1), 60-99.

Kanero,J., Demir-Lira, O. E., Koskulu, S., Oranc, G., Franko, I., Kuntay, A.C., & Oksun, T. G. *“How do robot gestures help second language ¨ learning?”* in Earli SIG 5 Abstract book, 2018.

Kelly, S. D., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and cognitive processes*, *24*(2), 313-334.

Kim, Y. K., Curby, T. W., & Winsler, A. (2014). Child, family, and school characteristics related to English proficiency development among low-income, dual language learners. *Developmental psychology*, *50*(12), 2600.

Laevers, F., Daems, M., De Bruyckere, G., Declercq, B., Silkens, K., Snoeck, G., ... & Van Kessel, M. (2005). Zelfevaluatie-instrument voor welbevinden en betrokkenheid van kinderen in de opvang (ZiKo).

Leite, I., Castellano, G., Pereira, A., Martinho, C., & Paiva, A. (2014). *Empathic robots for long-term interaction*. International Journal of Social Robotics, *6*(3), 329-341.

Linnenbrink, E. A., & Pintrich, P. R. (2003). The role of self-efficacy beliefs instudent engagement and learning intheclassroom. *Reading &Writing Quarterly*, *19*(2), 119-137.

Macedonia, M., Müller, K., & Friederici, A. D. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human brain mapping*, *32*(6), 982-998.

Mastin, J. D., & Vogt, P. (2016). Infant engagement and early vocabulary development: a naturalistic observation study of Mozambican infants from 1; 1 to 2; 1. *Journal of child language*, *43*(2), 235-264.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago press.

Mubin, O., Stevens, C. J., Shahid, S., Al Mahmud, A., & Dong, J. J. (2013). A review of the applicability of robots in education. *Journal of Technology in Education and Learning*, *1*(209-0015), 13.

Mulder, H., Hoofs, H., Verhagen, J., van der Veen, I., & Leseman, P.P.M.*"Psychometric properties and convergent and predictive validity of an executive function test battery for two-year-olds,"* Frontiers in Psychology, vol. 5, p. 733, 2014.

Nakano, Y. I., & Ishii, R. (2010, February). Estimating user's engagement from eye-gaze behaviors in human-agent conversations. In *Proceedings of the 15th international conference on Intelligent user interfaces* (pp. 139-148). ACM.

De Nooijer, J. A., Van Gog, T., Paas, F., & Zwaan, R. A. (2013). Effects of imitating gestures during encoding or during retrieval of novel verbs on children's test performance. *Acta Psychologica*, *144*(1), 173-179.

Roth, W. M. (2001). Gestures: Their role in teaching and learning. *Review of educational research*, *71*(3), 365-392.

Schumann, J. H. (1975). Affective factors and the problem of age in second language acquisition. *Language learning*, *25*(2), 209-235.

Serholt, S., Basedow, C. A., Barendregt, W., & Obaid, M. (2014, November). Comparing a humanoid tutor to a human tutor delivering an instructional task to children. In *Humanoid Robots (Humanoids), 2014 14th IEEE-RAS International Conference on*(pp. 1134-1141). IEEE.

Sidner, C. L., Lee, C., Kidd, C. D., Lesh, N., & Rich, C. (2005). Explorations in engagement for humans and robots. *Artificial Intelligence*, *166*(1-2), 140-164.

Stevanoni, E., & Salmon, K. (2005). Giving memory a hand: Instructing children to gesture enhances their event recall. Journal of Nonverbal Behavior, 29, 217–233.

Tai, Y. (2014). The application of body language in English teaching. *Journal of Language Teaching and Research*, *5*(5), 1205.

Tellier, M. (2005). How do teacher's gestures help young children in second language acquisition?. In *International Society of Gesture Studies, ISGS* (pp. en-ligne).

Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture*, *8*(2), 219-235.

Vogt, P., van den Berghe, R., de Haas, M., Hoffmann, L., Kanero, J., Mamus, E., Montanier, J.-M., Oranç, C., Oudgenoeg-Paz, O., Papadopoulos, F., Schodde, T., Verhagen, J., Wallbridge, C.D., Willemsen, B., de Wit, J., Belpaeme, T., Bergmann, K., Göksun, T., Kopp, S., Krahmer, E., Küntay, A.C., Leseman & Pandey, A.K., "Second language tutoring using social robots: A large-scale study," to appear in the *2019 ACM/IEEE International Conference on Human-Robot Interaction*, 2019.

Webber, T. (2004). Orientations to learning in mid-career management students. *Studies in Higher Education*, *29*(2), 259-277.

Westlund, J. K., Dickens, L., Jeong, S., Harris, P., DeSteno, D., & Breazeal, C. (2015). A comparison of children learning new words from robots, tablets, & people. In *Conference Proceedings New Friends 2015*.

de Wit, J., Schodde, T., Willemsen, B., Bergmann, K., de Haas, M., Kopp, S. & Vogt, P. (2018). The Effect of a Robot's Gestures and Adaptive Tutoring on Children's Acquisition of Second Language Vocabularies.

de Wit, J., Schodde, T., Willemsen, B., Bergmann, K., de Haas, M., Kopp, S., Krahmer, E. & Vogt, P., “The effect of a robot’s gestures and adaptive tutoring on children’s acquisition of second language vocabularies,” in Proceedings of *the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2018, pp. 50–58.

Yang, G. Z., Bellingham, J., Dupont, P. E., Fischer, P., Floridi, L., Full, R., ... & Nelson, B. J. (2018). The grand challenges of Science Robotics. *Science Robotics*, *3*(14), eaar7650.

Appendix A

Task Engagement

De schaal voor child-task engagement		
Niveau	Engagement	Voorbeelden
1	Uitgesproken laag	Het kind vertoont nagenoeg geen activiteit: „ Geen concentratie: staren, wegdromen; „ Een afwezige, passieve houding; „ Geen gerichte activiteit, doelloze handelingen, niets teweegbrengen; „ Alleen bezig met de experiment leider en niet met de taak; „ Geen tekenen van exploratie en interesse; „ Niets in zich opnemen, geen mentale activiteit.
2	Laag	Het kind vertoont enige activiteit, maar deze wordt geregeld onderbroken: „ Beperkte concentratie: wegkijken, prullen (friemelen), dromen; „ Makkelijk afgeleid; „ Taken worden in beperke mate uitgevoerd.
3	Matig	Er is de hele tijd activiteit, maar niet echt geconcentreerd. „ Het kind is routinematig, vluchtig bezig; „ Is beperkt gemotiveerd, voelt zich niet uitgedaagd, toont geen echte inzet; „ Doet geen diepgaande ervaring op; „ Is niet opgeslorpt door wat het doet; „ Gebruikt zijn capaciteiten maar met mate; „ De activiteit raakt de verbeelding en het denkvermogen van het kind niet. „ De meeste taken worden uitgevoerd.

4	Hoog	Er zijn doorgaans signalen van engagement: „ Het kind gaat globaal op in zijn spel; „ Er is doorgaans concentratie, maar soms verslapt de aandacht; „ Het kind voelt zich uitgedaagd, er is een zekere gedrevenheid; „ Gebruikt zijn capaciteiten; „ Spreekt de verbeelding en het denkvermogen aan.
5	Uitgesproken hoog	Het kind is gedurende de hele tijd ononderbroken bezig en gaat sterk op in zijn activiteit: „ Is ononderbroken geconcentreerd, opgeslorpt door de activiteit, vergeet de tijd; „ Is heel gemotiveerd, voelt zich sterk aangesproken; „ Is niet af te leiden; „ Kijkt aandachtig naar de taak, heeft aandacht voor details; „ Spreekt voortdurend al zijn capaciteiten en mogelijkheden aan; „ Er is een sterke mentale activiteit: de verbeelding en het denkvermogen draaien op volle toeren; „ Doet diepgaande nieuwe ervaringen op; „ Geniet van zo gedreven bezig te zijn.

Robot Engagement

De schaal voor child-robot engagement		
Niveau	Engagement	Voorbeelden
1	Uitgesproken laag	Het kind vertoont nagenoeg geen interactie: „ Geen concentratie: staren, wegdromen; „ Negeert de robot volledig; „ Heeft een gesloten (lichaams)houding richting de robot; „ Een afwezige, passieve houding; „ Geen gerichte activiteit, doelloze handelingen, niets teweegbrengen; „ Geen tekenen van exploratie en interesse; „ Niets in zich opnemen, geen mentale activiteit.
2	Laag	Het kind vertoont enige interactie, maar deze wordt geregeld onderbroken: „ Beperkte concentratie: wegkijken, prullen (friemelen), dromen; „ Kijkt beperkt richting de robot; „ Makkelijk afgeleid; „ Handelingen leiden maar tot beperkt resultaat.
3	Matig	Er is de hele tijd activiteit, maar niet echt geconcentreerd. „ Het kind is routinematig, vluchtig bezig; „ Is beperkt gemotiveerd, voelt zich niet uitgedaagd, toont geen echte inzet; „ heeft een open (lichaams)houding richting de robot; „ Doet geen diepgaande ervaring op; „ Is niet opgeslorpt door wat het doet; „ Gebruikt zijn capaciteiten maar met mate; „ De activiteit raakt de verbeelding en het denkvermogen van het kind niet. „ Doelloos aanraken van de robot

4	Hoog	Er zijn doorgaans signalen van engagement: „ Het kind gaat globaal op in zijn spel met de robot; „ Er is doorgaans sprake van joint attention; „ Er is doorgaans concentratie, maar soms verslapt de aandacht; „ Het kind voelt zich uitgedaagd, er is een zekere gedrevenheid; „ Gebruikt zijn capaciteiten; „ Spreekt de verbeelding en het denkvermogen aan.
5	Uitgesproken hoog	Het kind is gedurende de hele tijd ononderbroken bezig en gaat sterk op in zijn activiteit met de robot: „ Is ononderbroken geconcentreerd, vergeet de tijd; „ Is heel gemotiveerd, voelt zich sterk aangesproken; „ Is niet af te leiden; „ Kijkt aandachtig naar robot, heeft aandacht voor details; „ Praat tegen de robot; „ Gebaren na doen (alleen in de iconische gebaren conditie); „ Er is sprake van joint attention; „ Er is een sterke mentale activiteit: de verbeelding en het denkvermogen draaien op volle toeren; „ Doet diepgaande nieuwe ervaringen op; „ Geniet van zo gedreven bezig te zijn.