



Second Language Tutoring using Social Robots



Project No. 688014

L2TOR

Second Language Tutoring using Social Robots

Grant Agreement Type: Collaborative Project
Grant Agreement Number: 688014

D7.1 Evaluation report number domain

Due Date: **31/03/2018**
Submission Date: **DD/MM/YYYY**

Start date of project: **01/01/2016**

Duration: **36 months**

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

Responsible Person: **Ora Oudgenoeg-Paz**

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)	

Contents

Executive Summary	3
Principal Contributors	4
Revision History	5
1 Introduction	6
2 Pilot Study Number Domain	7
3 Pilot Study Space Domain	9
4 Design of Large-Scale Evaluation Study	10
5 Study with Turkish-Dutch children	12
Appendix I: Rintjema et al. (2018)	14
Appendix II: Protocol of the Large Evaluation Study	17
Appendix III: Pre-registration Aspredicted	90
Appendix IV: Master thesis H. Leeuwestein	93

Executive Summary

This deliverable reports about the main projects within WP7 as a preparation towards the large-scale evaluation study for which data collection is still ongoing. The start of the large-scale study was delayed based on the revised objectives set for the L2TOR project, as approved by the project officer and the reviewers. First, we report about two pilot studies conducted to test the systems for the number and space domain in schools. Second, we report about the design and preparation for the large-scale evaluation study. Finally, we report about a study testing the use of the L2TOR system with Turkish-Dutch children in the Netherlands learning Dutch as a second language.

Principal Contributors

<i>Utrecht</i>	Rianne van den Berghe, Hanneke Leeuwestein, Paul Leseman, Ora Oudgenoeg-Paz, Josje Verhagen
<i>Tilburg</i>	Mirjam de Haas, Emiel Krahmer, Emmy Rintjema, Paul Vogt, Bram Willemsen, Jan de Wit



Revision History

Version 1.0 (OO 17-05-2018)

1 Introduction

Given the revised objectives of the study as approved by the project officer and the reviewers, the large-scale evaluation study now includes only children in the Netherlands. Moreover, the timeline has been adjusted and the study is being conducted from February 2018 till June 2018. Therefore, the current deliverable will not include a final evaluation of the number domain as the title suggests, but rather a report of our pilot studies with the number and space domain and the design of the large-scale study running at the moment. The study protocol is also included in an appendix. This protocol can be used to guide future studies in this field. Additionally, we will report about the results of a study evaluating the L2TOR system with Turkish-Dutch pre-schoolers learning Dutch as second language.

In preparation of the large scale evaluation study, the system designed for both the number and the space domain was tested in two separate pilots. Following each pilot study adjustments were made to obtain the final system that is being used in the large-scale evaluation study. At the same time, the design of the final study was agreed upon, test instruments were developed and tested and an extensive study protocol was written to ensure standard administration of the study. Moreover data management plans were made and basic data files were created. A total of 9 schools were recruited to participate in the study (all schools that do not teach English in the kindergarten class). In these schools a total of 208 children were recruited for participation. At this moment, after some attrition (e.g., children who knew more than half of the target words during the pre-test, children who during the study indicated they did no longer want to participate) 195 children participated in the study. Out of which about 156 already completed all the lessons and tests. Additional 37 children are taking part in the lessons at this moment. In the summer months we are planning to run the analyses and write the first publications based on the data of the large-scale study.

Besides the large-scale evaluation study within WP7, we have also conducted a study to evaluate the L2TOR system with a group of Turkish-Dutch children. We tested whether these children learn Dutch words better with a robot speaking only Dutch or with a bilingual robot speaking both Turkish and Dutch. Data collection of this study has been completed and initial results are included in this report. Further analyses are still planned and will be included in future reports and publications.

2 Pilot Study Number Domain

The lessons designed for the number domain were tested in a pilot study with 15 children. In this pilot study the lessons were used in the form also planned for the large evaluation study with several important differences. First, the feedback system was not yet implemented. Therefore, children received a standard feedback sentence when their response was not correct (Dutch equivalent of ‘this is not entirely correct’). Second, no iconic gestures were used. So basically, all the children received lessons in the non-iconic gestures condition. Finally, only the lessons from the number domain were included. Thus, children completed 4 lessons in total. In the first three lessons they were taught target words in English from the number domain and in the final lesson (which was shorter) all words were repeated. This last lesson was used for recap and consolidation of the words taught.

To test word knowledge right after each lesson, a short task was added to the lessons described in previous deliverables. In this task children had to choose out of three pictures or short animations the picture or animation that corresponds with an English word said by the robot. All target words from each lesson were included in this task and each word was included in two test items. This form of immediate post-test was also integrated in the large-scale evaluation study later on. The main results of this pilot have been published in Rintjema et al. (2018). This paper is added in Appendix I. A summary of the main findings is provided below.

The study by Rintjema et al. included 15 Dutch children (5 girls) aged between 5 and 6 years ($M=5.5$ years, $SD=4.6$ months). During the lessons the children were taught 17 words related to the number domain. See deliverable 1.1 for a list of target words. The structure of the lessons and set up have been described extensively in deliverable 1.1. At the end of each lesson (first three lessons) a short post-test was conducted to test what the children learned in each lesson. Before the start of the lessons series and after then last lesson children were asked to translate the target words from English to Dutch. Two additional testes (not included in the report of Rintjema et al.) in the post-test included asking the children to translate the target words from Dutch to English and a comprehension test in a format similar to the post-test conducted at the end of each lesson. The tests conducted at the end of each lesson and this comprehension test varied by that at the tests at the end of each lesson animations and images from the lessons were used. In the post-test, however, we used photos of objects and films of object and a human actor to display the target words. The items used were pilot tested ($N=15$) to ensure that 5-6 years old children could choose the right item when they heard the words in Dutch. An additional post-test including both translation tasks and the comprehension task was administered one week after the last lesson was completed to test retention. Finally, eleven adults (Mean age=25 years, $SD=2.8$ years) rated the task engagement of children on a five-point scale. Two fragments of 5 seconds from the beginning and the end of each lesson were rated. The fragments included only image and no audio. A total of 117 fragments were rated (3 videos were missing due to technical difficulties) by asking students from an HRI course to indicate how well the children were engaged with the task of interacting with the robot and tablet. Interrater reliability of coding of engagement was high

(ICC=.886). It should be noted that this way of coding engagement is not without limitations. For the pilot study we considered it to be sufficient, but for the large scale study we are currently working on developing a detailed coding scheme based on existing instruments such as the In-CLASS (Slot, Bleses, & Downer, 2018).

The results reported in Rintjema et al. (2018) show that in the post-test done immediately after the last lesson no learning gains were seen, but in the delayed post-test scores were higher than in the pre-test indicating that children did learn some words. However, given the small sample size and the fact that 8 out of 15 children were multilingual, out of which 4 spoke some English, no firm conclusions can be drawn based on this study. Moreover, the tests done during the final part of each lessons suggest that performance increased between the first and second lesson and then remained stable. This trend is remarkable as the words included in the first lesson (simple counting words mainly) seem to be easier than the other words. Engagement was lowest during the third lesson and did not vary between the other lessons. However, as mentioned above, given the sample size and other limitations no conclusions can be drawn based on these results. Nevertheless, the main goals of this pilot was not to draw conclusions but rather to test the system and the pre and post-test.

During this pilot the system proved to work relatively well, however some adjustments still needed to be made. These adjustments concerned: (1) the speech of the robot, to make the interaction more fluent and to improve the pronunciation, (2) the head and arm movements of the robot and (3) the tablet interface design, such as the size of objects or the speed of moving objects. In one of the lessons, the animals on the screen moved so quickly that children were not able to click on the animal.

3 Pilot Study Space Domain

The lessons designed for the space domain were tested in a pilot study conducted two months after the pilot study of the number domain lessons. The system had received several updates based on the observations of the number domain pilot, and the latest version of the system was tested with the space domain lessons. We conducted a pilot study with twelve five-year-old children. These children all received the lessons in the non-iconic gesture conditions. The iconic gestures had not been implemented yet in the system at the time of the pilot. The children received a total of four lessons: three lessons which were aimed at teaching a total of 17 English words related to the space domain, and one recap lesson which repeated all target words. The recap lesson had already been piloted during the number domain pilot, but with the number domain words. During the space domain pilot, we piloted the recap lesson with the space domain words, so that we would be able to merge the two recap lessons into one overall recap lesson that could be used for the evaluation study. The three lessons aimed at teaching the space domain words were piloted including the short tasks at the end of the lesson, developed to measure children's word knowledge right after each lesson. However, we only included this task to test the system, and we did not look into children's learning gains. We also did not include pre- or post-tests, as the goal of the pilot was not to measure children's learning from the system, but rather, to thoroughly test the system itself.

During this pilot, an experimenter observed the lessons and listed problems or breakdowns in the system. Also, potential improvements were noted. For example, children were asked in one of the lessons to act out 'running'. Children generally did not act out the word during the pilot. This is not crucial to the lesson, as children are allowed to not act out the word, but it would be a shame if most children do not act out the words. The lack of acting out during the pilot suggested that the request was not clear enough or that children needed more encouragement to act out the words. We clarified the request and added encouragement by adding "you can stand up if you'd like" to the storyboards.

Most improvements to the system were minor changes to the storyboards, such as the timing or pronunciation of utterances. A larger improvement to the system concerned feedback on children's repetition of the target words. During the pilot, children received either positive or negative feedback, depending on whether they repeated the target word correctly or incorrectly. We noticed that repeating the target words is a difficult task for children, and that it is important to encourage them to try and repeat the target words. Any attempt should be rewarded with praise, rather than being followed by negative feedback if children's pronunciation was not quite right. This could discourage them from trying to repeat the target words during the remainder of the lessons. Therefore, we decided to always provide positive feedback if the child made any attempt at repeating the target words, regardless of whether they repeated the target word correctly or incorrectly. If they did not repeat the target word, the system would simply repeat its request and try to help the child, like it does when the child does not perform a task on the tablet correctly.

Overall, no major problems or breakdowns occurred during this pilot study. The observations were solely used to further improve the system.

4 Design of Large-Scale Evaluation Study

For the large-scale evaluation study an extensive protocol has been written. This protocol was partially based on previous work from the team of Utrecht University (e.g., Mulder et al., 2014; Oudgenoeg-Paz et al., 2015). Further development of the protocol was done in collaboration between the team at Utrecht and in Tilburg. The full protocol is included in Appendix II. As the study was conducted only in the Netherlands the protocol was written in Dutch. On request we would be happy to provide an English translation of the document. Below we first describe the design of the study. Second we specify the main issues discussed in the protocol and provide the rationale to the decisions taken regarding the design of the study. Please also note that we pre-registered this study at [Aspredicted.org](https://aspredicted.org). The pre-registration document is attached in Appendix III and can also be found on: <https://aspredicted.org/6k93k.pdf>.

The study included Dutch children aged 5 to 6.5 years. Children were divided in four conditions. To prevent effects of schools all conditions were administered at each school. The four conditions included: 1) children working with the robot (and the tablet) while the robot performed only diectic but no iconic gestures; 2) children working with the robot (and the tablet) where the robot performed an iconic gesture with every target word; 3) children working with the tablet only; 4) control group of children who received 3 activities with the robot not involving learning English words. We aimed to have about 50 children in total in each experimental condition and 30 in the control group. To ensure this numbers are obtained we applied oversampling. 9 schools and 210 children were recruited to participate in the study. At the moment, data collection is still ongoing, therefore no results will be reported in this deliverable.

All children first received a pre-test. During this test we measured their knowledge of the 34 target words in English. Children who knew more than half of the words were excluded from participation. These children did do one of the control group activities with the robot, in order to not disappoint them, but these data will not be used in the analyses. During the pre-test some general skills related to word learning were measured for which we would like to control in the final analyses. These were: Dutch vocabulary level, verbal short-term memory and selective attention (Dunn & Dunn, 2005; Gathercole, 2006; Messer et al., 2010; Mulder et al., 2014; Verhagen & Leseman, 2016). Finally, during the pre-test children were asked about their perception of the robot using a newly developed questionnaire (roughly based on Jipson & Gelman, 2007). The pre-test was always done after children had received the group introduction of the robot and therefore had seen the robot once and could have formed some idea about it.

Following the pre-test children received 7 lessons with the robot and the tablet, or just with the tablet (depending on the conditions they were assigned to). In the first 6 lessons they were taught English words as specified in previous deliverables and in the final lesson they heard all the target words again and were reminded of the activities they did during the other lessons. During the final part of every lesson children did a short picture selection task to test receptive knowledge of the target words of this lesson (similar to the task described in Rintjema et al., 2018).

One to three days after the last lesson the first post-test was done in which children's productive knowledge of the target words was tested by asking them to translate the words from English to Dutch and vice versa. Receptive knowledge was measured using a picture selection task similar to the one used in the study by Rintjema et al., (2018). Moreover, in this post-test children were again asked about their perception of the robot. Two to three weeks after the first post-test a delayed post-test was administered that included the tests of receptive and productive vocabulary.

To make the administration as standard as possible a detailed protocol was written where we detailed issues such as how should you approach young children, how to arrange the room for a test and for lessons with the robot and listed behaviours children could display and how to cope with them (e.g., child that does not want to answer questions during the test). Additionally, some rules were set for when to stop a lesson (e.g., when a child clearly indicates they do not want to proceed and if, even after three attempts to motivate the child, they cannot be motivated); how many lessons can be done per week (usually 2 with a maximum of 3); time span between last lesson and post-test (1 to 3 days, never on the same day of the recap lesson); time span between the first and delayed post-test (2 to 3 weeks). Additionally, it was agreed that if for some reason (e.g., illness of children) the post-test cannot be administered on time, the test is conducted at the first possible moment. If children missed out on more than 3 lessons (e.g., due to illness) we excluded them from further participation as we no longer had time to catch up with the missing lesson. Note also that we originally wanted to leave 4 to 5 weeks between the first and delayed post-test, but given the constraints of school holidays in the Netherlands and other activities taking place we shortened this period to 2 to 3 weeks. This time period is also often used in studies (Marulis & Neuman, 2010). Finally, as we noticed that children in the tablet only condition were somewhat disappointed that they did not get to do the lessons with the robot, we let all the children in this condition do one of the activities of the control group after the first post-test. This way, they also got a chance to do an activity with the robot, but we did not risk influencing their perception of the robot before they answered the questions about it for the second and last time. All these design decisions included in the protocol were taken based on our previous experience and on discussions between the researchers involved.

Additionally, the protocol contains detailed information about the setup and working with the system with an extensive trouble shooting system. We used this as a dynamic document, where every time a new problem was diagnosed and solved it was added to the trouble shooting section. This section will be used in the future when we will work on guidelines for implementation of such a system in schools. Moreover, a detailed system was set up for data storage and registrations to ensure that all that data is safely stored (in accordance with EU and local regulations) and that the overview is maintained. This is an essential part of data collection of this scale and complexity.

All the people who were involved in data collection (including the researchers, students, assistants, interns) were trained either during a training day held in Utrecht or through watching a video recording of the training day. During data collection close contact was kept between all team members and one of the researchers was always available to provide online support to people in the field.

5 Study with Turkish-Dutch children

Part of the L2TOR project involves employing the L2TOR system with Turkish children learning Dutch as a second language (see our revised objectives). One of the main advantages of using a social robot with this population offers great promise, as - unlike teachers - social robots can talk in the child's L1 and therefore use L1 as support for learning L2. Some empirical evidence suggests that children learning a second language transfer their skills in L1 into L2 and therefore L1 is supportive for learning of L2 (e.g., Cummins, 1991; Leseman, Henrichs, Blom, & Verhagen, in press). Moreover, the use of L1 in the classroom may also contribute to engagement, self-esteem and positive identity development (Holmes, 2008; Pulinx, van Avermaet, & Agirdag, 2017).

In this study 67 Turkish-Dutch children aged between 48 and 71 months ($M = 57.16$ months, $SD = 6.28$) participated. Using a NAO robot and a Microsoft Surface Pro tablet, children were taught six Dutch words for which they knew the Turkish word but not the Dutch word. Detailed information about the method and initial results of this study were included in a thesis written by a research master student in Utrecht, Hanneke Leeuwestein, under the supervision of Paul Leseman, Josje Verhagen and Ora Oudgenoeg-Paz (see Appendix IV). The teams of UU and TIU are currently working on further analyses and plan to submit this work as a joint publication in the summer months. Here we provide a short summary of the information provided in the thesis.

The study used a within subjects design with two conditions. Half of the target words were taught by a bilingual robot that provided Turkish translations of the target words (L2-L1 condition). The other half of the target words was taught by a monolingual robot that used only Dutch (L2 only condition). All children were presented with both robots, in counterbalanced order. After the lesson, children's knowledge of the target words was measured as well as their vocabulary in both L1 and L2, their enjoyment of the lesson and which of the two robots they preferred. A week after the lesson knowledge of the target words was again measured.

Previous empirical studies (e.g., Mayo & Leseman, 2008; Demir-Vegter, Aarts, & Kurvers, 2014) have shown that the group of Turkish-Dutch children is extremely heterogeneous in terms of Dutch and Turkish proficiency levels. Therefore, it was not possible to use the same set of target words for all children (as these had to be words the children know in Turkish but not yet in Dutch). Thus, during a pre-test conducted up to one week before the experiment target words were chosen for each child separately. Specifically, first, during a pilot study, 40 words (all prepositions or spatial verbs) were chosen from a standard list developed by the municipality of Amsterdam as a guideline for teaching Dutch as second language in kindergartens ('*Basiswoordenlijst Amsterdamse Kleuters*'; Mulder, Timman, & Verhallen, 2009). Based on pilot testing 20 words out of this list were selected to be used in the final study. During the pretest children heard the words in Dutch and Turkish using recordings of native speakers and were asked to select which out of three pictures matched the word. Words for which the child knew the Turkish label but not the Dutch label were selected as target words. All children with at least four possible target words were included in the study. For about 25% of the children who participated in the pre-test had fewer than four possible target words could be

obtained/included. These children were excluded, leaving a final sample size of 67 children.

This design entails that children received different lessons as each child had a different set of words. The post-test included a translation task of the target words and a picture selection task. This picture selection task was also individually constructed per child so that for each item one picture represented the target word, one distractor a different target word the child was taught and the third picture was a target word the child was not taught.

Results show that children did learn on average about two thirds of the words in each condition. Moreover, word knowledge during the delayed post-test was significantly higher than during the immediate post-test. No difference was found between the two conditions, which indicates that children did not learn more words (or fewer words) when translations in L1 were offered. Further analyses showed that the level of Turkish vocabulary knowledge, but not Dutch vocabulary knowledge moderated the effect of condition: Children who scored lower on Turkish vocabulary knowledge learned more words in the L1-L2 condition than in the L2 only condition. The children with higher scores on Turkish vocabulary showed no difference between the two conditions.

This interaction effect is somewhat counterintuitive. Therefore we will explore it further before drawing definite conclusions and looking for possible explanations. We are currently analysing the amount of exposure to each target word. Dependent on children's performance during the lessons they might have heard target words more often if they required more feedback. We are now checking if children differ in the amount of feedback they received in each of the conditions. Additionally we aim to analyse the data again using methods (such as mixed-effect regression modeling) that can account for the fact that each child was taught different target words. In such analyses the items (target words) can be added as random factors allowing us to take into account differences between items. This will enable us to explore if this interaction effect might be the result of difference in difficulty between items.

Almost all children indicated they enjoyed working the robot. However, when asked which robot they preferred, the majority of the children (71.6%) indicated that they would prefer to play again with the bilingual robot. Children's robot preference did not moderate the effect of condition.

In sum, given the complex design of this study we are still running further analyses and would therefore not draw any firm conclusions at this point. We do believe that the fact that the children preferred the bilingual robot is meaningful as it points to a possible additional advantage of using L1 speaking social robots in educational settings with migrant children.

Appendix I: Rintjema et al. (2018)

Session: Late-Breaking Reports

HRI'18 Companion, March 5-8, 2018, Chicago, IL, USA

A Robot Teaching Young Children a Second Language: The Effect of Multiple Interactions on Engagement and Performance

Emmy Rintjema
Tilburg University
P.O. Box 90153, 5000 LE Tilburg
The Netherlands
emmyrintjema@gmail.com

Rianne van den Berghe
Utrecht University
P.O. Box 80125, 3508 TC Utrecht
The Netherlands
m.a.j.vandenbergh@uu.nl

Anne Kessels
Tilburg University
P.O. Box 90153, 5000 LE Tilburg
The Netherlands
kesselsanne@gmail.com

Jan de Wit
Tilburg University
P.O. Box 90153, 5000 LE Tilburg
The Netherlands
j.m.s.dewit@uvt.nl

Paul Vogt
Tilburg University
P.O. Box 90153, 5000 LE Tilburg
The Netherlands
p.a.vogt@uvt.nl

ABSTRACT

This paper explores the use of a social robot for one-on-one tutoring, in a study in which 15 children participated in four second-language tutoring sessions. Specifically, changes across sessions are measured on two dimensions: engagement and performance. Results have revealed a significant positive change in performance as well as a significant pattern in engagement across the interactions¹.

KEYWORDS

Child-robot interaction, language tutoring, education

ACM Reference format:

Emmy Rintjema, Rianne van den Berghe, Anne Kessels, Jan de Wit and Paul Vogt. 2018. A Robot Teaching Young Children a Second Language: The Effect of Multiple Interactions on Engagement and Performance. In *HRI'18 Companion: Conference on ACM/IEEE International Conference on Human-Robot Interaction, March 5-8, 2018, Chicago, IL, USA*. ACM, NY, NY, USA, 2 pages. DOI: <https://doi.org/10.1145/3173386.3177059>

1 INTRODUCTION

In recent years, increasing effort is made to design social robots as second language (L2) tutors [1-2]. The potential for robots to be effective tutors comes from various aspects, including their ability to tutor one-on-one [3], and to interact with children over multiple sessions for a longer period of time. However, most L2 learning studies thus far involving one-on-one tutoring lasted only one session [4], while those that have carried out long-term studies allowed one-on-many tutoring [2]. The current study explores the effects of a robot tutor over multiple child-robot L2

tutoring sessions. Specifically, this study investigates potential changes in two dimensions: children's engagement and their performance during the tutoring sessions.

2 OUR APPROACH

This study has been carried out as an extensive pilot study as a part of the L2TOR project, which aims to develop a robot tutor that helps young children learn an L2 [5]. To explore the changing relation between child and robot over the course of multiple one-on-one tutoring sessions, we conducted an experiment that consists of four sessions. During these sessions, the robot taught English (L2) vocabulary to Dutch (L1) children (age 5-6). Specifically, we examined performance and engagement. Performance has been measured as the degree to which children managed to complete tasks during the sessions and a word-knowledge task during a post-test. Task engagement has been measured as the degree to which children are actively involved in the tutoring session. This way, the current study explores how performance and engagement change over time. Based on findings by previous studies [4-7], it is expected that engagement will decrease over the course of the sessions. However, children's performance on the tasks will not be affected, as children will likely become more relaxed and familiarized to the robot and the task setting over time.

3 METHODS

The participants were 15 Dutch children (5 girls and 10 boys) with an average age of 5 years and 6 months ($SD = 4.6$ months). All parents gave their consent. The experimental setup contained a SoftBank Robotics NAO robot, which interacted autonomously with the child in four L2 tutoring sessions. Prior to the sessions, an introduction session was organized at the school to introduce the robot to the children. During the first three sessions, the robot taught the children a total of 17 words related to math such as numbers, adjectives (e.g., big) and verbs (e.g., add), via several games that the child and the robot played together on a Microsoft Surface Pro 4. The fourth session was a recap lesson to repeat and consolidate all 17 target words. In all sessions, the robot had the role of a peer-tutor, that is, the child and the robot

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

HRI'18 Companion, March 5 - 8, 2018, Chicago, IL, USA

© 2018 Copyright is held by the owner/author(s).

ACM ISBN 978-1-4503-5615-2/18/03.

<https://doi.org/10.1145/3173386.3177059>



together played games on the tablet and ‘learned’ the words. At the end of each session (except for the recap lesson), the child had to complete a task. During this task, several items appeared on the tablet and the robot asked the child to tap on a specific item on the screen to which it referred in L2. The robot automatically logged children’s answers to measure their performance on the tasks. In addition, to measure overall improvement of the child’s English skill, a pre-test and two post-tests were administered. The pre-test was administered right before the first interaction. One post-test was completed immediately after the recap lesson and the other a week later. In these tests, we asked the children to translate the 17 target words from English to Dutch. The experiments were recorded using a camera. The changes in performance over the course of the sessions were measured by comparing the scores on the tasks that were completed at the end of each of the three learning sessions (the recap lesson did not contain a task part and was not included in this measure). In addition, the scores on the pre- and post-tests were compared to assess the child’s knowledge of the target words. To measure changes in engagement over the course of the sessions, a perception study was conducted. Eleven participants (Mean age = 25, $SD = 2.8$) rated the task-engagement of children on a five-point differential scale for a total of 117 short video clips (5 seconds) without audio (2 per child per lesson, 3 videos were missing due to technical difficulties). A high degree of interrater reliability was found between the participants of the perception study. The average measured ICC was .886 ($F(116, 1160) = 8.74, p < .001$). The fragments were taken at specific moments in the robot’s script, the first a few minutes after the start and the second a few minutes before the end of the lesson.

3 RESULTS

On average, scores on the immediate post-test ($M = 3.64, SD = 3.08$) were higher than scores on the pre-test ($M = 2.43, SD = 2.41$), but this difference was not significant ($Mdif = 1, t(14) = 1.46, p = .165$). The scores on the delayed post-test ($M = 4.21, SD = 3.14$) were significantly higher than the scores on the pre-test ($Mdif = 1.79, t(13) = 2.29, p = .039$), indicating that children’s knowledge of the target words improved over time (max. possible score: 17). To test the relation between the amount of sessions with the robot and performance on the tasks and engagement, a one-way repeated measures ANOVA was performed for both dependent variables. For performance, the overall ANOVA was significant ($F(1,14) = 22.65, p < .001, \eta^2 = .72$), revealing a significant effect of session on performance. Performance increased between the first ($M = 0.64, SD = 0.35$) and second lesson ($M = 1.40, SD = 0.46$) ($Mdif = 0.96, p < .001$), but not between lesson two and lesson three ($M = 1.39, SD = 0.24$) (max. possible score: 2). For engagement, the overall ANOVA was also significant ($F(1,14) = 4.61, p = .014, \eta^2 = .02$), revealing a significant effect of session on engagement. No significant change was observed between the first ($M = 3.55, SD = 0.46$) and second lesson ($M = 3.64, SD = 0.36$). Engagement decreased between the second and third lesson ($M = 3.09, SD = 0.63$) ($Mdif = -0.55, p = .034$), and increased between the third and

fourth lesson ($M = 3.62, SD = 0.62$) ($Mdif = 0.53, p = .039$) (max. possible score: 5).

4 CONCLUSIONS

In this study, we carried out an extensive pilot to study long-term effects of L2 tutoring using a social robot. Results have revealed a positive relationship between time spent with the robot and performance on the learning tasks. Children improved their learning achievements after spending more time with the robot, possibly because they get more used to the robot as a tutor. However, we cannot know this for sure as the content of the individual lessons might have influenced the performance. Furthermore, results showed a decrease of engagement between the second and third session and an increase of engagement between the third and fourth session. The downward trend may be explained by familiarization with the robot. The positive change between the third and fourth session may have been caused by the content of the recap session being different from the three learning sessions or by the fact that children knew that it was the last time they got to play with the robot. While more extensive studies on changes in performance and engagement in longitudinal one-on-one tutoring need to be conducted, this explorative study has thus taken first steps and found promising results regarding how performance and engagement evolve over time in long-term child-robot L2 tutoring sessions. Moreover, this study has set the stage for a larger evaluation study planned in the near future involving more lessons, more children and different conditions.

Acknowledgements

We are grateful for the support of all L2TOR project members. In addition, we thank the school, parents and children for their participation. The research was financially supported by the EU H2020 L2TOR project (grant 688014).

REFERENCES

- [1] Gordon, G., Spaulding, S., Westlund, J. K., Lee, J. J., Plummer, L., Martinez, M., Das, M., Breazeal, C. (2016). Affective Personalization of a Social Robot Tutor for Children’s Second Language Skills. Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16), 3951 – 3957.
- [2] 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, 61-84.
- [3] VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educational Psychologist, 46(4):197-221.
- [4] De Wit, J., Schodde, T., Willemsen, B., Bergmann, K., De Haas, M., Kopp, S., Krahmer, E., Vogt, P. (2018). The Effect of a Robot’s Gestures and Adaptive Tutoring on Children’s Acquisition of Second Language Vocabularies. In Proceedings of ACM / IEEE International Conference on Human Robot Interaction, Chicago, Illinois USA, March 2018 (HRI 2018).
- [5] Belpaeme, T. et al. (2015) L2TOR - Second Language Tutoring using Social Robots. In Proceedings of 1st Int. Workshop on Educational Robots. Springer.
- [6] Leite, I., Martinho, C., & Paiva, A. (2013). Social Robots for Long-term Interaction: A Survey. International Journal of Social Robotics, 5(2), pp 291- 308. DOI: <https://doi.org/10.1007/s12369-013-0178-y>
- [7] Nalin, M., Baroni, I., Kruijff-Korbayova, I., Canamero, L., Lewis, M., Beck, A., Cuayahuitl, H., Sanna, A. (2012). Children’s adaption in multi-session interaction with a humanoid robot. RO-MAN, 2012 IEEE, DOI: 10.1109/ROMAN.2012.6343778

Appendix II: Protocol of the Large Evaluation Study

Protocol Evaluatiestudie L2TOR



Josje Verahgen, Ora Oudgenoeg-Paz, Mirjam de Haas, Jan de Wit & Rianne van den Berghe



Inhoudsopgave

1. Algemene informatie over de studie
 - 1.1 Doel
 - 1.2 Methode
 - 1.3 Onderzoeksopzet
2. Onderzoek bij jonge kinderen
 - 2.1 Opzet tijdens de lessen
 - 2.2 Testruimte
 - 2.3 Omgaan met jonge kinderen
3. Lessen
 - 3.1 Vooraf
 - 3.2 De afname
 - 3.3 Trouble shooting
 - 3.3.1 Tablet
 - 3.3.2 Robot
 - 3.3.3 Laptop
 - 3.4 Rol van de testleider
4. Tests
 - 4.1 Algemene informatie
 - 4.2 Instructies tussen de taken door
5. Pre-test
 - 5.1 Volgorde van de taken in de pre-test
 - 5.2 Peabody Picture Vocabulary Test
 - 5.3 Pre-test doelwoorden
 - 5.4 Aandachtstaak
 - 5.5 Nonwoordrepetitietaak
6. Post-tests
 - 6.1 Algemene instructies post-test 1
 - 6.2 Vertaaltaak T2 > T1
 - 6.3 Vertaaltaak T1 > T2
 - 6.4 Begripstaak
 - 6.5 Kindvragenlijst
 - 6.6 Algemene instructies post-test 2
 - 6.7 Aandachtspunten om op te letten tijdens de post-test
7. Proefpersoonwerving, dataopslag en overige informatie voor testleiders
 - 7.1 Werving en informed consent
 - 7.2 Proefpersoonnummers
 - 7.3 Oudervragenlijst
 - 7.4 Dataopslag
 - 7.5 Logboek
 - 7.4 Een sessie afbreken of uitstellen
 - 7.5 Video-opnames
8. Trainingsprocedure testleiders

Appendices

- I Pre-test doelwoorden
- II Vertaaltaak T2 > T1
- III Vertaaltaak T1 > T2
- IV Scoreformulier begripstaak
- V Stickerdieren
- VI Scoreformulier nonwoordrepetitietaak
- VII Perceptievragenlijst
- VIII Lijst van materiaal voor lessen en tests
- IX Meeneemlijst

1. Algemene informatie over de studie

1.1 Doel

Het doel van de evaluatiestudie in het L2TOR-project is het testen van de effectiviteit van het gebruik van een sociale robot voor het aanleren van woorden in een vreemde taal aan kleuters. De robot zal gecombineerd worden met een tablet en vergeleken worden met het gebruik van alleen een tablet, en 'no treatment'. De combinatie van een robot en een tablet zal gebruikt worden in twee condities. In de ene conditie gebruikt de robot alleen deictische gebaren en in de tweede gebruikt de robot deictische *en* iconische gebaren om het leren van woorden te ondersteunen.

1.2 Methode

Nederlandse kinderen van vijf jaar oud krijgen in zeven individuele sessies een aantal woorden in het Engels aangeboden met ofwel een robot en een tablet (robotconditie), ofwel alleen een tablet (tabletconditie). Een kleiner aantal kinderen doet drie keer een andere activiteit met de robot (bijvoorbeeld een dansje) ('no treatment' conditie). Tijdens de experimentele condities krijgen kinderen een woordenschattraining aangeboden, waarin zij spelletjes doen op een tablet. Tijdens deze spelletjes krijgen ze in een speelse context Engelse woorden aangeboden en wordt aan de kinderen gevraagd deze na te zeggen en kleine opdrachten uit te voeren op de tablet die gericht zijn op het aanleren van deze woorden. In de robotconditie is hierbij een robot aanwezig. De robot gedraagt zich als een leeftijdsgenoot die samen met het kind de Engelse woorden leert. In de tabletconditie werken kinderen alleen op de tablet.

In de training krijgen kinderen woorden aangeboden uit twee domeinen: (voorbereidend) rekenen en ruimtelijke taal. Voorbeelden van woorden uit het voorbereidend rekendomein zijn telwoorden, bijvoeglijk naamwoorden zoals 'high' en 'low', en werkwoorden zoals 'add' en 'take away'. Voorbeelden van woorden uit het ruimtelijk domein zijn voorzetsels zoals 'on' en 'next to' en werkwoorden zoals 'jumping' en 'running'. Zie tabel 1 voor een overzicht van alle lessen en activiteiten.

1.3 Onderzoekopzet

In een introductiesessie die voorafgaat aan de dataverzameling maken kinderen kennis met de robot en de tablet. Ze krijgen ook uitleg over hoe ze moeten werken met de robot en de tablet. Daarna doen alle kinderen, in een tweede sessie die voorafgaat aan de training, een pre-test. In deze pre-test wordt kennis van de doelwoorden getest. Ook doen kinderen in deze sessie een aantal andere taken waarin meer algemene vaardigheden worden getest, zoals fonologisch geheugen en selectieve aandacht.

Op de eerste dag van de lessen kom je weer de klas in voor een korte herinnering aan de kinderen. Vertel weer hoe je heet en laat de robot zien (deze hoeft je dan niet aan te zetten).

Zie tabel 2 voor een overzicht van alle sessies in het onderzoek.

Tabel 1. Setting, activiteiten en doelwoorden per les

Les	Setting	Activiteit	Doelwoorden
1	Dierentuin	Dieren in hun kooi doen	One, two, three, add, more, most
2	Bakkerij	Broden bakken	Four, five, take away, fewer, fewest
3	Dierentuin	Dieren zoeken	High, low, big, small, heavy, light
4	Fruitwinkel	Winkel opruimen	On, above, below, next to, falling
5	Bos	Sporen zoeken	In front of, behind, walking, running, jumping, flying
6	Speeltuin	Speeltuin maken	Left, right, catching, throwing, sliding, climbing
7	Recap	Een fotoalbum maken	Herhaling van alle doelwoorden

Table 2. Planning van alle sessies en activiteiten in het evaluatieonderzoek

Week	Activiteit	Tijdsindicatie
Week 1	Introductie van de robot aan de hele klas	30 min (incl. 15 min voorbereidingstijd)
	Pre-test	30 min per kind
	Uitzetten van de oudervragenlijst (ca. 2 weken voor aanvang van de lessen)	
Week 2	Les 1 (eerste dag weer herinneren aan introductie - 5 minuten hele klas).	30 min per kind
	Les 2	30 min per kind
Week 3	Les 3	30 min per kind
	Les 4	30 min per kind
Week 4	Les 5	30 min per kind
	Les 6	30 min per kind
Week 5	Recap	30 min per kind
	Post-test 1	30 min per kind
Week 9	Post-test 2	30 min per kind

2. Onderzoek bij jonge kinderen

Bij het testen van jonge kinderen kunnen kleine aanpassingen in de testopstelling, testafname en testruimte de resultaten beïnvloeden. Zo kunnen kleine verschillen in hoe ver een kind van de tablet zit van invloed zijn op de prestaties van een kind en op hoe leuk hij of zij het vindt om aan het onderzoek mee te doen. Het is dan ook heel belangrijk dat je je in elke sessie houdt aan onderstaande richtlijnen.

mea2.1 Opzet tijdens de lessen

In de taken moet het kind vaak een plaatje op de tablet verslepen. Het is belangrijk dat het kind dit zo moeiteloos mogelijk kan doen. Het is daarom belangrijk dat het kind niet te ver van de tablet af zit. Het kind moet niet hoeven reiken naar de tablet. Dit leidt af van de taak en kost onnodig veel energie.

In de robot condities zitten het kind en de robot in een hoek van 90 graden ten opzichte van elkaar (zie Figuur 2). De robot is altijd aan de rechterkant van het kind. De robot zit vlakbij de tablet en in een gehurkte positie.



Houd je verder steeds aan de volgende regels:

- Zorg ervoor dat het scherm van de tablet goed zichtbaar is. Pas eventueel de helderheid van het scherm aan, doe de gordijnen dicht of ga naar een andere hoek van de kamer.

2.2. Testruimte

- Neem de lessen en tests het liefst af in een ruimte waarmee het kind vertrouwd is. Als het kantoortje van de directeur de enige optie is kun je dit ook gebruiken, maar realiseer je dan wel dat kinderen zich waarschijnlijk minder op hun gemak voelen, in ieder geval tijdens de eerste sessie(s).
- Zorg dat er geen andere kinderen in de ruimte zijn.
- Als de school geen rustige ruimte heeft, dan kun je de lessen en tests eventueel in een rustige hoek op de gang afnemen. Bedenk echter wel dat dit niet de voorkeur heeft! Mocht het toch een keer gebeuren, noteer dan altijd in het logboek dat de testafname niet in een rustige ruimte heeft plaatsgevonden.
- Probeer om de verschillende les- en testsessies zoveel mogelijk in dezelfde ruimte te laten plaatsvinden. Dit kan niet altijd, maar heeft sterk de voorkeur boven verschillende ruimtes.
- Kies een plek in de ruimte die zo leeg mogelijk is, zodat kinderen niet afgeleid raken door speelgoed, interessante posters, etc. Ga ook niet te dicht bij een raam zitten waar kinderen afgeleid zouden kunnen raken door voorbijgangers. Kinderen moeten bij voorkeur met hun rug naar het raam zitten. Zet afleidende spullen in een andere hoek neer.
- Doe de deur dicht, zodat je zo min mogelijk last hebt van omgevingslawaaï.
- Hang een papier met “stilte alstublieft: experiment bezig” aan de buitenzijde van deze deur.

2.3. Omgaan met jonge kinderen

Vijfjarigen zijn over het algemeen gewend om de opdrachten van een leerkracht uit te voeren. Maar deelname aan een onderzoek is wel wat anders. Voor jonge kinderen kan het indrukwekkend zijn om mee te doen met een onderzoek in een individuele sessie (met een robot!). Let goed op de volgende punten in het contact met de kinderen, bijvoorbeeld wanneer je ze ophaalt uit de klas of wanneer je tussen de taken door een gesprekje met ze voert:

- Voel aan of een kind verlegen is (of juist helemaal niet) en speel daarop in. Als het nog wat angstig of afwachtend is, kijk het dan niet teveel aan en praat niet te direct, maar geef het kind de tijd om rond te kijken en aan je te wennen. Zeg een paar algemene dingen om het kind op zijn/haar gemak te stellen. Je kunt bijvoorbeeld zeggen “Wat heb jij een mooie schoenen aan!” of “Ik zag je net buiten op het schoolplein spelen”.
- Als je het idee hebt dat het kind een gesprekje wil voeren, stel dan gemakkelijke vragen als ‘Heb jij ook een tablet?’ en stel simpele vervolgvragen op antwoorden van het kind.
- Als een kind geïntimideerd lijkt door de robot als hij of zij de testruimte binnenkomt en ontdekt dat hij/zij alleen met de robot aan het werk moet, kan het helpen om te zeggen dat de robot een beetje verlegen is en dat dat helemaal niet erg is. Dit kan helpen om het



kind zich meer verwant te laten voelen met de robot en geeft meteen aan dat het niet erg is om even afwachtend/verlegen te zijn.

3. Lessen

Er zijn 4 verschillende condities: (1) tabletconditie, (2) robot met deictische gebaren, (3) robot met deictische en iconische gebaren en (4) een no-treatment conditie. Sectie 3.1 t/m 3.5 gelden voor de eerste drie condities. Sectie 3.6 gaat over de no-treatment conditie.

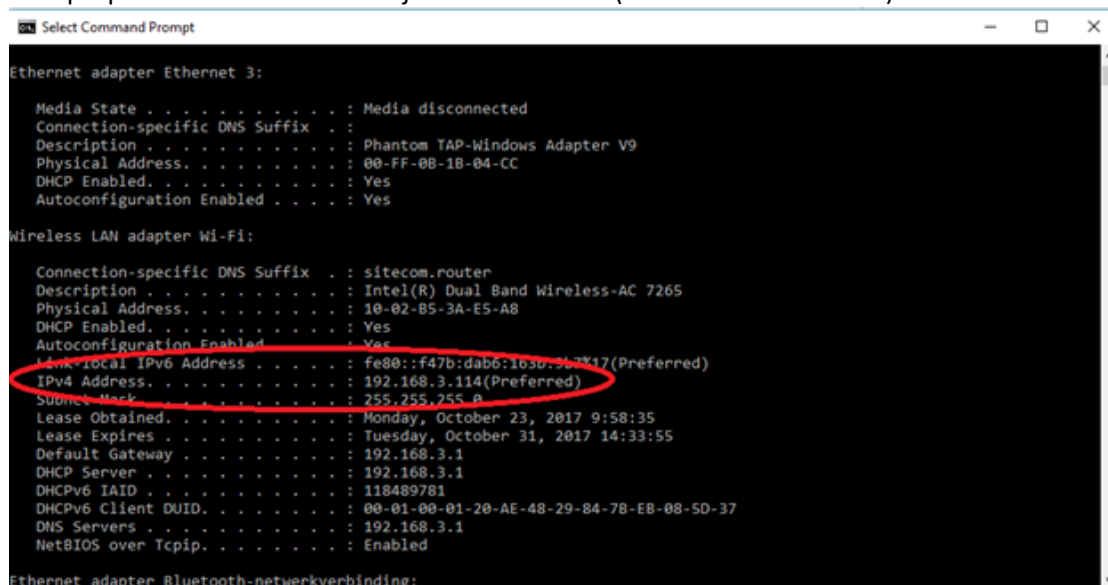
In het kort: hoe ziet de dag eruit:

1. Zet alle spullen klaar (robot, tablet, camera's)
2. Test de les in 1 conditie volledig
3. Test alle kinderen
4. Lunchpauze? -> Backup alle data
5. Einde van de dag? -> Backup alle data
6. Verwijder alle filmpjes van de sd kaartjes en stop deze terug in de camera

red3.1 Vooraf:

1. Vind alle stopcontacten, vind een plek voor de router, robot, tablet en de camera's. (Het wachtwoord van de UU routers en tablets is L2t0r!UU).
2. Stop de router in het stopcontact en zet hem aan door kort op de knop op zijn buik te drukken. Sluit de netwerkkabels op de niet afgeplakte netwerkpoorten aan. Sluit de netwerkkabels op de tablet en de robot aan. De laptop hoeft niet met een kabel op de router aangesloten te worden. De router hoeft niet met een kabel vanaf de blauwe poort op een internetuitgang in de muur aangesloten te worden.
3. Zet de robot aan (druk op de knop op zijn buik, plaats hem in een gehurkte positie met zijn armen op de knieën). Het is belangrijk dat je EERST de robot aanzet, dan pas de laptop en tablet (wacht met de laptop + tablet tot de robot opgestart is).
4. Plaats alle camera's op de juiste plek.
5. Nadat de robot zijn naam heeft gezegd + IP adres staat hij volledig aan. Is het IP adres 192.168.0.100? (op de Borne: 192.168.1.99, op de Meander: 192.168.3.100, op de Wegwijzer: 192.168.60.1) Dan hoef je verder niets te doen. Iets anders? Ga dan naar Troubleshooting, sectie 3.4.
 - a. Als de robot geen naam + IP adres gezegd heeft maar wel al voor een lange tijd aan staat, druk dan even op de aan-/uitknop om alsnog zijn naam+IP te horen.
 - b. Als de robot niet kan verbinden met het netwerk, ga naar sectie 3.4 'Troubleshooting'.
 - c. In alle andere gevallen, ga naar sectie 3.4 'Troubleshooting'.
6. Zet de laptop en de tablet aan (dus nadat de robot is opgestart).
7. Verbind de laptop en de tablet met de router (de laptop via wifi en de tablet met de netwerkkabel). De laptop en de tablet zullen (soms) aangeven dat ze geen verbinding hebben met het internet (door een internetpagina te openen), dat is geen probleem want ze hebben wel verbinding met de robot.
8. Ga nu naar de tablet.
 - a. Zorg dat het **volume** van de tablet op **74** staat (dit kan met de twee knoppen bovenop naast de aan/uit knop) -- deze lijkt regelmatig terug omlaag te springen dus houd dit in de gaten!

- b. Zorg dat de Wi-Fi **uit** staat (dit kan door rechts onderin op het icoontje te klikken en op de Wi-Fi knop te duwen (blauw = aan, grijs = uit)
- c. Ga naar start-menu, typ cmd (dit opent een zwart scherm) → typ ipconfig/all → zoek het IP adres achter IPv4 address (kijk even naar het voorbeeld hieronder, het ip adres is 192.168.3.114). Dit is het IP adres van de tablet, dit heb je nodig voor de laptop. Het zou iets moeten zijn als 192.168.x.x (zoals in het voorbeeld)



```

Select Command Prompt

Ethernet adapter Ethernet 3:

    Media State . . . . . : Media disconnected
    Connection-specific DNS Suffix . . : 
    Description . . . . . : Phantom TAP-Windows Adapter V9
    Physical Address. . . . . : 00-FF-00-1B-04-CC
    DHCP Enabled. . . . . : Yes
    Autoconfiguration Enabled . . . . : Yes

Wireless LAN adapter Wi-Fi:

    Connection-specific DNS Suffix . . : sitecom.router
    Description . . . . . : Intel(R) Dual Band Wireless-AC 7265
    Physical Address. . . . . : 10-02-B5-3A-E5-A8
    DHCP Enabled. . . . . : Yes
    Autoconfiguration Enabled . . . . : Yes
    Link-local IPv6 Address . . . . . : fe80::f47b:dab6:1b3b:5771%17(Preferred)
    IPv4 Address. . . . . : 192.168.3.114(Preferred)
    Subnet Mask . . . . . : 255.255.255.0
    Lease Obtained. . . . . : Monday, October 23, 2017 9:58:35
    Lease Expires . . . . . : Tuesday, October 31, 2017 14:33:55
    Default Gateway . . . . . : 192.168.3.1
    DHCP Server . . . . . : 192.168.3.1
    DHCPv6 IAID . . . . . : 118489781
    DHCPv6 Client DUID. . . . . : 00-01-00-01-20-AE-48-29-84-7B-EB-08-5D-37
    DNS Servers . . . . . : 192.168.3.1
    NetBIOS over Tcpip. . . . . : Enabled

Ethernet adapter Bluetooth-netwerkverbinding:
  
```

9. Nu kan je alles aan gaan zetten voor het experiment.
 - a. Op de tablet → dubbelklik op Start_experiment (staat op het bureaublad).
 - i. Let op, de robot gaat goed zitten. Als dit niet gebeurt, ga naar sectie 3.4 'Troubleshooting'.
 - ii. Als het goed is, krijg je op de tablet een lege internetpagina te zien van het experiment. Vergeet niet de internetpagina vol scherm te maken! (Gebruik F11)
 - b. Op de laptop → dubbelklik op WoZControl.exe (staat op het bureaublad in het mapje l2tor experiment). Vul bij IP: het ip adres van de tablet in (dus in het voorbeeld hierboven is dat 192.168.3.114). Klik op "connect". Er moeten 5 dingen online komen: Interactionmanager, outputmanager, tabletgame, controlpanel en underworlds. KinectVAD hoeft niet online te komen. Als dit niet zo is, ga naar sectie 3.4 'Troubleshooting'.
 10. Is het de tabletconditie? Plaats de tas over de robot heen. **Let op: voor de tablet conditie moet je eerst even de tablet conditie 'aanzetten' voordat je het test met het kind.** Maak iets van een test_tablet memory aan, en start eerst een keertje de les met het test_memory. Je merkt vanzelf dat de tablet inderdaad nog niets zegt de eerste keer, dus doe het dan daarna voor de tweede keer en voila! de tablet praat wel :-).
- Let op - als je na de tablet conditie een robot conditie gaat draaien, staat de robot op mute. zet de volume weer aan door het ip adres van de robot op de laptop in te toetsen en de volume weer aan te zetten via het Zora control panel.

3.2 De afname

Als je een kind voor het eerst test, moet je even uitproberen of de robot de naam van het kind goed zegt. Dat kan via het Control Panel of via de Zora Portal (dus door op je laptop Chrome te openen en het IP adres in te typen, en via Composer een tekstballonnetje te gebruiken). Of in het control panel bij robot speech (nadat je verbonden bent) de naam van het kind in te typen en op robot speech te klikken.

Is het de eerste keer testen? Vertel het kind dan even hoe de tablet werkt, hoe de robot met de tablet werkt en dat ze duidelijk moeten praten:

“Zo meteen ga je samen met Robin de robot een spelletje spelen op de tablet, en dan gaan jullie samen Engelse woorden leren! Robin de robot speelt een beetje anders met de tablet dan wij, want Robin de robot is natuurlijk een robot. Robin de robot kan zó (beeld uit) doen en dan beweegt er wat op de tablet. Als wij met de tablet spelen moeten we hem natuurlijk wel aanraken, dus echt zó (veeg over de tablet). Weet je nog dat Robin de robot niet zo goed kan horen? Als je zo meteen tegen Robin de robot wilt praten, moet je duidelijk aan de voorkant praten. Nou, volgens mij gaat dat hartstikke leuk worden als jullie zo meteen samen gaan spelen!”

1. Zet alle camera's op recording. Haal het kind op uit de klas (vraag of hij/zij even naar de WC moet gaan voordat jullie beginnen)
2. Op de laptop in het controlpanel:
 - a. Eerste keer testen? Ga dan naar de **tablet** en open het control panel. Verbind gewoon meteen zonder het IP adres aan te passen. Klik op “New” en vul daar de naam van het kind in, de fonetische uitspraak van de naam van het kind, het geslacht (0 voor een jongen, 1 voor een meisje), de geboortedatum, de school en de geboortetaal (wat voor al onze kinderen als het goed is Nederlands is) en het proefpersoonnummer in. Selecteer ook de conditie waarin het kind getest wordt.
 - b. Tweede (+ derde...) keer testen? Zoek het kind op in Child-Memory op je **laptop**. En klik op “Load”. Als je hier niet op klikt kan je niet op Start lesson drukken.
3. Kies het lesnummer bij session-file. Let op, we hebben twee domeinen en 3 of 4 lessen per domein. Je begint bij domein 1, en als je alle drie de lessen hebt gedaan, ga je verder naar domein 2. Les 4 is dus domein 2 les 1, les 5 is domein 2 les 2, enzovoorts.
4. Klik op Start Lesson
5. Tijdens het experiment kan je het experiment op pauze zetten door op Pause te klikken. Let op: de robot zal nog wel even zijn zinnen + gebaren afmaken, dus de pauze begint niet direct). Wil je weer verder gaan? Klik op Resume.
6. Tijdens het experiment moet je wel aangeven dat de kids het woord goed uitgesproken hebben, dat kan door te klikken op Correct Word op het control panel. Let op! Soms moet je meerdere keren klikken voordat het werkt (dit kan niet verkeerd gaan, het is niet mogelijk om te vaak te klikken).
7. Einde van de les? Zet de camera's uit (stop recording) en breng het kind terug naar de klas, vervolgens even je handen desinfecteren en begin maar weer van voor af aan.

3.3 Einde van de dag : Backup van de data

1. Was je handen op het einde van de test (ivm mogelijke ziektekiemen)

2. Eindig je met de tablet conditie? Zet dan even nog een robot conditie aan (tip: gebruik hiervoor een test memory en niet eentje van de kinderen) zodat het volume van de robot weer aangaat of zet het volume aan via de zora portaal (typ het ip adres van de robot in op een Chrome browser).
3. Zet alle elektronica uit en berg alle spullen op in de dozen/tassen en zet deze in de aangewezen ruimte/kast
4. Backup alle data
5. de film beelden van de SD kaart afhalen (of in ieder geval backuppen, en verwijderen op de SD kaart) op de laptop/externe harddisk en op Yoda uploaden. Deze lokale kopie moet bewaard blijven totdat Ora heeft kunnen dubbel checken dat alles goed staat op Yoda zodat we niet per ongeluk beelden kwijt raken.
6. De log bestanden van de tablet (kan je vinden onder C:\l2tor\logs en C:\l2tor\memories, of via de snelkoppelingen op het bureaublad) opslaan op Yoda (hier een lokale kopie van houden). **Verwijder de bestanden in c:\l2tor\memories niet**
7. Logboek updaten op Sharepoint en eventueel opmerkingen voor de volgende dagen noteren.
8. Geef aan op sharepoint welke kinderen je getest hebt.
9. papieren logboek die op school blijft updaten en eventueel opmerkingen voor de volgende dagen noteren.

3.4. Troubleshooting

Help het experiment start niet op?

check eerst of alle kabels overal inzitten en of je laptop verbonden is met het NaoNetwork / NaoNetwork2 / Utrecht L2tor netwerken.

3.3.1 Tablet

Help, de tablet kan niet opstarten!

Is de tablet verbonden met de elektriciteit? **Houd de tablet 30 minuten lang verbonden met de stroom.** Als hij dan nog steeds niet opstart, probeer dan het volgende:

1. Druk op de aan-/uitknop en druk op hetzelfde moment de volume omhoog knop in en houd deze ingedrukt voor 3 seconden. Je zal dan het UEFI scherm te zien krijgen (een zwart scherm met witte letters). Klik vervolgens op Exit en Restart. Klik NIET op de andere dingen!
2. Nog steeds niets? Bel de Microsoft helpdesk: 020-5001500.

3.3.2 Robot

Help, de robot verbindt niet met het netwerk!

Heb je zijn startknop ingedrukt en de robot zegt nog steeds dat het niet kan verbinden? Start dan Choregraphe op (programma op de tablet) en ga naar Connection -> Connect to -> daar zie je in het groen het robot IPadres. Verbind hier niet mee maar bekijk het IP alleen! Is het IP adres 192.168.0.100 (voor de Borne, Tilburg: 192.168.1.99)? Dan hoef je niets meer te doen, dan werkt het straks wel gewoon. Anders kijk hieronder.

Help, het IP adres is niet 192.168.0.100!

We gaan het IP adres zelf toevoegen aan ons experiment. Klik met de rechtermuisknop op start_experiment op het bureaublad van de tablet. Klik op properties. Hier zie je aan het einde van de geselecteerde tekst een ip adres (ziet er ongeveer uit als : 192.168.1.100), verander dat in het juiste IP adres van de robot.

Help, de robot start niet op!

Is de robot verbonden met de elektriciteit? Soms valt de oplader eruit of zakt uit de robot's oplaadpunt. Druk deze stevig aan. Houd de robot 30 minuten lang verbonden met de stroom. Gaan de lampjes van de robot wel aan? Kijk hieronder wat de kleuren van de robot betekenen. (overigens: in de tablet conditie gaan de lampjes van de robot uit, dat is normaal!) Geen lampjes aan? Bel QBMT/Softbank of ruil je robot om.

Help, de robot viel om!

Is er een kind in de buurt? Bekommer je eerst om het kind. Sluit zo snel mogelijk alles af. Breng het kind desnoods terug naar de klas en vertel dat ook de robot weleens valt. Start het experiment vervolgens opnieuw op.

Help, de robot beweegt raar!

Druk 2 keer SNEL achter elkaar op de aan-/uitknop. De robot zal een zucht geven en gaan hurken (of soms zal hij gaan staan, dat is ook niet erg). Mocht je te langzaam achter elkaar drukken zegt de robot zijn naam + IP, druk dan nog 2 keer.

Voor uitleg over de kleuren van de start knop van de robot, zie:

http://doc.aldebaran.com/1-14/nao/led_messages.html

3.3.3 Laptop**Help, de managers komen niet online!**

Probeer eerst alles opnieuw op te starten. Nog steeds niet?

We gaan het experiment starten vanaf de tablet in plaats van de externe laptop.

1. Ga naar de tablet en open 'start_experiment' en open "control panel" op het bureaublad.
2. Klik meteen op 'connect' (zonder IP te veranderen).
3. Er moeten 5 dingen online komen: Interactionmanager, outputmanager, tabletgame, controlpanel en underworlds. KinectVAD hoeft niet online te komen.

Werkt het nog steeds niet? Dan gaan we het experiment uitgebreider starten.

Sluit alles af op de tablet.

- a. Open de map "Experiment" (op het bureaublad van de tablet).
- b. Open 1. Start_connectionManager.
- c. **Open nu het control panel op je laptop.** Vul het IP adres van de tablet linksboven in en druk op Connect. Even wachten, en dan wordt ControlPanel groen in het control panel, want het komt online.
- d. Open 2. Start_underworlds (dit is weer op de tablet).
- e. Open 3. Start_interactionmanager.

- f. Open 4. Start_outputmanager (robot beweegt). Let op dat het IP adres van de robot gelijk is aan hoe het in dit bestandje staat. Je kunt dit vooraf controleren door met je rechtermuisknop erop te klikken, 'Edit' te doen en dan zie je het IP adres staan. Als het anders is, pas het dan aan, druk op Save en sluit het bestandje weer.
- g. Open 5. Start_tabletgame (de browser gaat open). Let op, sluit de browser eerst voor dat je Start_tabletgame opent.
- h. Bekijk het control panel op je laptop. Er moeten 5 dingen online komen/zijn: Interactionmanager, outputmanager, tabletgame, controlpanel en underworlds. KinectVAD hoeft niet online te komen.

4. Werkt het nog steeds niet? Test dan dezelfde stappen alleen via de tablet, dus:

Sluit alles af op de tablet.

- a. Open de map "Experiment" (op het bureaublad van de tablet).
- b. Open 1. Start_connectionManager.
- c. **Open nu het control panel op de tablet.** Druk direct op Connect. Even wachten, en dan wordt ControlPanel groen in het control panel, want het komt online.
- d. Open 2. Start_underworlds (dit is weer op de tablet).
- e. Open 3. Start_interactionmanager.
- f. Open 4. Start_outputmanager (robot beweegt). Let op dat het IP adres van de robot gelijk is aan hoe het in dit bestandje staat. Je kunt dit vooraf controleren door met je rechtermuisknop erop te klikken, 'Edit' te doen en dan zie je het IP adres staan. Als het anders is, pas het dan aan, druk op Save en sluit het bestandje weer.
- g. Open 5. Start_tabletgame (de browser gaat open). Let op, sluit de browser eerst voor dat je Start_tabletgame opent.
- h. Bekijk het control panel op de tablet. Er moeten 5 dingen online komen/zijn: Interactionmanager, outputmanager, tabletgame, controlpanel en underworlds. KinectVAD hoeft niet online te komen.
- i. **Let op: je kan het experiment niet draaien op deze manier. Dit is gewoon om te testen of het op deze manier wel werkt.**

3.5 Rol van de testleider

Tijdens de lessen zit de testleider in een hoek van de kamer en laat deze het kind met de robot werken. Probeer uit het zicht van het kind te zitten en zorg dat het lijkt alsof je met je eigen dingen bezig bent (kijk op de laptop). Houd ondertussen de interactie wel steeds in de gaten. Probeer niet in het zicht van de robot te gaan zitten om te voorkomen dat het je gezicht gaat volgen. Soms kan het niet anders, probeer dan te zorgen dat de robot niet jouw gezicht gaat volgen. Dit kan je doen door:

- 1. Je haren los te dragen (als je lang haar hebt) en dit voor je hoofd doen
- 2. Je hand voor je gezicht houden als de robot richting jou kijkt
- 3. Een boek/tijdschrift voor je hoofd houden.
- 4. Op het moment dat de robot jou blijft volgen, klik je op 'reset gaze' in het control panel. Dan gaat de robot opnieuw zoeken naar gezichten vanuit de positie waar hij ongeveer denkt dat het kind zit (links van hem).

Het is de bedoeling dat het kind zelf met de robot aan de slag gaat, maar mocht er iets niet goed gaan (technisch gezien) moet je wel hulp bieden.

Let op tijdens alle lessen, maar voornamelijk tijdens de eerste les, of het kind begrijpt hoe hij/zij moet werken met de tablet en de robot. Waar nodig bied je hulp.

- Als het kind niets op de tablet aanraakt: ga naast het kind zitten en moedig het kind aan om dit wel te doen. Als na twee keer aanmoedigen het kind nog steeds niets aanraakt, doe het dan 1 keer voor. Als het kind het dan nog steeds niet doet, zet je de robot even op pauze (via control panel) en oefen je met het kind.
- Als het kind niet lukt om de objecten op de tablet goed te verslepen: leg aan het kind uit dat hij/zij de objecten moet aanraken en ook moet blijven aanraken tot ze op de juiste locatie staan. Als na twee keer uitleg het kind dit nog steeds niets goed doet, doe het dan een keer voor. Als dat nog steeds niet goed gaat, pauzeer je de robot en oefen je even met het kind. Je kan geen objecten op de tabletgame bewegen.
- Als het kind niets zegt als de robot vraagt om iets te zeggen: ga bij het kind zitten en probeer het kind aan te moedigen om toch wat te zeggen. Als na twee keer aanmoedigen het kind nog steeds niets zegt, doe het dan een keer voor. Zeg tegen het kind: “Zullen we het samen doen?” en zeg het woord in ieder geval zelf na. Als het kind daarna nog steeds niets zegt, zet je de robot even op pauze en ga je het nazeggen even ‘droog’ oefenen met het kind.
- Als het kind te zacht of in de verkeerde richting praat, zeg: “Praat maar tegen het gezicht van Robin!” en “Kijk, hier” (wijs dan het gezicht aan). Als het kind dit na twee keer uitleg nog niet doet, doe het dan een keer voor. Als het nog steeds niet goed gaat, pauzeer je de robot even en oefen je dit met het kind.

3.6 No-treatment conditie

In de no-treatment conditie spelen kinderen wel met de robot, maar doen ze geen Engels lesje. Met deze conditie willen we inzicht krijgen in of kinderen niet door andere omstandigheden dan onze lesjes de Engelse woorden leren, bijvoorbeeld door met andere kinderen over de lesjes te praten. Kinderen in de no-treatment conditie doen in tegenstelling tot de andere kinderen niet twee, maar één sessie per week. Zij doen in totaal 3 activiteiten met de robot. Deze neem je als volgt af:

- Haal de robot van de map af en zet de robot op een gladde ondergrond. Dus doe dit niet op tapijt bijvoorbeeld.
- Zet de camera's omhoog zodat je ook de kinderen filmt als zij staan!
- Zet robot aan.
- Maak verbinding met de robot (ofwel via de router, ofwel via de robots eigen wifi-netwerk).
- Open Google Chrome en typ het IP-adres van de robot in. Nu opent de Zora Portal.
- Zet het volume op 50%.
- Ga naar Sturing en zet automatisch gezichten volgen en met de ogen knipperen aan.
- Ga naar Composer (en indien de website het aangeeft, kies voor Simpele composer).
- Open de activiteit, selecteer het eerste spraak icoontje. Vul de naam van het kind in.

- Haal het kind op.
- Druk op play.

De activiteiten zijn als volgt:

Activiteit 1

Robot zegt: Hoi (naam kind)

Robot zegt: Vandaag gaan we samen een dansje doen! Ik hou heel erg van dansen en ik hou ook heel erg van de film Frozen. Laten we samen dansen op 'laat het los'! Doe maar mee!

Laat het los.

Robot zegt: Dat was leuk! Dat was het voor vandaag. Doei.

Activiteit 2

Robot zegt: Hallo (naam kind)

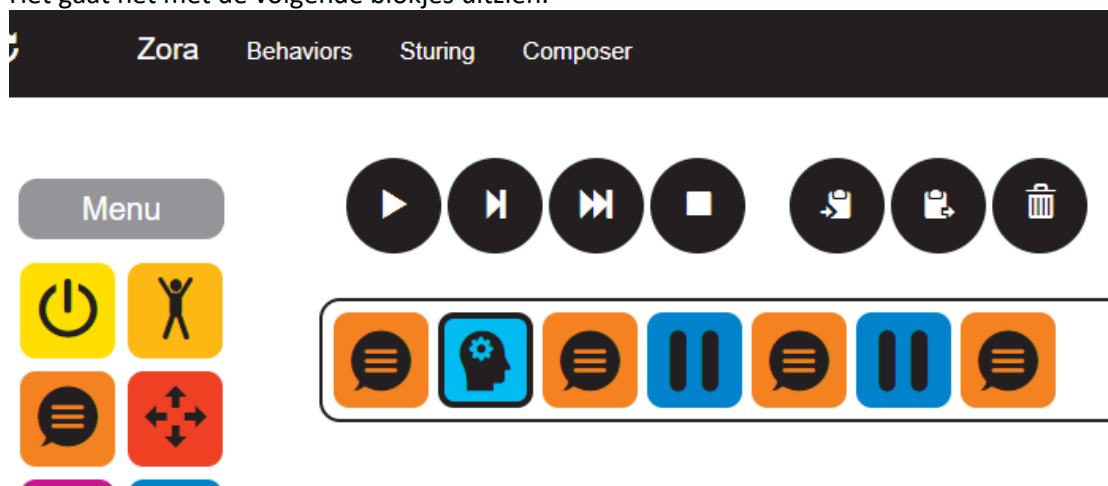
Robot zegt: Vandaag gaan we een ander dansje doen! Vandaag gaan we op hoofd schouders knie en teen dansen.

Hoofd schouders knie en teen

Robot zegt: Dat was leuk! Dat was het weer voor vandaag. Tot de volgende keer.

Activiteit 3

Het gaat het met de volgende blokjes uitzien:



Eerste spreekblokje:

Hallo <naam>, vandaag gaan we weer een dansje doen! Weet je nog dat we de vorige keer Frozen en hoofd schouders knie en teen hebben gedanst? Nu gaan we een ander dansje doen.

Tweede blokje (behavior):
Vogeltjesdans

Derde blokje (spreken):
dit was leuk. Ik vond het leuk om met jou te spelen. Hoe vond je het?

Vierde blokje (pauze -> stel timer in op 5 seconden, is een beetje gokken. Test dit even met 1 kind en pas het in verband nodig omhoog aan, verander dat ook hier):

Vijfde blokje (spreken):
Is er nog iets wat je tegen mij wilt zeggen?

Zesde blokje (pauze -> stel timer in op 10 seconden, beetje gokken weer. Test dit even met 1 kind en pas het in verband nodig omhoog aan, verander dat ook hier):

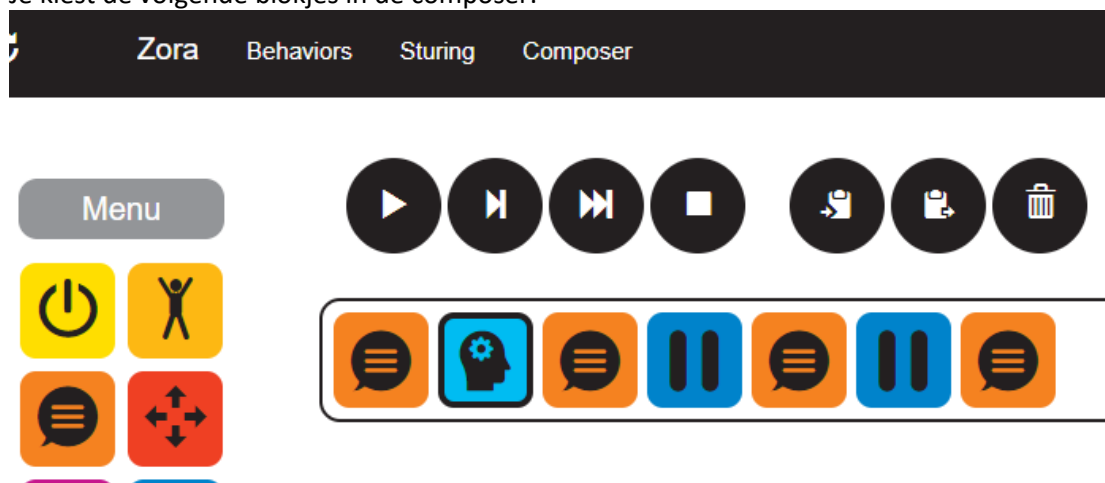
Zevende blokje (spreken):
oke. dit was de laatste keer dat wij samen hebben gespeeld. Dankjewel voor alles. Doe!

3.6.7 Outro voor kinderen in de tablet conditie

Om de kinderen in de tabletconditie ook een kans te geven om een activiteit met de robot te doen krijgen zij een 'outro': een korte groepsactiviteit met de robot. Deze doen we na de eerste post-test (dus nadat deze kinderen in ieder geval de eerste post-test hebben gehad, om hun perceptie van de robot niet beïnvloeden).

De outro wordt gedaan via het Zora portaal, op dezelfde manier als activiteit 3 (als robot en laptop aan zijn en verbonden met de router zijn open je google chrome op de laptop en daar toets je het IP adres van de robot in, dan kom je in het Zora portaal en daar kies je composer):

Je kiest de volgende blokjes in de composer:



Eerste spreekblokje:

Hallo <namen van alle kinderen in de groep (dus: Jantje, Pietje, ...) > wat leuk dat jullie er zijn!
Vandaag gaan we samen een dansje doen. Doen jullie mee? Sta maar op! Ik houd heel erg van de film Frozen, dus laten we dansen Laat het los!

Tweede blokje (behavior):

Laat het los

Derde blokje (spreken):

Dit was leuk. Ik vond het leuk om met jullie te dansen! Vonden jullie het ook leuk?

Vierde blokje (pauze -> stel timer in op 5 seconden.

Vijfde blokje (spreken):

Dit was de laatste keer dat ik met jullie heb gespeeld. Nu ga ik naar andere scholen. Is er nog iets wat jullie tegen mij willen zeggen?

Zesde blokje (pauze -> stel timer in op 10 seconden.

Zevende blokje (spreken):

Oke nu moet ik gaan. Dank jullie wel voor het spelen. Doe!

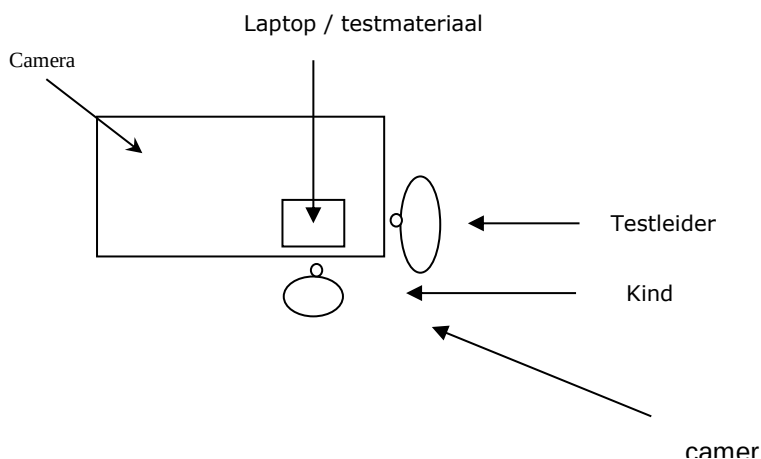
4. Tests

4.1. Algemene informatie

Er zijn drie testsessies per kind: één voorafgaand aan de training en twee post-tests. In elke testsessie worden meerdere tests afgenomen. De testafname is gestandaardiseerd. Dit houdt in dat de testjes altijd precies volgens de testinstructie en bij alle kinderen op dezelfde manier moeten worden afgenomen. Dit kan lastig zijn omdat elke testsituatie weer net even anders is, kinderen vaak anders op iets reageren, etc. Het is echter heel belangrijk om je te realiseren dat de data alleen betrouwbaar zijn als de taakafname altijd op dezelfde manier gebeurt. Sterker nog: de resultaten zijn niet bruikbaar als verschillende testleiders taken anders afnemen of als één testleider de taken op verschillende momenten anders afneemt. Over alle stappen in de afname en constructie van de testjes is nagedacht, ook over details en zaken die heel vanzelfsprekend lijken. Het is dus heel belangrijk dat je je bij de afname van de testjes precies houdt aan de testinstructie en daar nooit van afwijkt. Alleen dan kunnen we spreken van een gestandaardiseerde testafname.

Na elke test krijgen kinderen een sticker als beloning en zodat ze gemotiveerd blijven om mee te doen. Kinderen mogen de sticker zelf uitzoeken en op een vel papier met een afbeelding van een dier plakken (zie appendix).

In elke testsessie is er een vaste opstelling (zie de figuur hieronder). Deze is als volgt: de testleider zit in een hoek van 90 graden ten opzichte van het kind zodat hij/zij niet alleen het laptopscherm kan zien maar ook het gezicht van het kind. De videocamera is zo opgesteld dat de hele testopstelling goed zichtbaar is op de opnames, inclusief het gezicht van het kind.



4.2 Instructies tussen de taken door

Voordat je begint met de eerste taak van de sessie, geef je altijd de volgende instructie. Geef deze altijd letterlijk:

“Je gaat vandaag een paar spelletjes doen. Na elk spelletje krijg je een sticker. Je mag de sticker dan op dit blaadje, de sticker-kikker, plakken, OK?”.

Voordat je de volgende taak doet, zeg je altijd op een vriendelijke toon tegen het kind:

“We gaan nu het volgende spelletje doen!”

Let op: stel geen ja/nee-vragen aan het kind als “Zullen we nog een spelletje doen?” of “Wil jij nog een spelletje doen?”. Dit soort vragen stelt het kind in staat om ‘nee’ te zeggen en dat kan beter worden voorkomen. Over het algemeen vinden kinderen de spelletjes leuk om te doen en als kinderen het minder leuk vinden, is het beter om ze geen keuze te geven maar de testjes gewoon af te nemen. Je kunt wel zoiets zeggen als “Nu gaan we de laatste doen!” of “Het is bijna klaar” tegen het eind van een sessie als je ziet dat een kind het niet meer zo leuk vindt. Maar zorg ervoor dat je niet sneller door de test heen gaat omdat het kind zijn geduld kwijt raakt. Het is belangrijk om alle testen op dezelfde manier af te nemen.

N.B. Let op dat het altijd hebt over spelletjes tegen het kind, en nooit over tests of taken!

5 Pre-test

5.1 Volgorde van de taken in de pre-test

Een week voordat de training begint, vindt een testsessie plaats waarin een aantal tests worden afgenomen. Deze worden in een vaste volgorde afgenomen:

1. Peabody Picture Vocabulary Test (Nederlands)
2. Pre-test doelwoorden
3. Aandachtstaak
4. Nonwoordrepetitietask
5. Vragenlijst perceptie robot

Hieronder wordt elke taak apart beschreven.

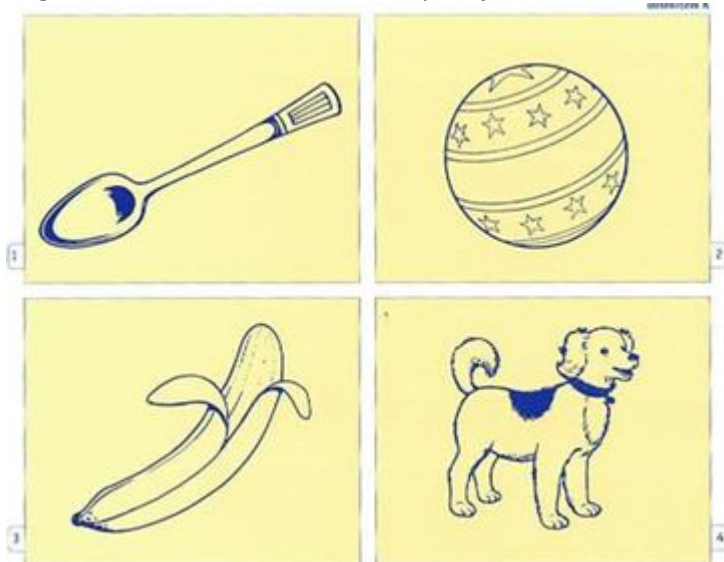
5.2 Peabody Picture Vocabulary Test (Nederlandse versie)

Nodig: PPVT-testplaten

Zie ook de PPVT handleiding.

De PPVT meet receptieve woordenschat (dus: Hoeveel woorden begrijpt het kind?)

Je gebruikt hiervoor een boek met plaatjes:



Gebruik de meegeleverde scoreformulieren. Noteer de kindcode en naam van het kind op elk formulier en reken de leeftijd in jaren en maanden uit.

- Neem de voorbeeldopgaven af. dit zijn er twee.

- Bij elk plaatje vraag je “Waar is: X?” (“ovaal”, “eiland”).
Let op – gebruik geen lidwoorden.
De test bestaat uit sets van ieder 12 woorden.
Er is een instapset, startset en afbreekset.
Volg de instap- en afbreekregels altijd precies op
- Begin bij de set voor de betreffende leeftijdsgroep. Dit is de instapset. (bij kinderen tussen 5 jaar en 5,5 jaar is dit set 5, bij kinderen tussen 5,5 en 6 jaar is dit set 6. Een kind van 5,5 begint bij set 6.
- Het kind kiest per testplaat steeds een afbeelding bij een mondeling aangeboden woord.
- Bij vijf of meer fout in de instapset, ga je een set terug. Je gaat terug tot een set is afgenomen waarin maximaal vier items fout zijn. Dit is de startset.
- Ga verder met de volgende set na de startset.
- De test wordt afgebroken na de set waarin negen of meer items fout zijn. Dit is de afbreekset.
- Je slaat de sets over die je al gehad hebt (bijvoorbeeld als je bij vijf begint, dan terug gaat naar vier en je zou weer terug naar vijf moeten, kan je die overslaan maar bij zes weer verder gaan).

Geef altijd positieve feedback: “Goed zo!” “Ja!” “Heel goed!”. Doe dit na elk item, dus ook als het antwoord fout is.

Geeft het kind bij een item geen antwoord? Of zegt het “ik weet niet”? --> scoor het antwoord apart, maar tel het als fout bij de beslissing over doorgaan/afbreken.

voorbeeld afname van PPVT:

Set 5 – 4 goed (8 fout) / instap set

Set 4 – 8 goed (4 fout) / startset (maximaal 4 fout)

Set 6 – 7 goed (5 fout)

Set 7 – 4 goed (8 fout)

Set 8 – 2 goed (10 fout) / afbreekset (meer dan 8 fout)

Scoor elk antwoord van een kind direct op het scoreformulier

Berekenen van de scores:

Tel alle woorden die het kind goed heeft

Lage sets die niet zijn afgenomen tellen voor 12 woorden per set.

Vul de totaalscore van het kind in op het scoreformulier / alternatief - trek het totaal aantal fout af van het nummer van het laatste item (= laatste item van de afbreekset).

Bereken de score niet in bijzijn van het kind maar direct na het testen of aan het einde van de dag.

Vul de totaalscore aan het eind van elke testdag in in het logboek van alle kinderen die je die dag hebt getest

Wat als...?

- Het kind niet wijst bij het oefenitem?
 - Stel de vraag nog een keer.
 - Als het kind nog niet wijst, wijs je zelf de hond aan en zeg je 'Kijk! Daar is de hond'.
 - Als het kind nog niet wijst, vraag je naar de andere voorwerpen op het plaatje ('bal', 'lepel').
- Het kind niet wijst bij een testitem?
 - Stel de vraag nog een keer.
 - Nog geen reactie? Stel de vraag nog een keer.
 - Als het kind nog niets aanwijst, scoor je geen antwoord (=0)
- Het kind alle plaatjes wil benoemen?
 - Laat het kind alles rustig benoemen.
 - Probeer de vraag 'Waar is X?' het vervolg te stellen voordat kind zelf de plaatjes benoemt.
- Het kind het plaatje eerder benoemt dan je de vraag hebt gesteld?
 - Stel de vraag nog een keer.
 - Scoor dan pas het antwoord.
- het Kind eerst een plaatje aanwijst en daarna een ander plaatje?
 - Scoor het laatste antwoord.

5.3 Pre-test doelwoorden

Het doel van deze pre-test is om te meten of kinderen de doelwoorden al weten.

Procedure

- Lees de instructie in Appendix A letterlijk voor aan het kind en kies dan de goede lijst, afhankelijk van de sessie (zie Tabel 3).
- De taak draait op powerpoint. Bij iedere slide zie je hetzelfde plaatje en bij het klikken op de speaker hoor je het betreffende doelwoord.
- Bij ieder item wordt het doelwoord door de computer gezegd. Jij leest wel de rest van de zin voor. Bijvoorbeeld (wat jij zegt is schuin en wat de computer zegt is dik gedrukt: *wat betekent **big** in het Nederlands?*
- Let goed op dat het kind geconcentreerd is en aan het luisteren voordat je het geluid van de computer afspeelt.
- Scoor elk item meteen na een antwoord van het kind door een kruisje te zetten in het vakje bij 'goed antwoord' of 'incorrect' of 'geen antwoord'. Bij een verkeerd antwoord noteer je wat het kind gezegd heeft.
- Herhaal de woorden niet zelf.
- Bereken de score later: 1 punt voor goede antwoorden, 0 voor foute antwoorden of geen antwoord.

- Feedback:
 - Als kinderen een antwoord geven, geef ze dan heel positieve feedback (“Goed zo”, “Super goed!”). Dit mag bij het goede antwoord, maar óók bij het verkeerde antwoord. We proberen de kinderen niets te leren, dus zeggen niet of het goed is of niet. Wel willen we ze motiveren om iets te proberen door de positieve feedback.
 - Als kinderen geen antwoord geven of zeggen dat ze het niet weten, zeg je dat dat helemaal geen probleem is dat ze het woord niet kennen omdat ze het zullen gaan leren.

Wat als...?

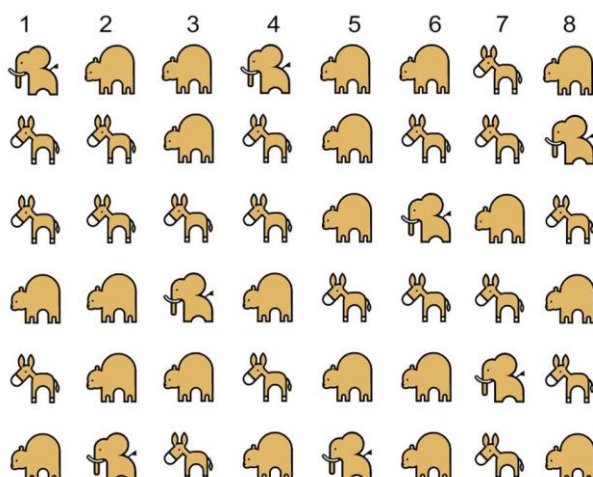
- Kinderen aangeven dat ze niet meer verder willen?
 - Vertel dat het helemaal geen probleem is dat ze de woorden nog niet kennen, maar eigenlijk juist wat we graag willen zien, omdat ze de woorden dan kunnen leren. Je kunt ook zeggen dat er niet meer zoveel woorden zijn (“*We zijn er bijna. Er zijn nog maar een paar woordjes*”).
- Kinderen sommige of zelfs (bijna) alle woorden kennen?
 - Als ze 17 woorden kennen, kunnen ze gewoon meedoen. Als ze er 18 of meer kennen, zeg je dat ze het heel goed hebben gedaan en dat ze even met de robot mogen spelen. Laat het kind even wat leuke dingen doen met de robot (eerste activiteit van de controle groep) en breng het daarna terug naar de klas. Deze kinderen worden uitgesloten van deelname aan het onderzoek maar mogen wel even kort met de robot spelen zodat ze niet gepasseerd worden. Schrijf altijd nauwkeurig op in het logboek (zie Sectie 7.5) welk kind wel was aangemeld voor de studie, maar niet kon meedoen vanwege te veel voorkennis, zodat we het overzicht houden van welke kinderen wel waren aangemeld maar moesten worden uitgesloten van deelname.

5.4 Aandachtstaak

De aandachtstaak is een visuele zoektaak waarin kinderen doelplaatjes zoeken en daarbij afleidende plaatjes moeten negeren. De taak meet selectieve aandacht. De taak draait op E-prime, versie 2.0. Zorg ervoor dat er een ‘runtime’-versie van dit programma is geïnstalleerd op de laptop waarop de tests worden afgenomen.

Taakbeschrijving

In deze taak ziet het kind een heleboel plaatjes van drie soorten dieren (beertjes, paardjes en olifantjes) tegelijk op het scherm. Het is de bedoeling dat het kind zo snel mogelijk de olifantjes zoekt en aanwijst. De testleider spoort het kind aan om heel snel te zoeken en alle olifantjes te vinden. De taak heeft vier testitems (d.w.z. vier keer een scherm vol met dieren).



Procedure

- Zeg: “Wij gaan olifantjes zoeken!”
- Start de taak en volg de aanwijzingen op het scherm.
- Gebruik de oefenitems om aan het kind duidelijk te maken dat hij/zij alleen de olifantjes moet aanwijzen en dat hij/zij dit zo snel mogelijk moet doen. Doe dit op de volgende manier:
 - Lees steeds de letterlijke tekst voor die op het scherm staat.
 - Om de tekst duidelijker te maken, ondersteun je deze door de plaatjes aan te wijzen wanneer dit wordt aangegeven op het scherm
 - Zodra het kind een olifant aanwijst, toets je het getal in dat boven de olifant staat. Hierdoor verschijnt er een streep door de olifant heen. Probeer dit altijd zo snel mogelijk te doen. Het kind leert zo dat er direct een streep door de olifant heen verschijnt als hij/zij deze aanwijst.



- Na de oefenitems moet het kind tijdens vier testitems van 40 seconden zoveel mogelijk olifantjes aanwijzen.
- Na het derde testitem volgt er een extra instructiegedeelte, om het vierde testitem in te leiden. Lees ook hier de instructie letterlijk voor.
- Geef veel feedback tijdens deze taak en doe dit altijd als volgt:
 - Als het kind een olifantje aanwijst, zeg je: “Goed zo, waar is er nog één?”
 - Als het kind een ander beest dan een olifantje aanwijst, zeg je: “Nee, waar is een olifantje?”

- Als het kind hetzelfde olifantje meerdere keren aanwijst, zeg je: “Nee, waar is nog een ander olifantje?”
- Verder is het belangrijk om het kind gedurende de hele taak aan te blijven sporen snel nog meer olifantjes aan te wijzen. Zeg bijvoorbeeld:
 - “Snel zoeken!”
 - “Waar is een olifantje?”
 - “Zoeken zoeken zoeken!”
 - “Zie je nog een olifantje?”
- Probeer ervoor te zorgen dat kinderen gedurende de testitems niet weggijken van het scherm. Dit is erg belangrijk voor een nauwkeurige meting, omdat kinderen die weggijken minder tijd hebben waarin ze olifantjes kunnen vinden.
- Na afloop van elk item zie je een scherm met een grote vrolijke olifant. Kijk hiernaar met het kind als leuke onderbreking van de test en als ontspanningsmoment.

Scoring

- Tijdens de eerste drie testitems krijgt het kind een scherm te zien met allemaal dieren verdeeld over 8 kolommen (zie de figuur op de vorige bladzijde). In elke kolom staat één olifantje. Boven elke kolom staat een getal (1 t/m 8). Als het kind een olifantje aanwijst, toets je als testleider het getal in van de kolom waarin het olifantje staat. Dus als het kind bijvoorbeeld het olifantje aanwijst in de meest linker kolom op het scherm, kolom 1, dan toets je als testleider het getal 1 in op de computer. Hierna verschijnt er een streep door het olifantje in kolom 1.
- Bij het vierde testitem zijn er meer kolommen aanwezig. Deze hebben steeds de benaming 1a, 1b, 2a etc. Hierbij geldt dat het getal van de kolom ingevoerd moet worden als het kind in die kolom een olifantje aanwijst. Ook dan verschijnt er een streep door het olifantje in die kolom. De letter (a of b) achter het getal kun je negeren. Ook bij dit testitem zijn er in totaal 8 olifanten.
- Als het kind hetzelfde olifantje meerdere keren aanwijst, toets je elke keer opnieuw het getal in van de kolom waar dit olifantje in staat. Deze gegevens worden opgeslagen als ‘repetitie-fout’.
- Als het kind een ander dier aanwijst (beertje of paardje), toets je ‘0’ in. Deze gegevens worden opgeslagen als ‘gewone fout’.
- LET OP: deze taak meet de aandacht en verwerkingssnelheid van het kind. Het is daarom essentieel dat je als testleider ook zo snel mogelijk en toch nauwkeurig elke reactie van het kind scoort.
- Van elke score die je invoert, wordt de reactietijd automatisch gemeten!
- Na elk testitem krijg je de volgende vraag:

Keek het kind de hele tijd naar het scherm tijdens dit item?

1 = ja

2 = nee, kind keek heel even weg (< 5 sec)

3 = nee, kind keek een tijdje weg (5 - 20 sec)

4 = nee, kind keek meer dan de helft van de tijd weg (> 20 sec)

5 = nee, kind keek helemaal niet naar het scherm

Niet van toepassing, want ik heb dit item niet of maar gedeeltelijk kunnen afnemen door:

6 = storende factoren in de omgeving

7 = taalprobleem

8 = gedrag van kind (bijv. huilen of weglopen)

9 = technisch probleem (bijv. zwart scherm)

- Geef bij deze vraag aan of het kind tijdens het item de hele tijd naar het scherm keek of niet. Als het kind wegkeek tijdens het item, maak je een inschatting van de duur van het wegstijgen. Als het item echter onderbroken werd door iets in de omgeving (bijvoorbeeld de leerkracht kwam opeens binnen) of je kon de taak niet afnemen door een taalprobleem, door gedrag van het kind, of een technisch probleem, dan geef je dit aan bij één van de opties onder 'niet van toepassing'.
- LET OP: kruis alleen optie '8' aan als het kind bijvoorbeeld ging huilen of wegliep van de laptop en het daardoor niet mogelijk was het item (volledig) af te nemen, maar niet als het kind weinig aandacht voor de taak had en daarom veel wegkeek. In dit laatste geval kies je altijd één van de opties 2 t/m 5.
- Vervolgens krijg je de volgende vraag:

Wees het kind duidelijk aan tijdens dit item?

1 = altijd duidelijk

2 = meestal duidelijk

3 = soms duidelijk, soms niet

4 = meestal onduidelijk

5 = altijd onduidelijk

6 = niet van toepassing, want ik heb dit item niet of maar gedeeltelijk kunnen afnemen

- Geef bij deze vraag aan of het kind tijdens het item duidelijk aanwees of niet. (Kies alleen optie '6', niet van toepassing, als je bij de vraag over het kijkgedrag optie '6', '7', '8' of '9' had gekozen).

Wat als...?

[bij de oefenitems]

- Het kind weigert om iets aan te wijzen op het scherm bij de oefenitems?
 - Pak op vriendelijke wijze de hand van het kind en wijs samen een olifantje aan tijdens de oefenitems. Zeg dan 'tikkie!' Let op: dit vergt een goed inschattingsvermogen, met name bij verlegen kinderen. Als testleider moet je dus goed inschatten en aftasten of het kind dit wil toelaten.

- Een kind niet duidelijk aanwijst bij de oefenitems?
 - Wijs tijdens de oefenitems zelf een keer een olifantje aan en zeg *'tikkie!'*. Als het kind daarna zelf een olifantje aanwijst, kun je zeggen *'goed zo, doe maar tikkie!'*
 - Pak op vriendelijke wijze de hand van het kind en wijs samen een olifantje aan tijdens de oefenitems. Zeg ook dan *'tikkie!'* Let op: ook hier geldt dat dit een goed inschattingsvermogen vergt, met name als een kind verlegen is.

[bij de testitems]

- Een kind wegkijkt tijdens een testitem?
 - Doe je uiterste best om zo snel mogelijk de aandacht van het kind terug te brengen naar de taak.
 - Probeer een grove inschatting te maken van hoe lang het kind wegkeek van het testscherm en voer deze informatie na afloop van het item in.
- Een kind niet duidelijk aanwijst bij de testitems?
 - Je kunt het kind aansporen duidelijk te wijzen door *'tikkie!'* te zeggen als een kind een olifantje aanwijst.
 - Zeg *'goed wijzen!'*
 - Let op: bij de testitems mag je het kind niet meer helpen!
- Het kind alle dieren aanwijst bij de testitems, en niet alleen de olifantjes?
 - Scoor elk dier dat het kind aanwijst zoals aangegeven onder 'scoring'.
 - Zeg: *"Nee, je moet alleen de olifantjes zoeken in dit spelletje!"*
 - Spoor het kind zoveel mogelijk aan om de olifantjes aan te wijzen.
- Het kind zo snel de olifantjes aanwijst dat je het niet meer bij kan houden met invoeren van de scores?
 - Als dit gebeurt, is het heel belangrijk dat je het kind niet tegenhoudt bij het zoeken; blijf positieve feedback geven en het kind aanmoedigen verder te gaan met zoeken. Voer zo snel als je kan alsnog alle scores in. Zorg in elk geval dat je alle aangewezen olifanten hebt ingevoerd voordat de 40 seconden voorbij zijn en het testitem vanzelf afloopt. Na afloop van de taak geef je bij de vraag 'is het gelukt om het invoeren van de scores bij te houden?' aan dat dit niet (of niet bij elk item) het geval was.
- Het kind zo snel de olifantjes aanwijst dat het ze allemaal heeft gevonden voordat de 40 seconden voorbij zijn?
 - Zeg: *'Goed zo! Jij hebt ze allemaal al gevonden! Nu wachten we even totdat de grote olifant weer komt.'* Wacht rustig met het kind totdat het item vanzelf afloopt.
- Het kind een olifantje aanwijst op het allerlaatste moment (dus net voordat het testscherm vol dieren verdwijnt), waardoor je de score niet meer in kunt voeren?
 - Je kunt gewoon verder gaan met de taak en hoeft hier niets mee te doen.

5.5 Nonwoordrepetitietaak

Taakbeschrijving

Deze taak meet met name het fonologisch geheugen van een kind, d.w.z., het vermogen om spraakklanken tijdelijk vast te houden in het verbaal kortetermijngeheugen. In de taak wordt aan kinderen gevraagd om nonwoorden na te zeggen, zodat de rol van eerder opgedane taalkennis zo klein mogelijk is. Er wordt een zgn. quasi-universele nonwoordrepetitietaak gebruikt die voor kinderen met verschillende moedertalen, en ook tweetalige kinderen, kan worden gebruikt.

Procedure

Alle instructies zijn opgenomen en worden afgespeeld als je de powerpoint opent. Als testleider hoef je alleen op de muis of het rechterpijltje te drukken om naar de volgende slide te gaan. Doe dit pas als je zeker weet dat het kind de instructies heeft begrepen.

De twee oefenitems mogen zo vaak herhaald worden totdat het kind de woordjes goed nazegt. Een woord herhalen doe je door twee keer op het linker pijltje te drukken en daarna weer op het rechterpijltje.

Bij de echte test mag dit NIET! Een kind mag het woordje maar één keer horen. Als je het idee hebt dat dit niet duidelijk is (omdat het kind bijvoorbeeld vraagt of zij/hij het woord nog een keer mag horen), leg dit dan uit aan het kind.

Tussen elk item staat een lege slide met alleen het buitenaards wezen. Deze slide staat erin zodat je hier kan checken of het kind echt goed luistert. Het item mag namelijk niet een tweede keer herhaald worden. Zorg dus na elk item dat je checkt of het kind nog aandachtig is. Mocht dit niet het geval zijn, stimuleer het kind dan om nog even door te gaan.

De items in de taak lopen op in lengte en moeilijkheid. De kinderen beginnen met nonwoorden met twee lettergrepen en eindigen met nonwoorden met 5 lettergrepen. Per lengte moet het kind steeds vier nonwoorden nazeggen (16 items in totaal dus). Als je merkt dat een kind wat gefrustreerd raakt na een moeilijk item, mag je aangeven dat sommige woordjes moeilijk zijn en sommige makkelijker omdat ook oudere kinderen de taal van het buitenaards wezentje moeten leren. Zeg daarna wel dat het kind zo goed mogelijk zijn/haar best moet doen.

Het kan voorkomen dat een kind helemaal niks zegt (nonrespons). Dit moet je zoveel mogelijk proberen te voorkomen. Probeer altijd een uiting te ontlokken, al is het maar een klein stukje van het woord.

Procedure per slide:

1. Lege slide met planeet. Als het kind er goed klaar voor zit => klik verder.
2. Buitenaards wezen komt binnenvliegen met muziekje en zegt daarna *"Hoi! Ik ben een buitenaards wezen en ik woon op de planeet Wikela. Waar ik vandaan kom, spreken we een andere taal dan jij. Ik wil je nu wat woordjes in mijn taal leren. Luister goed en herhaal elk woord nadat je het hebt gehoord. Als ik bijvoorbeeld zeg: 'loen', dan zeg jij, 'loen'. Ben je er klaar voor?"* => klik verder.
3. *"Loen"* => als het kind het goed herhaalt, klik verder. Als het kind het niet goed herhaalt, klik twee keer op het linker pijltje en vervolgens op het rechterpijltje om het item nog

- eens te laten horen. Eventueel kan je het zelf nog een keer nazeggen en vragen of het kind het daarna kan nazeggen.
4. *"Heel goed! En nu nog een keer."* => zorg dat het kind goed oplet en klik verder.
 5. *"Peik"* => als het kind het goed herhaalt, klik verder. Als het kind het niet goed herhaalt, klik twee keer op het linker pijltje en vervolgens op het rechterpijltje om het item nog eens te laten horen. Eventueel kan je het zelf nog een keer nazeggen en vragen of het kind het daarna kan nazeggen.
 6. *"Heel goed! ! Je hoort zo nog meer woordjes uit mijn taal. Herhaal elk woord nadat je het hebt gehoord en wacht daarna op het volgende woord. Ben je er klaar voor om mijn taal te leren?"*. => kijk of het kind er klaar voor is en klik 2x verder.
 7. Nu komen er een aantal slides met items die het kind maar één keer mag horen. Ze lopen op in lengte en moeilijkheid. Stimuleer het kind dat het goed zijn/haar best moet doen. Als je frustratie ziet, mag je aangeven dat je het een heel moeilijke taal vindt of dat ook oudere kinderen deze taal moeten leren dus dat het niet erg is als het moeilijk is. Als het kind weer goed oplet, ga je verder (zeg bijvoorbeeld: 'oke, we gaan er weer voor, heel goed luisteren').
 8. Na de set items zegt het wezentje: *"Wat leer jij mijn taal snel!"* => complimenteer het kind en klik verder.
 9. Aan het eind hoort het kind: *"Dat heb je heel goed gedaan! We vonden het leuk om je te leren kennen, maar moeten er nu weer vandoor. Tot ziens!"*. De buitenaardse wezentjes vliegen weer weg en je kan de powerpoint afsluiten.

Scoring

- Deze taak wordt met de hand gescoord op een scoreformulier. Noteer voor elk item of het kind het correct heeft nagezegd (scoor een 1), of het kind het nonword incorrect heeft nagezegd (scoor een 0), of het nonwoord niet heeft nagezegd (scoor 999) of dat je de respons van het kind niet kon scoren als correct of incorrect [NA].
- Je scoort een 1 (correct) als het kind alle klanken van het item in de goede volgorde nazegt. Dus als het kind geen klanken weglaat, klanken vervangt door andere klanken, of nieuwe klanken toevoegt aan het item. Als het kind een of meer van deze drie dingen wel doet, scoor je een 0 (incorrect). Ook als een kind maar een deel van het nonwoord nazegt, scoor je een 0 (incorrect).
- N.B. Intonatie en klemtoon hoeven niet geïmiteerd te worden door het kind. Het gaat alleen om de afzonderlijke klanken (/p/, /b/, etc.) van het nonwoord!
- Probeer om zo aandachtig naar het kind te luisteren en zo nauwkeurig mogelijk te scoren! De scores zullen later gecontroleerd worden op basis van geluidsopnamen.

Wat als...?

- Kinderen gefrustreerd raken door de moeilijkheidsgraad van de items?
 - o Als kinderen het moeilijk vinden, kun je zeggen dat sommige woorden gemakkelijk zijn en andere moeilijk omdat oudere kinderen de taal van de alien ook moeten leren. Vertel kinderen ook dat ze gewoon zo goed mogelijk moeten proberen om de gekke woorden na te zeggen.

- Kinderen niets nazeggen?
 - o Probeer om toch een antwoord te ontlokken, al is het maar een deel van een item.
 - o Als kinderen fluisteren probeer dan toch te horen wat ze zeggen en dit te scoren. Als het echt niet lukt kan je het kind vragen om harder te praten. Het is wel mogelijk dat verlegen kinderen zullen dan niets meer zeggen.
 - o Als kinderen niets zeggen bij vier opeenvolgende items, ook niet nadat je ze flink hebt aangemoedigd, mag je de taak afbreken. Probeer dit echter te vermijden door de kinderen aan te moedigen de woorden na te zeggen en/of extra te oefenen met de oefenitems!

5.6 Perceptievragenlijst

Deze vragenlijst wordt gebruikt om te meten hoe de kinderen denken over de robot (zie de appendix voor de vragenlijst). De vragenlijst neem je als volgt af:

- Pak de vragenlijst.
- Vul het proefpersoonnummer, de datum, en jouw ID in.
- Lees de instructie die bovenaan staat letterlijk voor.
- Lees de vragen één voor één letterlijk door en omcirkel het antwoord dat het kind geeft.
- Bij alle vragen (behalve vraag 2 t/m 5) vraag je door op het antwoord van het kind. Als het kind 'ja' antwoordt, vraag je 'waarom?', en als het kind 'nee' antwoordt vraag je 'waarom niet?'. Probeer het antwoord dat het kind geeft zo goed mogelijk te noteren, zodat je alleen de videobeelden hoeft te bekijken als het echt niet lukte om het antwoord te noteren.
- Reageer neutraal op ieder antwoord dat het kind geeft.
- Bij vraag 1 vraag je of het kind Robin de robot meer een jongetje of een meisje vindt. Als het kind 'jongetje' antwoordt, zeg je bij vraag 2 t/m 4 'meester' en 'vriendje'. Als het kind 'meisje' antwoordt, zeg je bij vraag 2 t/m 4 'juf' en 'vriendinnetje'.

Let op:

- Zeg tijdens het afnemen van deze vragenlijst altijd 'Robin de robot'. Niet 'Robin' of 'de robot', want dan stuur je het kind in de richting van een mens of een machine.
- Zeg tijdens het afnemen van deze vragenlijst niet 'hij' of 'zij' als je het over de robot hebt, maar zeg altijd 'Robin de robot'.
- Probeer altijd een ja/nee-antwoord te krijgen van het kind, en gebruik de 'weet-niet'-optie pas als laatste middel. Als een kind meteen zegt dat hij/zij het niet weet, moedig het kind dan aan om te antwoorden en herhaal de vraag. Als het kind na twee pogingen het nog steeds niet weet, omcirkel je 'weet niet' en ga je door naar de volgende vraag.
- Als een kind aan jou vraagt "Wat denk jij?", zeg dan "Dat weet ik niet, wat denk jij?". Geef in ieder geval nooit je eigen mening!

6. Post-tests

Er zijn twee post-tests in de studie. Het doel van deze post-tests is om te bekijken of kinderen de doelwoorden die tijdens de training zijn aangeboden hebben geleerd. Kennis van de woordvormen wordt getoetst met twee vertaaltaken. Kennis van de woordbetekenissen wordt getoetst met een begripstaak. De volgorde van deze taken is als volgt:

1. Vertaaltaak T2 > T1
2. Vertaaltaak T1 > T2
3. Begripstaak
4. In post-test 1: perceptievragenlijst

Na deze taken zal in post-test 1 een korte vragenlijst aan het kind worden voorgelegd.

6.1. Algemene instructies post-test 1

Voordat je post-test 1 afneemt, geef je de volgende instructie aan het kind:

“De afgelopen weken heb je Engelse woordjes geleerd. Vandaag gaan we kijken of je deze woorden nog weet. We gaan een paar spelletjes doen. Na elk spelletje mag je een sticker op het sticker-konijn plakken”.

Neem hierna de hierboven genoemde drie taken in bovenstaande volgorde af. Vergeet niet om na elke taak een sticker te geven. Na de laatste taak zeg je dat ze het supergoed hebben gedaan (ongeacht hoe goed het kind het echt deed) en bedank je ze dat ze hebben meegedaan.

6.2 Vertaaltaak T2 > T1

je hebt hier 2 versies van Let op dat het goede pakt. Als op het scoreformulier bovenop de taak staat lijst 1 dan moet je versie 1 pakken, als er lijst 2 staat moet je versie 2 pakken.

Het doel van deze taak is om te bekijken of kinderen de Nederlandse vertaling van de Engelse woorden hebben geleerd. De test is precies hetzelfde als de pre-test, behalve de tekst die ervoor gezegd moet worden.

Procedure

- Lees de instructies in Appendix B letterlijk voor aan het kind. Kies dan de goede lijst, afhankelijk van de sessie (zie Tabel 3).
- Scoor elk antwoord van het kind meteen door een 1 in de juiste kolom op het formulier te zetten. Let erop dat je elk antwoord scoort (en niet bijv. alleen de goede antwoorden), zodat kinderen niet kunnen achterhalen of ze de antwoorden goed of fout hebben, doordat je soms wel en soms niet iets opschrijft.
- Je mag het woord nog 2 keer herhalen (dus maximaal 3 keer afspelen) als het kind daarom vraagt, als je denkt dat het het niet goed heeft kunnen horen.
- Feedback:

- Geef bij elk antwoord positief-neutrale feedback aan het kind (“ja”, “goed zo”, “jij doet goed je best!”) op positieve toon. Wees echter niet te enthousiast; dit kan ervoor zorgen dat kinderen zich ongemakkelijk voelen als ze het antwoord niet weten. Bovendien zou het kunnen voorkomen dat kinderen in de gaten hebben dat ze ook complimenten krijgen voor verkeerde antwoorden en expres verkeerde antwoorden gaan geven.
- Bij kinderen die de antwoorden niet weten: geef algemene positieve feedback ergens halverwege de taak en aan het eind: “Ik zie dat jij heel erg goed je best doet! Super!”, “Ik zie dat je goed zit na te denken, hoor. Heel knap!”

6.3 Vertaaltaak T1 > T2

je hebt hier 2 versies van Let op dat het goede pakt. Als op het scoreformulier bovenop de taak staat lijst 1 dan moet je versie 1 pakken, als er lijst 2 staat moet je versie 2 pakken.

Het doel van deze taak is om te zien of kinderen de Engelse vertalingen hebben geleerd van de doelwoorden. De procedure is precies hetzelfde als bij de vertaaltaak T2 > T1, behalve dat de woorden andersom vertaald moeten worden. Zie hierboven, in sectie 6.1, voor de procedure.

6.4 Begripstaak

je hebt hier 2 versies van Let op dat het goede pakt. Als op het scoreformulier bovenop de taak staat lijst 1 dan moet je versie 1 pakken, als er lijst 2 staat moet je versie 2 pakken. De taak is verdeeld in twee powerpoint presentaties, omdat het anders te groot is. Let op dat je beide delen afneemt.

Deze taak meet de receptieve kennis van doelwoorden die geleerd zijn tijdens de lessen. In totaal worden 18 doelwoorden getest. Elk doelwoord komt drie keer voor, dus er zijn 54 items in totaal.

In deze taak krijgen kinderen steeds drie plaatjes of drie korte filmpjes naast elkaar te zien. Bij elk item wordt de vraag gesteld “Waar zie je: X?”. De taak draait op powerpoint en de vragen zijn opgenomen op de computer.

De eerste twee items zijn oefenitems. Deze moet je ook scoren op het formulier. Gebruik deze items om het kind de taak te leren. Deze mag je herhalen zo vaak als nodig. Je mag het kind er ook bij helpen. Ongeacht of deze items wel of niet goed zijn gedaan ga je door met de testitems. Bij de testitems mag je het kind uiteraard niet helpen. De films worden meteen afgespeeld als de dia in beeld komt. Let op: wacht altijd even totdat ze allemaal zijn begonnen en het kind de kans heeft gehad om ze allemaal te zien (ze draaien in een loop). Speel dan pas de audio af.

Bij sommige laptops worden de films niet goed tegelijkertijd afgespeeld (meestal wat oudere laptops of als er veel op staat). In dat geval moet je zelf met de muis op de films klikken om ze op te starten. Wacht ook hier goed totdat alle films afspelen voordat je de audio aanzet.

Op het scoreformulier noteer je het nummer van het plaatje dat het kind heeft aangewezen.

- Laat het kind altijd zelf een plaatje of video aanwijzen, d.w.z. help het niet bij de

keuze!

- Als je de indruk hebt dat het kind het niet goed heeft verstaan, herhaal dan de vraag.
- Herhaal de vraag maximaal twee keer (dus drie keer vragen in totaal). Markeer GA ('geen antwoord') in als er na drie keer vragen geen antwoord van het kind is gekomen.
- Laat het kind rustig zoeken: sommige kinderen hebben snel een overzicht van de plaatjes, maar bij andere duurt het wat langer.
- Geef positieve feedback als het kind een plaatje heeft aangewezen. Zeg dingen als 'Ja!', 'Goed zo!', 'Dat kan jij goed!' etc.
- Geef net zo vaak positieve feedback bij goede en foute antwoorden zodat het kind het niet merkt als hij/zij fouten maakt.
- Varieer af en toe in intonatie en toonhoogte bij het stellen van de vragen zodat het niet eentonig wordt.

Wattedoenals...?

- Een kind niets aanwijst in het testgedeelte?
Stel de vraag nog een keer – bij geen antwoord – nog een keer
Als het kind dan nog niets aanwijst, scoor je GA – geen antwoord
- Een kind de andere plaatjes wil benoemen?
Laat het kind de plaatjes op zijn gemak bekijken en benoemen
Creëer gerust wat ruimte om naar de andere plaatjes te kijken en daarover te praten als je het gevoel hebt dat het kind de taak daardoor leuker vindt om te doen en het langer volhoudt.
- Een kind het 'doelplaatje' benoemt voor je de vraag hebt kunnen stellen?
 - Stel de vraag nog een keer en scoor dan pas het antwoord.
- Een kind eerst het verkeerde plaatje aanwijst en daarna het goede?
 - Scoor altijd het laatste antwoord: in dit geval dus 'goed'.
- Een kind eerst het goede plaatje aanwijst en daarna het verkeerde?
 - Scoor altijd het laatste antwoord: in dit geval dus 'fout'.

6.5 Perceptievragenlijst

Deze vragenlijst is identiek aan de lijst die we in de pre-test gebruiken. Zie instructies onder 5.6.

6.6 Post-test 2

Deze sessie is precies hetzelfde als de sessie van post-test 1. De algemene instructie verschilt wel iets en is als volgt:

"Weet je nog dat je een tijdje terug allemaal Engelse woordjes hebt geleerd? Vandaag gaan we bekijken of je ze nog weet. We doen dezelfde spelletjes als de vorige keer. Weet je die nog? Net als de vorige keer mag je na elk spelletje een sticker uitzoeken. Je mag die dan op de sticker-aap plakken".

Doe hierna de taken van de post-test in de vaste volgorde die hierboven staat aangegeven. Vergeet niet om na elke taak een sticker te geven. Na de laatste taak, geef je de kinderen een cadeautje en bedank je ze heel hartelijk dat ze zo goed hun best hebben gedaan!

6.7 Aandacht punten voor tijdens het testen (post-test)

Algemeen:

- Leg de laptop aan de rand van de tafel en ga zitten naast het kind in een 90° hoek. Zodat je goed mee kan kijken maar ook kan scoren een beetje uit het zicht van het kind.
- Tafel verder zo leeg mogelijk houden.
- Bij het scoren - altijd wat schrijven om niet de indruk te geven dat we het schrijven alleen als ze het goed of slecht hebben. Probeer dit onzichtbaar voor ze te maken (lijnen en punten bijvoorbeeld in plaats van krulletjes en x'jes of zo iets).
- Praten over spelletjes of taakjes maar niet over testjes

Begrip taak

- oefen items aan het begin – let op dat het kind de taak begrijpt zo niet kan je deze herhalen.
- Veel positieve feedback geven om kinderen te motiveren om door te gaan. Alleen we willen niet suggereren dat ze het goed hebben als het niet zo is. Dus probeer een beetje neutrale inhoud maar wel positief te kiezen (je doet je best, knap hoor enz.).
- Let op dat je even goed positief blijft ook bij kinderen die de taak niet goed doen.
- Kinderen hoeven niet altijd een antwoord te geven, ze mogen zeggen ik weet het niet. Ze mogen ook gokken. Het verschilt nogal tussen kinderen wat ze doen.

Vertaal taak:

- Houd je meer aan de tekst zoals die staat in het boekje – staat ook op je scoreformulier
- Zeg (bij de eerste items – zolang dat nodig is:
- Wat betekent...[geluid uit de computer] in het NL (of in het Engels voor de NL-ENG taak)
- Leg dit nog een keer uit als het kind het niet begrijpt.
- Geef complimenten als kinderen het proberen – Ja, je doet goed je best! (maar uiteraard niet goed zo of zo iets zeggen voor woorden die het kind niet goed heeft).
- Als je een woord moet herhalen, spel het nog een keer af, herhaal de woorden niet zelf.

7. Proefpersoonwerving, dataopslag en andere belangrijke informatie voor testleiders

7.1 Werving en informed consent

Drie weken van te voren:

- Alle scholen moeten op tijd (d.w.z., drie weken voor begin van de studie) informatiebrieven en informed consent formulieren krijgen. Stem met de school van tevoren af of ze de brieven digitaal of op papier willen ontvangen. Informed consent formulieren moeten altijd op papier. Vraag de scholen om de informatiebrieven en informed consent formulieren aan de ouders uit te delen en als het kan nog een keer actief naar te vragen en ouders hieraan te herinneren.

Een week voor de start van de studie:

- Neem contact op met de school om te horen hoeveel kinderen er mee gaan doen. Herinner de leerkrachten er zo nodig aan om ouders nog een keer te herinneren aan de studie zodat zij zich kunnen opgeven.
- Ga op bezoek bij de school (of overleg met ons wie er naartoe gaat). Plan een afspraak met de kleuterleerkracht. Punten om door te nemen tijdens het bezoek:
 - Ruimte bekijken om te zien of deze geschikt is, informeren of deze beschikbaar is voor de hele periode
 - Wanneer mag er wel en niet getest worden met de kinderen?
 - Rooster
 - Ouderbrieven en informed consent procedure – hoe willen de scholen de informatie voor ouders verspreiden en kunnen ze ons helpen om ouders hieraan te herinneren?
 - Waar kunnen we de robot achterlaten op school?
 - Wanneer hebben ze de CITO toetsen?
 - Al onze medewerkers hebben een VOG.

Op de dag van de introductie:

- Vraag aan de leerkracht alle getekende formulieren terug. Bekijk welke kinderen er meedoen. Als ouders na de introductie alsnog toestemming willen geven kan dat nog (introductie wordt gedaan met de hele klas).

Op elke school testen we 18 kinderen, maar om eventuele uitval te vangen willen we beginnen met meer kinderen. Probeer in iedere school minimaal 22 kinderen te testen: 6 kinderen voor per testconditie en 4 voor de controlegroep.

De ondertekende formulieren moeten door de onderzoeker worden getekend en gekopieerd. De kopieën geef je via de leerkracht aan de ouders terug. Alle getekende formulieren moet je goed bewaren en het proefpersoon nummer erop noteren (na het kopiëren). Maak er foto's van en upload ze op het Yoda systeem (zie uitleg aan het eind van dit document). Lever de papieren formulieren in samen met de scoreformulieren bij Mirjam/Jan (Tilburg) en Ora/Rianne (Utrecht). Noteer in het logboek bij elk deelnemer of we wel of geen toestemming hebben om beelden te gebruiken voor onderwijsdoeleinden.

Sorteer de lijst van de kinderen die meedoen binnen een school op alfabetische volgorde. Vervolgens pak je uit het Excelbestand proefpersoonnummers de lijst onder het tabblad van jouw school. Het eerste nummer in de lijst geeft je aan het eerste kind in je lijst en zo verder. De nummers horen al bij een bepaalde conditie. Er staat ook in dit bestand welke versie van de pre- en post-test het kind moet doen. Na de pre-test gaat Ora de resultaten van de pretest bekijken om eventueel verdeling in condities bij te sturen waar nodig.

Ieder kind krijgt een proefpersoonnummer. Houd goed bij hoeveel deelnemers je hebt in de verschillende fases en welke data binnen is. Dit doe je op papier en digitaal (zie voor verdere informatie de sectie over dataopslag).

7.2 Proefpersoonnummers

Ieder kind krijgt een uniek nummer dat zijn proefpersoonnummer wordt. Dit nummer heeft 6 cijfers. Het eerste cijfer staat voor de conditie waar het kind toe behoort:

- 1 = controle
- 2 = alleen tablet
- 3 = robot zonder iconische gebaren
- 4 = robot met iconische gebaren.

De volgende twee cijfers staan voor de school (loopt van 01 tot 10) en de laatste 3 cijfers verwijzen naar het kind zelf. Dit is een uniek getal tussen 000 en 250.

Voorbeeld: kind nummer 003, uit school 04 dat meedoet in de tabletconditie krijgt proefpersoonnummer 204003.

Je krijgt van tevoren te horen wat het nummer is van de school waar jij naartoe gaat en welke unieke kindnummers je kunt gebruiken in deze school. Het is heel belangrijk om deze nummers goed in te voeren in het logboek en in het systeem van de robot. Aan het begin van iedere les, film je een blad met het proefpersoonnummer en datum erop (maak een stapel met de juiste nummer van tevoren klaar. Zo kunnen we altijd achterhalen welk kind op welke film staat. Als iedereen zich hier strikt aan houdt voorkomen we fouten in de data die veel tijd kosten om te herstellen of zelfs helemaal niet meer zijn op te lossen.

7.3 Oudervragenlijst

Deze lijst wordt gestuurd naar ouders via de mail en ze kunnen hem online invullen. Zodra je de instemmingsverklaringen binnen hebt, noteer je alle mailadressen van de ouders in een excel-bestand met de naam van hun kind en het proefpersoonnummer erbij. Dit bestand moet je zo snel mogelijk uploaden in Sharepoint (in ieder geval voor het einde van de week van de pre-test). De link naar de oudervragenlijst zal dan verstuurd worden naar deze mailadressen.

Als ouders na een herinneringsmail de vragenlijst nog niet hebben ingevuld, vragen we je om op school na te gaan of je de leerkracht of de ouders hieraan kunt herinneren.

7.4 Dataopslag

Alle data- and videobestanden moeten worden geupload op YODA, in de daarvoor bestemde map (zie aparte informatie over YODA in appendix VIII and op <https://yoda.sites.uu.nl>). Logboeken worden geupload op Sharepoint. Hier krijgt iedereen een link met toegang tot de relevante map.

Alle papieren formulieren, zoals logboekinformatie en scoreformulieren, moeten ingevoerd worden in de daarvoor bestemde digitale bestanden op de dag waarop de desbetreffende kinderen zijn getest!

Naamgeving van bestanden - in elke bestand moet altijd het proefpersoonnummer van het kind staan. Namen van bestanden hebben altijd het volgende formaat:

- Pre-test:
 - Excel bestand met ingevoerde data per school op basis van schoolnummer: Schoolnummer_school_pretest.extensie (uploaden op Yoda)
 - Aandachtstaakbestanden worden in de map van de taak op de laptop opgeslagen. Houd de naam aan die E-prime aan de bestanden geeft en zet ze bij elkaar in de daarvoor bestemde map op Yoda.
 - Films per kind: proefpersoonnummer_pretest.exetensie (uploaden op Yoda)
- Post-test 1:
 - Excel bestand met ingevoerde data per school op basis van school nummer: Schoolnummer_school_posttest1.extensie (opslaan op Yoda)
 - Films per kind: proefpersoonnummer_posttest1.exetensie (uploaden op Yoda)
- Post-test 2:
 - Excel bestand met ingevoerde data per school op basis van school nummer: Schoolnummer_school_posttest2.extensie (uplodedn op Yoda)
 - Films per kind: proefpersoonnummer_posttest2.exetensie (uploaden op Yoda)
- Lessen:
 - Logbestanden van de lessen: proefpersoonnummer_les#.extensie (uploaden op Yoda)
 - Films per kind: proefpersoonnummer_les_[nummer les].exetensie (uploaden op Yoda)
- Lobgoeken:

- Schoolnummer_school_logbook.extensie (houden op sharepoint en daarop updaten)

7.5 Logboek

Vul de informatie meteen na het testen in in het papieren logboek! Vul aan het eind van de dag als je klaar bent met testen deze informatie in in het digitale logboek. Dit logboek wordt op Sharepoint opgeslagen, update het ook op Sharepoint zodat iedereen hier gelijk toegang toe heeft. Bewaar het papieren logboek en lever het in. Bij het inleveren maak je een scan van de logboeken en alle andere papieren formulieren. Spreek met het lokale team af waar je deze scans opslaat.

Houd het logboek er tijdens iedere sessie (testen of lessen) bij en ook tijdens de introductie. Bij de introductie noteer je welke vragen hebben kinderen gesteld en wat er in de klas over Robin de robot is besproken.

7.6 Een sessie afbreken of uitstellen

Het is heel belangrijk om kinderen zoveel mogelijk te enthousiasmeren tijdens de lessen en tests.

Het kan echter een enkele keer voorkomen dat:

- Kinderen verbaal of non-verbaal heel duidelijk aangeven dat ze willen stoppen en dat het niet lukt om ze opnieuw te motiveren, zelfs niet na herhaaldelijke pogingen om ze weer mee te laten doen. Als het na drie nadrukkelijke pogingen nog niet is gelukt om de aandacht van een kind weer terug te krijgen bij de taak, mag je een sessie afbreken. Noteer in zo'n geval goed in het logboek waarom je de sessie hebt afgebroken.
- Kinderen herhaaldelijk weigeren om een opdracht uit te voeren, expres verkeerde antwoorden geven of op een andere manier niet coöperatief zijn, zodanig dat de testleider inschat dat de data die bij het kind verzameld worden niet betrouwbaar zijn. In dit geval mag je de sessie afbreken. Noteer goed waarom je de sessie hebt afgebroken.

Als een sessie is afgebroken, moeten de niet afgenomen lessen of tests op een later moment worden afgenomen. Doe dit zo snel mogelijk, bij voorkeur in dezelfde week! Als dit niet kan door ziekte of doordat kinderen opnieuw niet meewerken, dan kan de sessie opnieuw uitgesteld worden. Maar probeer dit soort uitstel zoveel mogelijk te voorkomen!

- **Lessen:** als kinderen lessen missen door bijvoorbeeld ziekte probeer deze les (of lessen) dan nog in dezelfde week of de week erna in te halen (bijvoorbeeld door 3 i.p.v. 2 lessen in de week erna te doen). Aan het einde van de lesperiode is er eerder ruimte voor wat inhaallessen als kinderen een eerdere sessie niet konden doen. Bij voorkeur zit er minimaal 1 dag tussen twee lessen maar als dat niet anders kan dan kunnen twee lessen op aansluitende dagen gedaan worden. We doen nooit twee lessen op 1 dag. Als kinderen drie of meer lessen missen dan stoppen ze met het onderzoek.

- **Post-test 1:** post-test 1 kan plaatsvinden vanaf de eerste dag na de recaples (mag niet op dezelfde dag als de recap les) tot maximaal 3 dagen erna. Er zitten dus maximaal 2 dagen tussen de recaples en post-test 1). Als het niet lukt (er zitten meer dan 2 dagen tussen), post-test 1 zo snel mogelijk gedaan worden. Deze test mag nooit overgeslagen worden!
- **Post-test 2:** post-test 2 is bij voorkeur 4 tot 5 weken na de recap les, maar als dat niet lukt, probeer de uitloop dan zo kort mogelijk te houden! Als het nodig is, ga dan terug naar de school voor het testen.

Let op! Het is belangrijk om goed te plannen om aan deze regels te kunnen voldoen. Probeer met name de tijd tussen de laatste les ("recap") en post-test 1 beperkt te houden. Dit laatste punt houdt in dat je bij voorkeur geen recap les plant op vrijdag omdat er dan een weekend tussen de recap en post-test 1 in valt.

7.7. Video-opnames

Alle trainings- en testsessies worden opgenomen op video. Er worden twee camera's gebruikt met een statief.

Zie Figuur 1 op voor een schematische weergave van waar de camera's neergezet moeten worden.

8. Trainingsprocedure testleiders

Er is een strikte trainingsprocedure waarin testleiders getraind worden op de taakafname en hiermee oefenen voordat ze beginnen met de dataverzameling. Deze procedure betreft een aantal stappen:

1. Testleiders worden geïnformeerd over een datum waarop de training zal plaatsvinden. Zij ontvangen dan meteen dit protocol en worden gevraagd om dit goed door te lezen. Ook wordt hen gevraagd om een afspraak te maken met een kind van 5 jaar ca. één week na de training om de taken mee te oefenen.
2. Tijdens een trainingsdag worden alle taken doorgenomen en kunnen de testleiders vragen stellen over het materiaal en de taakafname.
3. In de week na de training oefenen de testleiders met het materiaal. Ze lopen de tests door en oefenen met een andere volwassene. Als er niet genoeg test sets zijn, moeten ze sets delen.
4. Als zij vertrouwd zijn met het materiaal, oefenen zij met het kind met wie een afspraak is gemaakt. Zij nemen bij dit kind alle taken af in de juiste volgorde, en volgens dit protocol. Deze oefen-afname wordt gefilmd. De video-opname moet van goede audio-kwaliteit zijn zowel het kind als de testleider en het laptop scherm moeten hierop zichtbaar zijn. De opnames worden in de hiervoor bestemde map op YODA geüpload.
5. De coördinatoren bekijken de video en schrijven hierover een gedetailleerd feedbackverslag waarin voor elke taak apart wordt aangegeven op welke zaken de testleider nog moet letten.
6. In een telefonische afspraak neemt een van de coördinatoren samen met de testleider dit feedbackrapport door.
7. Als de afname over het algemeen goed ging en de coördinatoren denken dat de testleiders klaar zijn om de taken af te nemen, krijgen zij 'groen licht' om te starten met de dataverzameling.
8. Als de afname op een aantal punten niet volgens het protocol is, wordt testleiders gevraagd om een tweede video op te sturen van een tweede oefen-afname met een ander kind van 5 jaar. Deze tweede video wordt opnieuw bekeken, becommentarieerd middels een uitgebreid feedbackverslag en telefonisch doorgenomen met de testleider. Als deze tweede video voldoet, mag de testleider alsnog aan de slag.

Appendix I.

Pre-test doelwoorden

Lees de volgende instructie letterlijk voor:

De komende weken mag je spelletjes doen met [een tablet/Robin] waarin je Engelse woorden gaat leren! Maar, misschien ken je al wat Engels. Laten we eens kijken of je al een paar Engelse woorden kent. Ik hoop dat je ze nog niet kent, want dan kun je ze gaan leren met [de tablet/Robin]! Dus, denk maar goed na of je deze woorden kent. Het geeft helemaal niets als je ze niet kent.

De computer gaat een woord in het Engels zeggen en jij mag zeggen of je weet wat het woord in het Nederlands betekent.

Hier komen ze!

Voor de controle groep: Als je groter bent op school ga je Engelse woorden leren! Maar misschien ken je al wat Engels. Laten we eens kijken of je al een paar Engelse woorden kent. Ik hoop dat je ze nog niet kent, want dan kun je ze gaan leren op school! Dus, denk maar goed na of je deze woorden kent. Het geeft helemaal niets als je ze niet kent.

De computer gaat een woord in het Engels zeggen en jij mag zeggen of je weet wat het woord in het Nederlands betekent.

Hier komen ze!

Kies nu de juiste lijst en scoor elk item direct:

Lijst 1:

	Correct	Incorrect (noteer het antwoord)	Geen antwoord
1. climb	☐	☐	☐
2. flying	☐	☐	☐
3. most	☐	☐	☐

4. jumping	☐	☐	☐
5. behind	☐	☐	☐
6. sliding	☐	☐	☐
7. fewer	☐	☐	☐
8. light	☐	☐	☐
9. two	☐	☐	☐
10. more	☐	☐	☐
11. five	☐	☐	☐
12. below	☐	☐	☐
13. above	☐	☐	☐
14. low	☐	☐	☐
15. add	☐	☐	☐
16. big	☐	☐	☐
17. falling	☐	☐	☐
18. running	☐	☐	☐
19. small	☐	☐	☐
20. one	☐	☐	☐

21. heavy	☐	☐	☐
22. right	☐	☐	☐
23. catching	☐	☐	☐
24. three	☐	☐	☐
25. walking	☐	☐	☐
26. in front of	☐	☐	☐
27. on	☐	☐	☐
28. throwing	☐	☐	☐
29. fewest	☐	☐	☐
30. take away	☐	☐	☐
31. left	☐	☐	☐
32. next to	☐	☐	☐
33. four	☐	☐	☐
34. high	☐	☐	☐
Totaal aantal goed			

Lijst 2:

	Correct	Incorrect (noteer het antwoord)	Geen antwoord
1. high	☐	☐	☐
2. four	☐	☐	☐
3. next to	☐	☐	☐
4. left	☐	☐	☐
5. take away	☐	☐	☐
6. fewest	☐	☐	☐
7. throwing	☐	☐	☐
8. on	☐	☐	☐
9. in front of	☐	☐	☐
10. walking	☐	☐	☐
11. three	☐	☐	☐
12. catching	☐	☐	☐
13. right	☐	☐	☐
14. heavy	☐	☐	☐
15. one	☐	☐	☐

16. small	☐	☐	☐
17. running	☐	☐	☐
18. falling	☐	☐	☐
19. big	☐	☐	☐
20. add	☐	☐	☐
21. low	☐	☐	☐
22. above	☐	☐	☐
23. below	☐	☐	☐
24. five	☐	☐	☐
25. more	☐	☐	☐
26. two	☐	☐	☐
27. light	☐	☐	☐
28. fewer	☐	☐	☐
29. sliding	☐	☐	☐
30. behind	☐	☐	☐
31. jumping	☐	☐	☐
32. most	☐	☐	☐



33. flying	☐	☐	☐
34. climb	☐	☐	☐
<i>Totaal aantal goed</i>			

Appendix II.

Vertaaltaak T2 > T1

Lees onderstaande instructie letterlijk voor:

De afgelopen weken heb je spelletjes gedaan [op een tablet/samen met Robin] waarin je Engelse woorden hebt geleerd. Laten we eens kijken welke je nog weet. We gaan nu een vraag-antwoord spelletje doen. Ik zeg steeds een woord in het Engels en dan mag jij vertellen wat dit woord betekent in het Nederlands.

Daar gaan we!

Controle groep: Weet nog dat [ik /naam van onderzoeker van pretest] je heb verteld dat je als je groter bent Engelse woorden gaat leren op school? Maar misschien ken je al wat Engels. Laten we weer eens kijken of je al een paar Engelse woorden kent. Ik hoop dat je ze nog niet kent, want dan kun je ze later gaan leren op school! Dus, denk maar goed na of je deze woorden kent. Het geeft helemaal niets als je ze niet kent.

Kies de juiste lijst hieronder en scoor elk item direct:

Lijst 1:

	Correct	Incorrect (noteer het antwoord)	Geen antwoord
1. Wat betekent "low" in het Nederlands?	☐	☐	☐
2. Wat betekent "four"?	☐	☐	☐
3. Wat betekent "add"?	☐	☐	☐
4. Wat betekent "behind"?	☐	☐	☐
5. Wat betekent "climbing"?	☐	☐	☐

6.	Wat betekent “heavy”?	☐	☐	☐
7.	Wat betekent “jumping”?	☐	☐	☐
8.	Wat betekent “two”?	☐	☐	☐
9.	Wat betekent “next to”?	☐	☐	☐
10.	Wat betekent “catching”?	☐	☐	☐
11.	Wat betekent “on”?	☐	☐	☐
12.	Wat betekent “take away”?	☐	☐	☐
13.	Wat betekent “small”?	☐	☐	☐
14.	Wat betekent “most”?	☐	☐	☐
15.	Wat betekent “in front of”?	☐	☐	☐
16.	Wat betekent “fewest”?	☐	☐	☐
17.	Wat betekent “below”?	☐	☐	☐
18.	Wat betekent “sliding”?	☐	☐	☐
19.	Wat betekent “big”?	☐	☐	☐
20.	Wat betekent “above”?	☐	☐	☐
21.	Wat betekent “five”?	☐	☐	☐
22.	Wat betekent “left”?	☐	☐	☐

23.	Wat betekent "one"?	☐	☐	☐
24.	Wat betekent "running"?	☐	☐	☐
25.	Wat betekent "right"?	☐	☐	☐
26.	Wat betekent "walking"?	☐	☐	☐
27.	Wat betekent "throwing"?	☐	☐	☐
28.	Wat betekent "more"?	☐	☐	☐
29.	Wat betekent "light"?	☐	☐	☐
30.	Wat betekent "falling"?	☐	☐	☐
31.	Wat betekent "three"?	☐	☐	☐
32.	Wat betekent "fewer"?	☐	☐	☐
33.	Wat betekent "flying"?	☐	☐	☐
34.	Wat betekent "high"?	☐	☐	☐
<i>Totaal aantal goed</i>				

Lijst 2:

	Correct	Incorrect (noteer het antwoord)	Geen antwoord

1.	Wat betekent "high" in het Nederlands?	☐	☐	☐
2.	Wat betekent "flying"?	☐	☐	☐
3.	Wat betekent "fewer"?	☐	☐	☐
4.	Wat betekent "three"?	☐	☐	☐
5.	Wat betekent "falling"?	☐	☐	☐
6.	Wat betekent "light"?	☐	☐	☐
7.	Wat betekent "more"?	☐	☐	☐
8.	Wat betekent "throwing"?	☐	☐	☐
9.	Wat betekent "walking"?	☐	☐	☐
10.	Wat betekent "right"?	☐	☐	☐
11.	Wat betekent "running"?	☐	☐	☐
12.	Wat betekent "one"?	☐	☐	☐
13.	Wat betekent "left"?	☐	☐	☐
14.	Wat betekent "five"?	☐	☐	☐
15.	Wat betekent "above"?	☐	☐	☐
16.	Wat betekent "big"?	☐	☐	☐
17.	Wat betekent "sliding"?	☐	☐	☐

18.	Wat betekent “below”?	☐	☐	☐
19.	Wat betekent “fewest”?	☐	☐	☐
20.	Wat betekent “in front of”?	☐	☐	☐
21.	Wat betekent “most”?	☐	☐	☐
22.	Wat betekent “small”?	☐	☐	☐
23.	Wat betekent “take away”?	☐	☐	☐
24.	Wat betekent “on”?	☐	☐	☐
25.	Wat betekent “catching”?	☐	☐	☐
26.	Wat betekent “next to”?	☐	☐	☐
27.	Wat betekent “two”?	☐	☐	☐
28.	Wat betekent “jumping”?	☐	☐	☐
29.	Wat betekent “heavy”?	☐	☐	☐
30.	Wat betekent “climbing”?	☐	☐	☐
31.	Wat betekent “behind”?	☐	☐	☐
32.	Wat betekent “add”?	☐	☐	☐
33.	Wat betekent “four”?	☐	☐	☐
34.	Wat betekent “low”?	☐	☐	☐



<i>Totaal aantal goed</i>			

Appendix III.

Vertaaltaak T1 > T2

Lees de volgende instructies letterlijk voor:

Je deed het heel goed toen ik de Engelse woorden zei en jij moest zeggen wat ze in het Nederlands betekenden. Nu gaan we het andersom doen! Ik ga het woord in het Nederlands zeggen en dan mag jij zeggen wat het Engelse woord is.

Daar gaan we!

Controle groep: Heel knap dat je hebt geprobeerd om te zeggen wat woorden in het Engels betekenen! Nu gaan we het andersom doen! Ik ga het woord in het Nederlands zeggen en dan mag jij proberen te zeggen wat het Engelse woord is.

Kies de juiste lijst hieronder en scoor elk item direct:

Lijst 1:

	Correct	Incorrect (noteer het antwoord)	Geen antwoord
1. Wat is “eraf halen” in het Engels?	☐	☐	☐
2. Wat is “vier” in het Engels?	☐	☐	☐
3. Wat is “twee” in het Engels?	☐	☐	☐
4. Wat is “klimmen” in het Engels?	☐	☐	☐
5. Wat is “meest” in het Engels?	☐	☐	☐
6. Wat is “voor” in het Engels?	☐	☐	☐
7. Wat is “op” in het Engels?	☐	☐	☐
8. Wat is “erbij doen” in het Engels?	☐	☐	☐
9. Wat is “onder” in het Engels?	☐	☐	☐
10. Wat is “klein” in het Engels?	☐	☐	☐
11. Wat is “vangen” in het Engels?	☐	☐	☐
12. Wat is “minst” in het Engels?	☐	☐	☐
13. Wat is “naast” in het Engels?	☐	☐	☐
14. Wat is “springen” in het Engels?	☐	☐	☐
15. Wat is “zwaar” in het Engels?	☐	☐	☐
16. Wat is “laag” in het Engels?	☐	☐	☐

17. Wat is "achter" in het Engels?	☐	☐	☐
18. Wat is "glijden" in het Engels?	☐	☐	☐
19. Wat is "vliegen" in het Engels?	☐	☐	☐
20. Wat is "gooien" in het Engels?	☐	☐	☐
21. Wat is "vallen" in het Engels?	☐	☐	☐
22. Wat is "rechts" in het Engels?	☐	☐	☐
23. Wat is "boven" in het Engels?	☐	☐	☐
24. Wat is "vijf" in het Engels?	☐	☐	☐
25. Wat is "rennen" in het Engels?	☐	☐	☐
26. Wat is "groot" in het Engels?	☐	☐	☐
27. Wat is "hoog" in het Engels?	☐	☐	☐
28. Wat is "lopen" in het Engels?	☐	☐	☐
29. Wat is "links" in het Engels?	☐	☐	☐
30. Wat is "drie" in het Engels?	☐	☐	☐
31. Wat is "een" in het Engels?	☐	☐	☐
32. Wat is "licht" in het Engels?	☐	☐	☐
33. Wat is "meer" in het Engels?	☐	☐	☐
34. Wat is "minder" in het Engels?	☐	☐	☐
<i>Totaal aantal goed</i>			

Lijst 2:

	Correct	Incorrect (noteer het antwoord)	Geen antwoord
1. Wat is "minder" in het Engels?	☐	☐	☐
2. Wat is "meer" in het Engels?	☐	☐	☐
3. Wat is "licht" in het Engels?	☐	☐	☐
4. Wat is "een" in het Engels?	☐	☐	☐
5. Wat is "drie" in het Engels?	☐	☐	☐
6. Wat is "links" in het Engels?	☐	☐	☐
7. Wat is "lopen" in het Engels?	☐	☐	☐
8. Wat is "hoog" in het Engels?	☐	☐	☐
9. Wat is "groot" in het Engels?	☐	☐	☐
10. Wat is "rennen" in het Engels?	☐	☐	☐
11. Wat is "vijf" in het Engels?	☐	☐	☐
12. Wat is "boven" in het Engels?	☐	☐	☐

13. Wat is “rechts” in het Engels?	☐	☐	☐
14. Wat is “vallen” in het Engels?	☐	☐	☐
15. Wat is “gooien” in het Engels?	☐	☐	☐
16. Wat is “vliegen” in het Engels?	☐	☐	☐
17. Wat is “glijden” in het Engels?	☐	☐	☐
18. Wat is “achter” in het Engels?	☐	☐	☐
19. Wat is “laag” in het Engels?	☐	☐	☐
20. Wat is “zwaar” in het Engels?	☐	☐	☐
21. Wat is “springen” in het Engels?	☐	☐	☐
22. Wat is “naast” in het Engels?	☐	☐	☐
23. Wat is “minst” in het Engels?	☐	☐	☐
24. Wat is “vangen” in het Engels?	☐	☐	☐
25. Wat is “klein” in het Engels?	☐	☐	☐
26. Wat is “onder” in het Engels?	☐	☐	☐
27. Wat is “erbij doen” in het Engels?	☐	☐	☐
28. Wat is “op” in het Engels?	☐	☐	☐
29. Wat is “voor” in het Engels?	☐	☐	☐
30. Wat is “meest” in het Engels?	☐	☐	☐
31. Wat is “klimmen” in het Engels?	☐	☐	☐
32. Wat is “twee” in het Engels?	☐	☐	☐
33. Wat is “vier” in het Engels?	☐	☐	☐
34. Wat is “eraf halen” in het Engels?	☐	☐	☐
<i>Totaal aantal goed</i>			

APPENDIX IV

Noteer meteen na elke respons het **nummer van het plaatje (1,2,3)** in de kolom “antwoord kind”.
Tel na afloop van het testen het aantal goed.

LIJST 1

nr	item	goede antwoord	antwoord kind	goed
1	oefenitem 1	2		☐
2	oefenitem 2	1		☐
3	below	3		☐
4	behind	1		☐
5	take away	1		☐
6	sliding	3		☐
7	take away	3		☐
8	sliding	1		☐
9	on	2		☐
10	add	1		☐
11	fewest	2		☐
12	next to	3		☐

13	most	3		☐
14	heavy	2		☐
15	catching	2		☐
16	next to	1		☐
17	in front of	3		☐
18	sliding	2		☐
19	jumping	2		☐
20	catching	1		☐
21	two	3		☐
22	fewest	1		☐
23	climbing	2		☐
24	add	2		☐
25	next to	2		☐
26	jumping	3		☐
27	climbing	3		☐
28	heavy	1		☐
29	behind	2		☐

30	small	2		☐
31	most	1		☐
32	behind	3		☐
33	catching	3		☐
34	in front of	1		☐
35	below	1		☐
36	four	3		☐
37	below	2		☐
38	two	2		☐
39	on	1		☐
40	small	3		☐
41	low	2		☐
42	four	2		☐
43	climbing	1		☐
44	in front of	2		☐
45	on	3		☐
46	low	1		☐

47	low	3		☐
48	small	1		☐
49	take away	2		☐
50	heavy	3		☐
51	two	1		☐
52	most	2		☐
53	jumping	1		☐
54	add	3		☐
55	fewest	3		☐
56	four	1		☐
Totaal aantal goed				

LIJST 2:

nr	item	goede antwoord	antwoord kind	goed
1	oefenitem 1	2		☐
2	oefenitem 2	1		☐
3	four	1		☐

4	fewest	3		☐
5	add	3		☐
6	jumping	1		☐
7	most	2		☐
8	two	1		☐
9	heavy	3		☐
10	take away	2		☐
11	small	1		☐
12	low	3		☐
13	low	1		☐
14	on	3		☐
15	in front of	2		☐
16	climbing	1		☐
17	four	2		☐
18	low	2		☐
19	small	3		☐
20	on	1		☐

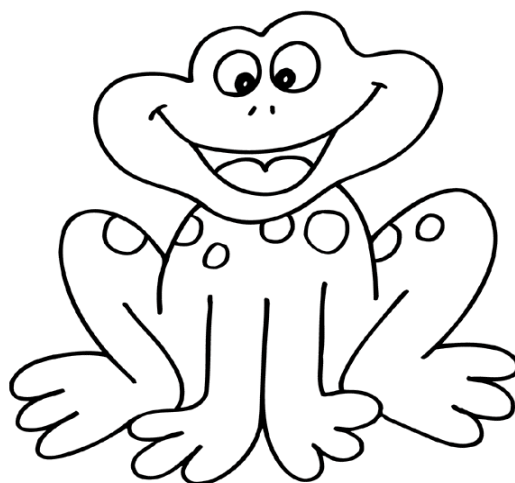
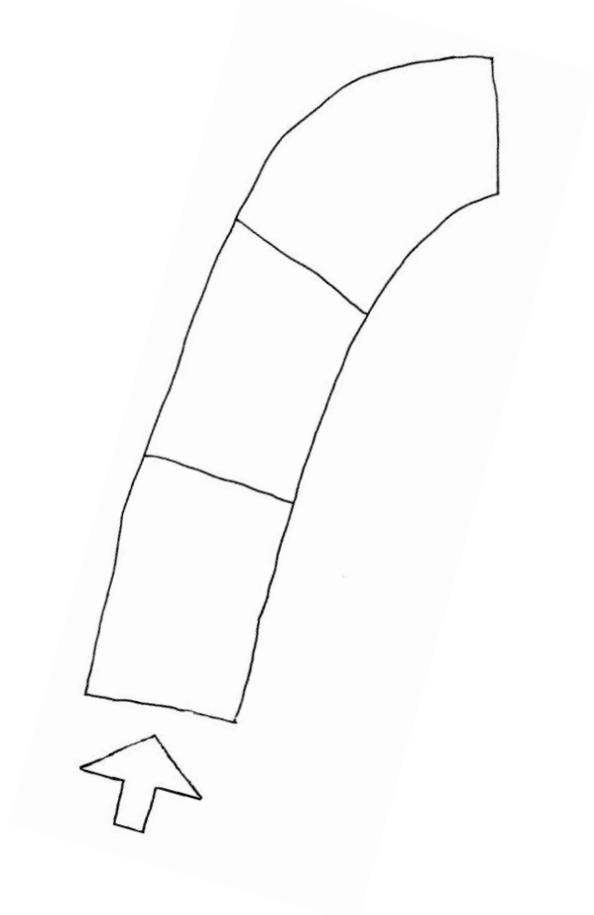
21	two	2		☐
22	below	2		☐
23	four	3		☐
24	below	1		☐
25	in front of	1		☐
26	catching	3		☐
27	behind	3		☐
28	most	1		☐
29	small	2		☐
30	behind	2		☐
31	heavy	1		☐
32	climbing	3		☐
33	jumping	3		☐
34	next to	2		☐
35	add	2		☐
36	climbing	2		☐
37	fewest	1		☐

38	two	3		☐
39	catching	1		☐
40	jumping	2		☐
41	sliding	2		☐
42	in front of	3		☐
43	next to	1		☐
44	catching	2		☐
45	heavy	2		☐
46	most	3		☐
47	next to	3		☐
48	fewest	2		☐
49	add	1		☐
50	on	2		☐
51	sliding	1		☐
52	take away	3		☐
53	sliding	3		☐
54	take away	1		☐

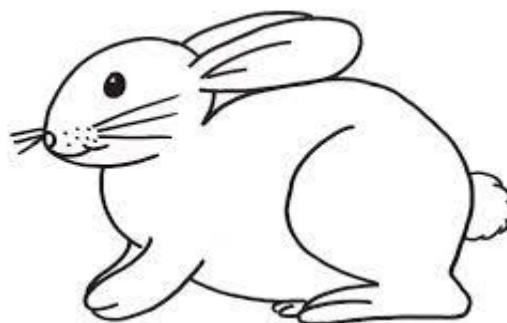
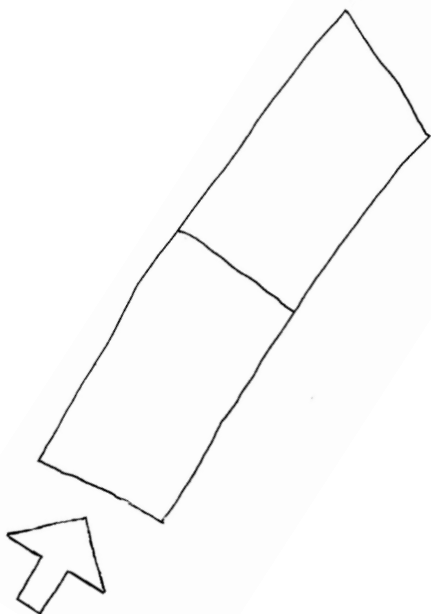
55	behind	1		☐
56	below	3		☐
Totaal aantal goed				

APPENDIX V

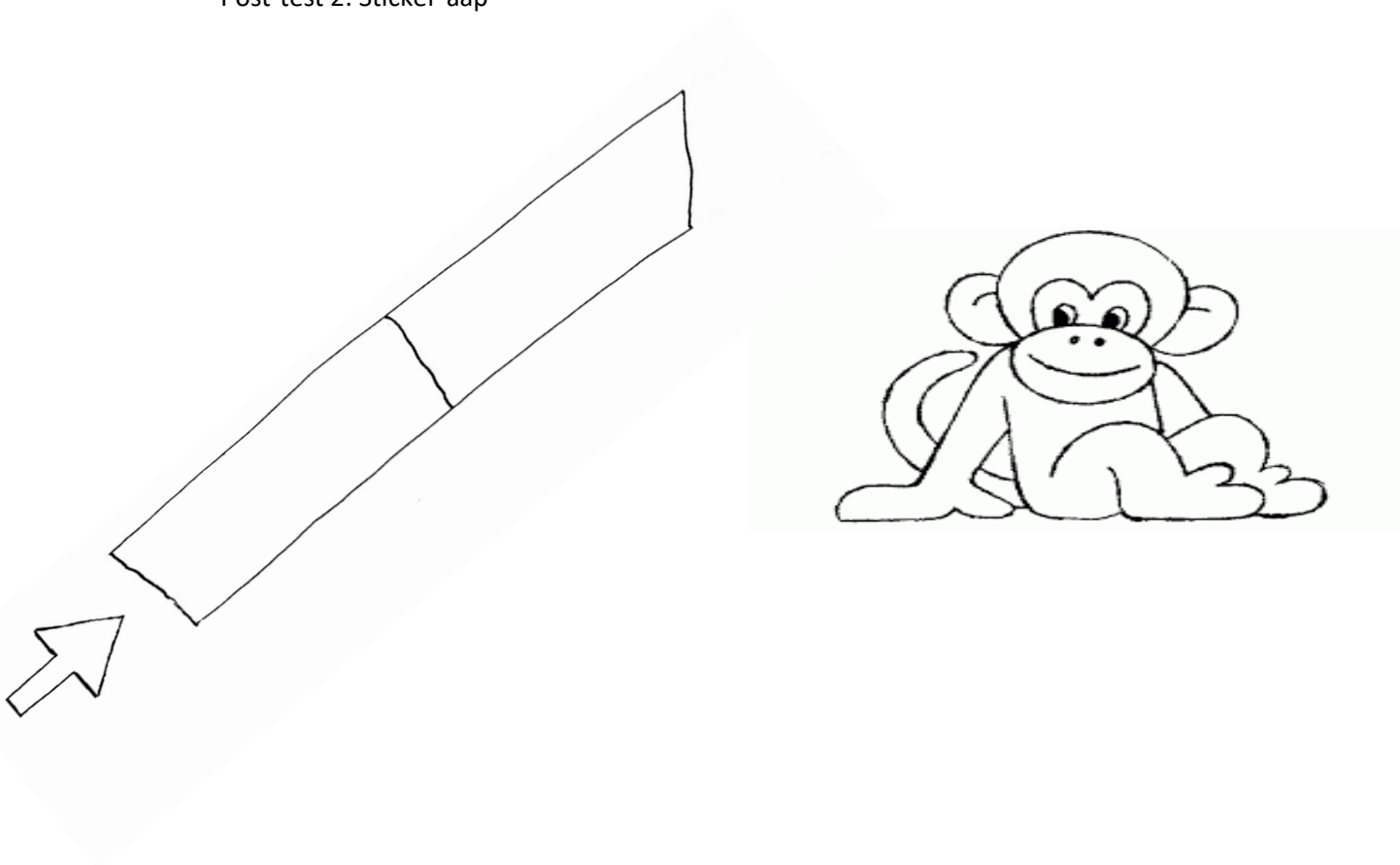
Pre-test: Sticker-kikker



Post-test 1: Sticker-konijn



Post-test 2: Sticker-aap



Appendix VI

Vragenlijst perceptie robot

Ik ga je nu een paar vragen stellen over Robin de robot. Ik ben benieuwd wat jij van Robin de robot denkt!

1. Vind je Robin de robot een jongen of een meisje?
Jongen/meisje/weet niet, omdat
2. Vind je Robin de robot meer een juf/meester* of meer een vriendje/vriendinnetje*?
Juf/vriendje/weet niet
3. Vind je Robin de robot meer een vriendje/vriendinnetje* of meer een ding?
Vriendje/ding/weet niet
4. Vind je Robin de robot meer een ding of meer een juf/meester*?
Ding/juf/weet niet
5. Vind je Robin de robot meer een mens of meer een ding?
Mens/ding/weet niet
6. Denk je dat Robin de robot dingen kan zien?
Ja/nee/weet niet, omdat
7. Denk je dat Robin de robot is gemaakt door een mens?
Ja/nee/weet niet, omdat
8. Denk je dat Robin de robot verdrietig kan zijn?
Ja/nee/weet niet, omdat
9. Denk je dat Robin de robot iets kan onthouden?
Ja/nee/weet niet, omdat
10. Denk je dat Robin de robot het voelt als je Robin kietelt?
Ja/nee/weet niet, omdat
11. Denk je dat Robin de robot kan denken?
Ja/nee/weet niet, omdat
12. Denk je dat Robin de robot moet eten?
Ja/nee/weet niet, omdat
13. Denk je dat Robin de robot het begrijpt als je iets zegt?
Ja/nee/weet niet, omdat
14. Denk je dat Robin pijn kan hebben?
Ja/nee/weet niet, omdat
15. Denk je dat Robin iets leuk kan vinden?
Ja/nee/weet niet, omdat

16. Denk je dat Robin de robot groeit?

Ja/nee/weet niet, omdat

17. Denk je dat Robin de robot blij kan zijn?

Ja/nee/weet niet, omdat

18. Denk je dat Robin jou kan herkennen?

Ja/nee/weet niet, omdat

19. Denk je dat Robin de robot kapot kan gaan?

Ja/nee/weet niet, omdat

* Hangt af van of het kind de robot een jongetje of een meisje vindt, dus het antwoord op vraag 1.

Appendix VII

Meeneemlijst:

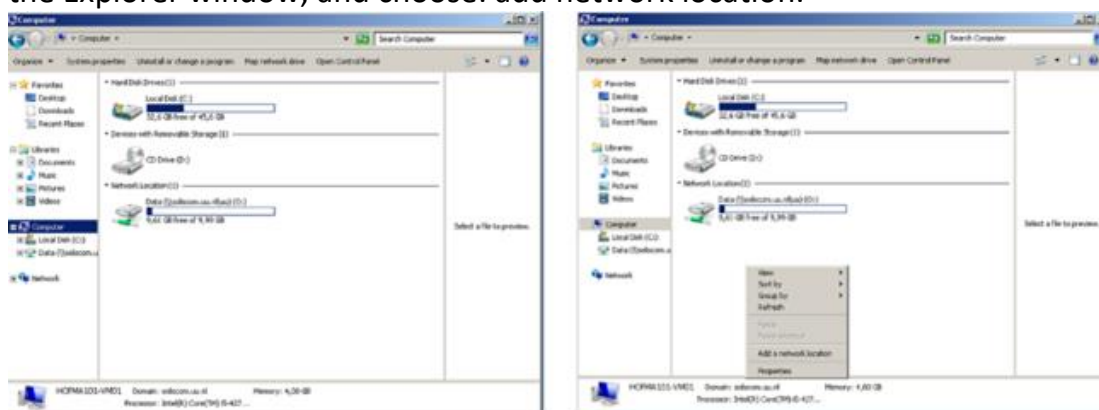
Lessen	testen
Robot + lader	Laptop (met juiste software en bestanden – let op voor pre- en post-test)
Router	Scoreformulieren (let op pre- of post-test)
Laptop met juiste software erop + lader	Stickers
Tablet met juiste software erop + lader	Stickerkikker (of ander dier)
USB - Ethernet adapter	PPVT test boek (2 boeken) (alleen pre-test)
Plastic doos	Lijsten met kinderen en test versies
2 netwerkkabels	Camera
Contactdoos	Statief
Plakband	2 SD kaarten (of meer als nodig)
Koffer met verpakkingsmateriaal	
2 camera's	
2 statieven	
4 SD kaarten (of meer als nodig)	
Flesje water	

Externe microfoon	
Lijst met kinderen, condities en nummers	
Billendoekjes + beeldschermdoekjes	
“Niet storen” papier + plakband	

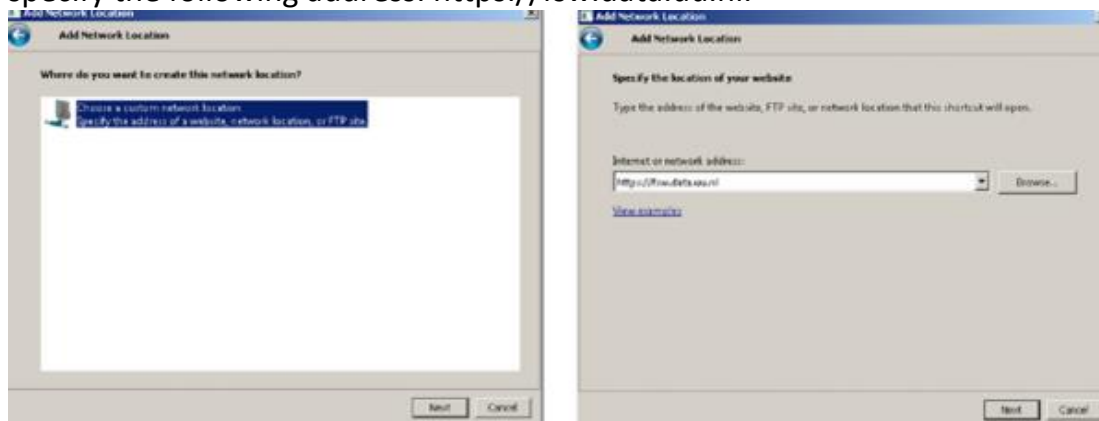
APPENDIX VIII - YODA

Storing data on YourData (YoDa),
faculty of Social and Behavioural Sciences, Utrecht University.
May 1, 2017.
Version 1.0

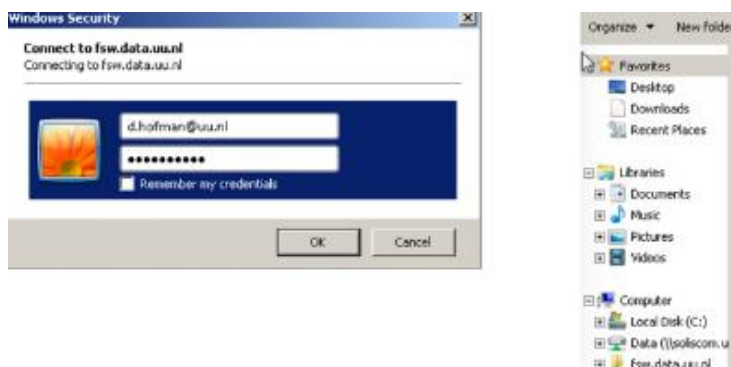
For connecting to your storage folder, follow these steps:
Working from the Uithof or home , on a Windows computer:
Log in as usual. In the explorer, click 'computer', right-click in an empty space of the Explorer window, and choose: add network location.



When prompted for a location, choose 'Choose a custom network location', and specify the following address: <https://fsw.data.uu.nl>.



Login using your e-mail address and solis-password (non UU people will receive their own login details). Now, you will find your encrypted storage for person data in your list of drives.



Note to windows users:

Windows (7, 8, 10) in combination with the WEBDAV protocol causes issues when using files larger than 50MB. This file size limit can be circumvented by one of the options described here:

<https://support.microsoft.com/nl-nl/help/2668751/you-cannot-download-more-than-50-mb-or-upload-large-files-when-the-upl>

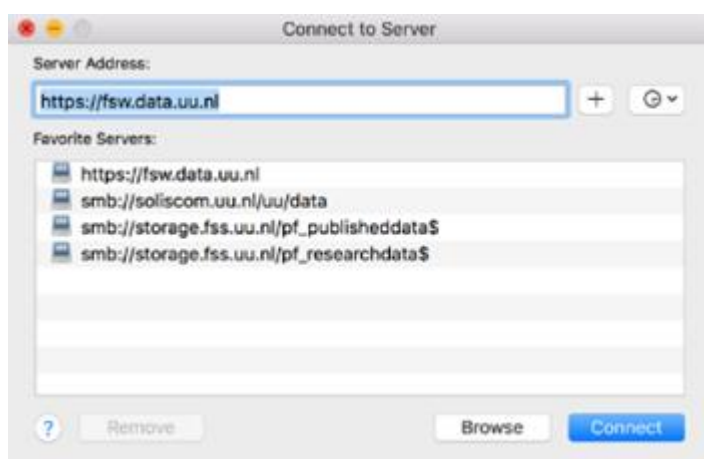
In short:

- you either download the 'easy fix wizard' and using this adjust the max flie size limit to 4294967295;
- you adjust the registry key
HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\WebClient\Parameters:
select 'FileSizeLimitInBytes', right-click and choose 'modify', click 'decimal' and enter 4294967295.

Restart.

Working from a mac:

Click finder, and use command-k to open the connect-to-server dialog box:



Enter `https://fsw.data.uu.nl`, and use enter your e-mail address and password to login. Next, double-click the folder you want to store data in, and copy-paste files to this location.

Appendix III: Pre-registration Aspredicted



L2TOR (#8181)

Author(s)

Emiel Krahmer (Tilburg University) - e.j.krahmer@tilburguniversity.edu
 Rianne van den Berghe (Utrecht University) - m.a.j.vandenbergh@uu.nl
 Mirjam de Haas (Tilburg University) - mirjam.deHaas@tilburguniversity.edu
 Ora Oudgenoeg (Utrecht University) - O.Oudgenoeg@uu.nl
 Jan de Wit (Tilburg University) - J.M.S.deWit@tilburguniversity.edu

Created: 02/02/2018 10:37 AM (PT)

Public: 02/05/2018 02:45 AM (PT)

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

In a large-scale, longitudinal evaluation experiment we will study whether young Dutch-L1 children can learn English-L2 words from a social robot. Children are between 5;0 and 5;11 years at the start of the experiment, and participate in a series of 7 lessons (6 "core" plus 1 general recap session), containing 34 English target words in total.

We hypothesise:

(H1) The robot will be effective at teaching children L2 target words: children will learn words from a robot (H1a) and will remember them better (H1b) than children who participate in a no treatment (control) condition.

(H2) Children will learn more words (H2a), and will remember them better (H2b) when learning from a robot than from only a tablet.

(H3) Children will learn more words (H3a), and will remember them better (H3b) when learning from a robot that produces iconic gestures than from one that does not produce such gestures.

3) Describe the key dependent variable(s) specifying how they will be measured.

Before the sessions start, we pre-test the target vocabulary (34 English words). Children are presented with each of the target words and asked what they mean in Dutch ("Wat betekent t in het Nederlands?"). In addition, we measure:

- general verbal ability (Peabody Picture Vocabulary Test),
- attention (visual search task),
- phonologic memory (non-word repetition task).

At the end of each of the 6 core lessons, we test in a brief exercise whether the words from that lesson were learned by the child.

Crucially, after all 7 lessons are completed, 6 posttests are conducted (a 3-part immediate posttest, which will be administered max. 2 days after the final lesson, and a 3-part delayed posttest, taking place after approx. 4 weeks). Both posttests involve the same 3 parts:

- Translation from English to Dutch (as in the pretest)
- Translation from Dutch to English
- Test of understanding of English words

For the first two all 34 target words will be tested, for the last a selection of 1/3 of the words will be tested in multiple instances. In all cases, the number of correct answers will be analysed.

4) How many and which conditions will participants be assigned to?

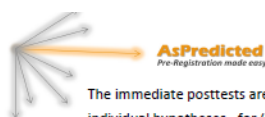
Participants will be assigned to 1 of 4 conditions: ☒

- (1) robot with iconic gestures + tablet,
- (2) robot without iconic gestures + tablet,
- (3) tablet-only,
- (4) no treatment.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

The basic analysis which we will conduct is a MANCOVA, with Condition as the independent variable (4 levels), the score on the Pre-test as a covariate, and the 6 Posttests as the dependent variables.





The immediate posttests are relevant for (H1a), (H2a) and (H3a); the delayed posttests for (H1b), (H2b) and (H3b). Planned comparisons allow us to address individual hypotheses - for (H1) we will compare conditions (1) and (2) with (4); for (H2) we compare conditions (1) and (2) with (3); and for (H3) we compare conditions (1) and (2).

Given that loss of power may be an issue for the basic MANCOVA we will conduct follow-up analyses to see whether it is possible to combine the 3 parts of the post-tests into a single latent outcome variable (one for immediate and one delayed) and analyse it using an ANCOVA.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

Children will be excluded from the experiment, if:

- they know more than half of the words in the pretest (which we deem unlikely),
- they miss three lessons (e.g., because they are ill for more than one week), or
- they are bilingual.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

We aim to collect data from 180 children: 50 in each of the conditions (1) - (3), and 30 in condition (4) (no treatment). Since some dropout is to be expected with this kind of experiment (longitudinal, with children), we oversample with 10 extra children per condition.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

We consider the following secondary analyses:

- We will explore the influence of age and gender, as well as of verbal ability, attention and phonological memory through additional ANCOVA analyses. ☐
- We plan to study children's engagement during the lessons, by manually scoring children's behaviour and via additional judgement studies, in which independent raters evaluate the perceived engagement of children in randomly selected brief fragments. We expect children to be more engaged when interacting with a robot than with a tablet, and more engaged with a robot that gestures than with one that does not. Additionally, we will explore the link between engagement and learning performance (hypothesising that more engaged children learn better). ☐
- We will explore the existence and nature of a novelty effect: for this purpose, we will analyse performance on the small tests at the end of each lesson as well engagement levels for the different lessons.
- Next to the (M)ANCOVAs described above, we will explore mixed effect analyses for the data
- In a separate questionnaire we will ask how children perceive the robot, and we may explore how their perceptions influence interactions with the robot. In addition, parents will be asked to fill in a questionnaire about their child, asking about the social-economic status (SES), family history, incidence of learning disorders and other risk factors, and tablet use at home. We may also explore the influence of these factors.

This preregistration is written by the following authors (in alphabetical order), participating in the design and running of the experiment:

- Rianne van den Berghe (m.a.j.vandenbergh@uu.nl)
- Mirjam de Haas (mirjam.deHaas@tilburguniversity.edu)
- Emiel Krahmer (e.j.krahmer@tilburguniversity.edu)
- Paul Leseman (P.P.M.Leseman@uu.nl)
- Ora Oudgenoeg (O.Oudgenoeg@uu.nl)
- Josje Verhagen (J.Verhagen@uu.nl)
- Paul Vogt (P.A.Vogt@tilburguniversity.edu)
- Jan de Wit (J.M.S.deWit@tilburguniversity.edu)
- Bram Willemsen (B.Willemsen@tilburguniversity.edu)

Appendix IV: Master thesis H. Leeuwestein

Research Master's programme: Educational Sciences: Learning in Interaction
Utrecht University, the Netherlands

MSc thesis Hanneke Leeuwestein

TITLE: Teaching Turkish-Dutch kindergartners Dutch vocabulary using
social robots: does use of Turkish translations benefit Dutch word learning?
March 2018

Supervisors:

Dr. Ora Oudgenoeg-Paz

Dr. Josje Verhagen

Prof. dr. Paul Leseman

Word count: 7734

Abstract

When bilingual children learn a second language (L2), they rely, for a great deal at least, on their native language (L1). Consequently, incorporating children's L1 into L2 language learning programs could benefit L2 learning. However, in linguistically diverse classrooms, teachers typically cannot provide L1 input to all pupils. Educational technologies such as multilingual social robots offer opportunities for one-to-one tutoring using L1 to learn L2. This study investigated whether providing L1 translations in an L2 vocabulary training facilitates L2 word learning. Turkish-Dutch kindergartners ($n = 67$) were taught six Dutch (L2) words for which they knew the L1 Turkish but not the L2 Dutch label. Using a within-subjects design, half of the target words were taught by a bilingual robot, alongside their Turkish translations; the other half were taught by a monolingual robot, using Dutch only. The results showed no significant differences in learning gains between the conditions, although children indicated they preferred the bilingual robot over the monolingual robot. However, a moderation effect of children's Turkish vocabulary knowledge was found. Children with well-developed Turkish vocabulary knowledge showed higher learning gains when L1 translations were provided, than children with lower levels of Turkish vocabulary knowledge who showed higher learning gains in the Dutch-only learning environment. Possible explanations for this moderation effect and suggestions for future research are discussed.

Keywords: second language teaching, cross-language transfer, robot-assisted language learning, bilingual kindergartners

Teaching Turkish-Dutch kindergartners Dutch vocabulary using social robots: do L1 translations benefit L2 word learning?

Classrooms are becoming increasingly diverse in terms of children's cultural and linguistic backgrounds. This offers both new opportunities and challenges (OECD, 2015). In the Netherlands, many children grow up with a native language (L1) other than the national language, and therefore have to learn Dutch as a second language (L2) when entering kindergarten. Of all children with a migration background in the Netherlands, the Turkish-Dutch children form the largest group (CBS, 2017). These children often lag behind their monolingual Dutch peers in academic performance (Duarte & Gogolin, 2013; OECD, 2015) and socio-emotional wellbeing (OECD, 2015). Scheele, Leseman, and Mayo (2010), using equivalent parallel tests of Turkish and Dutch vocabulary, found a significant vocabulary gap between bilingual Turkish-Dutch and monolingual Dutch children at the age of three, prior to kindergarten entry: both L1 and L2 vocabulary size of Turkish-Dutch children was smaller than the vocabulary size of their monolingual Dutch peers. Data from the Dutch central bureau of Statistics (CBS, 2016) show that these educational differences persist over time, as Turkish-Dutch students most often attend the lowest levels of secondary education. The relatively low Dutch language proficiencies of Turkish-Dutch students may explain this finding (CBS, 2016).

In Turkish-Dutch immigrant families the amount of Dutch language input is often limited, which can be explained by the lack of Dutch language knowledge of the parents and the high extent to which many families are focused on maintaining their Turkish language and culture (Extra, Aarts, van Der Avoird, Broeder, & Yagmur, 2002; Hartgers,

2012). When entering Dutch kindergarten, children from families in which mainly Turkish is spoken, thus, often face several difficulties; not only are they exposed to a language they do not master yet, but they also have to learn new academic concepts and skills in this language (Aarts, Demir, & Vallen, 2011).

The disadvantageous educational position of Turkish-Dutch children and their families' focus on L1 maintenance makes it relevant to develop approaches to support bilingual language development, taking into account children's L1 and L2. Currently, support of minority children's L1 is rather limited to non-existent in most kindergarten classrooms in the Netherlands. In addition to a predominant assimilationist education policy, this can also be explained by the fact that in the current superdiverse classrooms, teachers often lack the skills required to provide L1 input for all pupils (Ministry of Education, 2017).

The use of educational technologies offers opportunities to overcome this practical constraint on L1 use in the classroom. One of the newest forms of educational technology involves social robots. Robots can be programmed to speak multiple languages and could therefore be of added value for (second) language tutoring (Kanero et al., 2018). The current study, which is carried out within a larger project on L2 tutoring using social robots (L2TOR), explores opportunities for robot-assisted L2 learning in Turkish-Dutch immigrant children using a bilingual Turkish-Dutch robot. Specifically, we investigate whether providing L1 translations could facilitate L2 word learning using a bilingual robot. Furthermore, we study whether children prefer being taught by a monolingual Dutch or a bilingual Turkish-Dutch robot and if this is related to their word learning outcomes.

Herewith we contribute to the current scientific and societal debate on whether and how immigrant children's L1 should be incorporated into classrooms whilst exploring opportunities for integrating social robots into language education.

Role of L1 in L2 learning

So far, no consensus has been reached on the question as to whether and how providing forms and structures of children's L1 in educational settings could support their L2 acquisition (Swain & Lapkin, 2013). On the one hand, research of Scheele et al. (2010) has shown that L1 use may interfere in L2 learning by reducing the time for both L1 and L2 input. Although the study of Scheele and colleagues focused solely on language input at home, research by Pulinx, Agirdag and van Avermaet (2017) in Flemish Belgium has shown that this competition between L1 and L2 input is the most important reason for many teachers to interdict students to speak other languages than the language of instruction at school. These teachers have a so-called monolingual view, reasoning that multilingualism would hinder children's Dutch language proficiency. This view is also reflected in educational policies in the Netherlands. That is, on the governmental level, educational institutions are prescribed to mainly focus on Dutch language acquisition, which is described as the most crucial factor for integration in Dutch society (CBS, 2016; Ghorashi & van Tilburg, 2006).

Such monolingual orientations, sometimes referred to as parallel monolingualism (Heller, 1999), separate bilingualism (Creese & Blackledge, 2008) or two monolinguals in one body (Gravelle, 1996), are based on the idea that languages should be kept separate (i.e., spoken in different contexts) to avoid linguistic interference between the languages

and foster monolingual-like language competencies in both languages. In this perspective, it is argued that focus on L1 acquisition reduces the effectiveness of L2 teaching (for a discussion of this view, see Cummins, 2008; Gogolin, 2011).

Conversely, several studies have studied the positive interplay between L1 and L2 skills (e.g., Bouvy, 2000; Cummins, 1991; Verhoeven, 2007). In such studies, it is argued that bilingual children rely on one shared conceptual basis, enabling them to transfer their L1 skills more easily into their L2 (Cummins, 1981). In this way, bilingual children's vocabulary development generally starts slow because the available time for language learning has to be divided between the two languages. However, after reaching a certain threshold of knowing different concepts in their two languages, vocabulary knowledge in both languages starts to grow relatively fast because of the positive transfer skills between the L1 and L2 (Bee Chin & Wigglesworth, 2007). That is, when children already know a concept in their L1, they only have to learn the label instead of the conceptual meaning in their L2 (Ellis, 2006). Leseman, Henrichs, Blom, and Verhagen (in press) found such facilitating effects in Turkish-Dutch children between the ages of three and six years old. Their results showed that these Turkish-Dutch children's L1 vocabulary growth was positively related to their L2 vocabulary size and growth. Moreover, the input these children received in their L1 was associated with their L2 vocabulary size.

Apart from possible academic benefits, incorporating learners' L1 may also affect their socio-emotional wellbeing and identity development. Research showed that being part of an ethnic or cultural group becomes a part of children's identity already between the ages of three and five years (Ruble et al., 2004 in Vedder & van Geel, 2017). Children's

L1 is an important marker of their identity (Le Page & Tabouret-Keller, 1985), so actively discouraging children to use their L1 at school might therefore have negative consequences for the identity development of these young Dutch-Turkish immigrant children (Agirdag, 2017).

Finally, studies that specifically focused on the linguistic effects of L1 use in L2 contexts also showed that L1 use can enhance engagement (Holmes, 2008), task participation (Probyn, 2005) and self-esteem, contributing to positive identity development (Creese & Blackledge, 2015; Pulinx, van Avermaet & Agirdag, 2017). Concluding, several studies have pointed to potential academic and socio-emotional benefits of L1 use in interventions aimed at teaching L2.

Incorporating the L1 in classrooms: why (not)?

Despite previous findings indicating that L1 use benefits both L2 learning and children's socio-emotional wellbeing, immigrant children's L1s are rarely used in formal education (Gogolin, 2011). At multiple schools, Agirdag (2017) observed that sometimes children are even negatively approached by teachers when they use a language other than the language of instruction. This is in contrast to the daily practice of many bilinguals who flexibly switch between languages for communicative purposes (Creese & Blackledge, 2015; Lewis, Jones, & Baker, 2012).

Zooming in on the classroom, Cole (1998) therefore states that “the struggle to avoid L1 at all costs can lead to bizarre behavior: One can end up being a contortionist trying to explain the meaning of a language item, where a simple translation would save time and anguish ... if students understand the concept of a noun, it is much simpler to

translate the word “noun” than to describe it in L2” (p. 2). Celik (2003) studied this approach in a L2 English classroom, involving first year university students sharing the same L1. L2 vocabulary was taught in an L2-embedded story, and several target words that were expected to be unknown in the L2 were introduced in the learners’ L1. Celik found this to be an effective method for L2 vocabulary teaching. Note however that this study was conducted in a classroom where the teacher and students all shared the same L1. In linguistically diverse classrooms, it is virtually impossible for teachers to provide all children with L1 input (Gogolin, 2011). Therefore, social robots that can be programmed to speak multiple languages could be used in such classrooms.

Robot-assisted language learning

Besides the fact that robots can speak multiple languages, there are several other advantages of using this technology for (second) language tutoring. Robots can provide additional language input of high quality in one-to-one tutoring sessions (Belpaeme et al., 2018; Kanero et al., 2018). Moreover, Leyzberg, Spaulding, Toneva, and Scassellati (2012) found humanoid robots’ physical entity being an advantage over other digital learning environments. In a puzzle-solving task, participants who were instructed by a physically present robot solved puzzles faster than participants who were only provided voice recordings or a videotaped robot. Two explanations are proposed by the authors. First, the novelty of robots may increase enthusiasm to participate. Second, it is argued that a physically present robot may be perceived as more compelling than videotaped or recorded instructions, which may lead to taking the task instructions more seriously.

In recent years, the field of robot-assisted language learning (RALL) has gained popularity among researchers. Mixed results have been found on robot's effectiveness in vocabulary teaching (for a review, see van den Berghe, Verhagen, Oudgenoeg-Paz, van Der Ven, & Leseman, under review). Based on an experiment with nineteen participants between four and six years old, Kory Westlund and colleagues (2015) concluded that teaching vocabulary by robots can be as effective as by human teachers. A recent review study by Kanero and colleagues (2018) concluded that robots can successfully teach vocabulary to young children, but that no studies have shown robots to be better in teaching vocabulary compared to human teachers. However, in our study, a social robot is purposefully used as a tool that can provide high quality input in both learners' L1 and L2, contrary to most teachers. To the best of our knowledge, no previous RALL studies have focused on opportunities to stimulate positive transfer effects between children's L1 and L2 in a word learning intervention.

Present study

The focus of this study is specifically on Turkish-Dutch kindergartners for whom the majority language, Dutch, is their L2. In an experimental setting using social robots, these children were taught Dutch words for which they knew the Turkish but not the Dutch label. In a L2-L1 condition, a bilingual Turkish-Dutch robot was used to teach children Dutch words, alongside their Turkish translations. In a L2-only condition, a monolingual robot was used to teach Dutch words using Dutch only. We investigated whether the bilingual social robot can facilitate L2 word learning through providing L1 translations. Moreover, we studied whether children prefer being taught by the monolingual or the

bilingual robot, and if their preference is related to their word learning outcomes. Last, we explored whether learning gains between conditions are affected by children's vocabulary knowledge in Turkish and Dutch.

Since the amount of L1 involved only a few translations, we did not expect any negative effect of L1 reducing time for L2 learning (e.g., Scheele et al., 2010). In fact, based on theories of cross-language transfer effects between L1 and L2 (e.g., Cummins, 1981), it was expected that learning gains would be higher when L1 translations are provided. Moreover, we hypothesized that the higher children's Turkish vocabulary scores, the higher their learning gains in the L2-L1 condition compared to the L2-only condition. The opposite effect was expected regarding Dutch vocabulary scores: the higher children's Dutch vocabulary scores, the higher their learning gains in the L2-only condition. Last, we expected children to prefer being taught by the bilingual robot and that this preference moderated learning gains between the conditions, because L1 use may acknowledge children's cultural identity and increase both engagement and learning gains (Holmes, 2008; Pulinx, van Avermaet & Agirdag, 2017).

Method

Participants

A total of 95 Dutch children with a Turkish immigration background were recruited. However, 28 children did not meet the inclusion criteria regarding knowledge of the target words (see next sections) and were therefore excluded from the experimental task. The remaining 67 children (33 boys, 34 girls) were aged between 48 and 71 months ($M = 57.16$, $SD = 6.28$). They were recruited from nine primary schools in nine different

cities in the Netherlands. Six were public schools ($n = 32$), and three were Islamic schools ($n = 35$). A passive informed consent procedure was used, as approved by the ethics committees of both Utrecht University and Tilburg University.

Participants' parents all migrated to the Netherlands before the participating children were born. Participants were therefore labeled as second (83.1%) or third (16.9%) generation immigrants with a Turkish background. Three children were born in either Germany or Belgium, but moved to the Netherlands before kindergarten age. Most of the children came from low and middle socio-economic status households ($M = 2.50$, $SD = .92$, range 1 (primary school) to 5 (master's degree)) as indicated by the fact that 82% of parents attained primary, secondary or vocational education at most. As displayed in Table 1, all children are, in various degrees, exposed to the Turkish language at home. This information was obtained through a parental questionnaire, which five parents did not (completely) fill in.

Table 1

Frequencies of the language(s) parents communicate in with their child ($n = 62$)

Language(s)	Number of children
Turkish only	5
Mostly Turkish, and a little bit of Dutch	30
Equally both Turkish and Dutch	22
Mostly Dutch, and a little bit of Turkish	5

Experimental procedure

Using a Softbank Robotics NAO humanoid robot (58cm high) through an animated game on a Microsoft Surface Pro tablet (12.3-inch display), children were taught Dutch words. The experiment employed a within-subjects design with two conditions. Half of the target words were taught by a bilingual robot which provided Turkish translations (L2-L1

condition), and the other half of the target words were taught by a monolingual robot using Dutch only (L2-only condition). All children thus interacted with both a monolingual Dutch (L2-only condition) and a bilingual Turkish-Dutch (L2-L1 condition) robot. The presentation of both conditions was counterbalanced across participants. The measures were self-reported enjoyment, children's robot preference, target word retention, and children's L1 and L2 vocabulary knowledge.

Target word selection. Previous research showed an enormous heterogeneity in language proficiencies among our target group (e.g., Mayo & Leseman, 2008; Demir-Vegter, Aarts, & Kurvers, 2014). Therefore, it was not possible to select words to teach in the experimental task which *all* children knew in Turkish but not in Dutch. Instead, we decided to conduct a pre-test to test receptive vocabulary knowledge of twenty words in both Turkish and Dutch, and use the pre-test results to make an individual selection of target words for each child. The pre-test was conducted up to a maximum of one week prior to the experiment. It contained sixteen spatial verbs and four prepositions¹ that were all retrieved from the basic word list for kindergartners in Amsterdam (*'Basiswoordenlijst Amsterdamse Kleuters'*) (Mulder, Timman, & Verhallen, 2009). This word list has been developed by the municipality of Amsterdam and is meant to serve as a guideline for kindergarten teachers of Dutch L2 learners. It contains words that most monolingual Dutch kindergartners know already, and bilingual children should learn before the start of formal education in primary school. The list is divided in four levels of increasing difficulty.

¹ These are postpositions in Turkish.

During the pre-test, children were exposed to native speaker word recordings and were instructed to choose the corresponding picture out of four options. Target words of which the child knew the Turkish but not the Dutch label, were selected as target words for this child in the experiment. Children with at least four to six possible target words were selected for the experiment, and the ones with less than four possible target words were excluded from participation ($n = 28$). To keep the length of the training session equal, participants who had only four ($n = 20$) or five ($n = 9$) possible target words were supplemented with one or two extra ‘filler words’. These fillers were words children already knew in both languages, except for one child who did not know the Turkish label of one word. Each target word was alternately assigned to the conditions, resulting in an equal division of each target word between the two conditions. The order in which target words were taught within each condition was randomized.

Robot introductions. Previous studies on child-robot interaction have pointed to the importance of an introduction session prior to experimental tutoring sessions (e.g., Han, Jo, Jones, & Jo, 2008). This may reduce the novelty effect, as described above, which entails that children’s curiosity when being introduced to new technology could increase perceived usefulness and intentions to interact with the robot (Liu, Liao, & Pratt, 2009). Therefore, children were introduced to both the monolingual and the bilingual robot in groups, at least several hours up to two weeks before the experiment.

During this introduction, the children were told that there were two robots visiting the school: a monolingual Dutch robot and a bilingual Turkish-Dutch robot. To make the children believe that there were indeed two different robots, the experimenter changed the

robot's shirt and called the robot by a different name. Both introductions consisted of a background story followed by the robot performing a dance together with the children. Due to illness, five children did not attend the robot introductions.

Experiment. Participants took part in a one-to-one session with the robot and a Microsoft Surface Pro tablet (12.3-inch display), with the setup as displayed in Figure 1. The design of the lesson was based on other lessons designed within the L2TOR project (Oudgenoeg-Paz, Verhagen, & Vlaar, 2017). The experiment started with an introductory part in which, if necessary, the experimenter was actively involved to make sure participants understood what the robot expected them to do. Children practiced three actions they were instructed to do individually during the experiment: (1) selecting objects on the tablet, (2) repeating words produced by the robot, and (3) acting out movements with dolls.

The introduction was directly followed by two sessions, the L2-only and the L2-L1 condition (in counterbalanced order), with a short break in between to secretly change the robot's shirt. In these sessions, children were instructed to look for animals in the forest displayed on the tablet. The words were separately introduced through animal animations in which either animals performed action verbs or, in case of preposition target words, animals were placed relative to trees. Both conditions ended with a task in which participants heard a target word and had to select the correct animation.

As displayed in Box 1, the only difference between the conditions was that in the L2-L1 condition, a Dutch sentence was added twice: 'in Turkish we say X', referring to Turkish translations of the Dutch target words. In the L2-only condition, participants were

exposed to the Dutch target words at least ten times. In the L2-L1 condition, children heard the Dutch target word at least eight times, complemented with a Turkish translation which was provided twice. Some children who did not select the correct object at the first trial were additionally presented the Dutch target word during extra trials. The complete experiment lasted approximately half an hour and all interactions were videotaped.

Figure 1
Setup during the experiment



Box 1

*Model of the interaction script for both conditions***L2-only condition – monolingual robot**

Robot: Hey, look! I see a monkey. The monkey is shaking the tree.

Robot: Touch the monkey that is shaking the tree on the screen.

(Children select monkey on the tablet)

Tablet: Shaking. The monkey is shaking.

Robot: Did you hear that? So, in Dutch we say 'shaking'. The monkey is shaking the tree. Say after me: shaking.

(Children repeat after the robot)

Robot: Can you shake the dolls? Get them out of the bag. Come on, show me how you can shake the dolls.

(Children act out target word with the dolls)

Robot: Okay, let's put them in the bag again.

Task: Hey, where do you see a monkey shaking?

L2-L1 condition – bilingual robot

Robot: Look! Do you see the monkey that is shaking the tree? In Turkish we say *sallamak*.

Robot: Touch the monkey on the screen.

(Children select monkey on the tablet)

Tablet: Shaking. The monkey is shaking.

Robot: Oh, so it is *sallamak* in Turkish and shaking in Dutch. Say after me: shaking.

(Children repeat after the robot)

Robot: Can you shake the dolls? Get them out of the bag. Come on, show me how you can shake the dolls.

(Children act out target word with the dolls)

Robot: Okay, let's put them in the bag again.

Task: Hey, where do you see a monkey shaking?

Note. In the original (Dutch) interaction, all target words were presented as infinitives.

Measures

The dependent variables in the experiment were enjoyment, children's preference for either one of the two robots, and target word retention as assessed with comprehension and translation tests. Enjoyment was measured immediately after each condition, and children were asked about their preferences for either one of the robots directly after the tutoring session. Target word retention was measured immediately after the experiment and during a delayed post-test at least six days after the experiment ($M = 8.07$ days, $SD = 2.70$ days, range = 6 – 18 days). Assessing retention on both an immediate and a delayed

post-test is common in vocabulary learning interventions (Marulis & Neuman, 2010). Last, a standard test was used to measure children's vocabulary in both their L1 and L2.

Enjoyment. Participants' degree of enjoyment was measured after each condition, to investigate possible differences in enjoyment between conditions. Children were shown images of a happy and an unhappy child who supposedly played with the same robots. Children were asked to which child they were most similar too. Subsequently, they were asked whether they really resembled the child, or only a little. In this way, an ordinal variable was created with four values, ranging from 'absolutely not enjoyable' to 'very much enjoyable' (van den Berghe, van der Ven, Verhagen, Oudgenoeg-Paz, Papadopoulos, & Leseman, in preparation).

Robot preference. Immediately after the experiment, participants were shown a picture of the monolingual and bilingual robots, recognizable by their different t-shirts. The experimenter then asked the children which robot they would prefer to play with again: Dutch-speaking robot Robin, or robot Deniz who could speak both Turkish and Dutch. Children who started with the L2-only condition were told that they first played with Robin, followed by Deniz. The answer options were provided in the same order, keeping the counterbalanced order of the experimental conditions.

Comprehension test. Passive knowledge of the six target words was tested with a comprehension task with three items per target word. In each item, children saw three videos (or pictures in case of prepositions) simultaneously on a tablet screen. They had to choose the video that corresponded with the target word. The two distractors included another target word that that particular child had been exposed to, and a word that was not

included in the experiment for this particular child. As each child had a different set of target words, determined based on the pre-test, the comprehension test for the post-test was made individually for each child using a program script in Python.

For the children with only four or five target words, the filler words were also tested. However, these scores were not included in the calculation of the final scores. For each participant, the number of correct answers for the target words were calculated for each condition. Due to the different number of target words, the sum of correct answers of the target words was divided by the number of target words. For each participant, this resulted in two final scores ranging from zero to three, indicating the proportion of correct answers per condition.

Translation tests. Expressive vocabulary knowledge and translation abilities regarding the target words were tested using translation tasks (e.g., Vlaar, Verhagen, Oudgenoeg-Paz, & Leseman, 2017), starting with Turkish to Dutch followed by Dutch to Turkish. These tasks were embedded in a recorded story to activate the target language. In the Turkish to Dutch translation test, participants heard a short introduction in Dutch in which they were instructed to help a Turkish child learning Dutch. This imaginary child said the six target words in Turkish and then participants had to say the translation in Dutch. The Dutch to Turkish translation test contained a Turkish introduction in which participants were instructed to help a Dutch child learning Turkish through providing Turkish translations.

In the Turkish to Dutch translation test, two points were given for correct answers, meaning that the child said exactly the same word as the robot taught them. One point was

given for words with (almost) the same meaning and phonetics, for example the adjective form of the verb. In the Dutch to Turkish translation test, two points were given for the target words or equivalents. One point was given when children did not mention the Turkish verb, but described the situation, for example the target word '*shivering*' was explained as '*I'm feeling very cold so I'm doing this*'. Answers were judged in collaboration with a Turkish speaking researcher.

L1 and L2 receptive vocabulary tests. The Diagnostic Test of Bilingualism, developed by the Dutch national education testing service (Cito), was used to measure children's L1 and L2 receptive vocabulary knowledge (Verhoeven, Narrain, Extra, Konah, & Zerouk, 1995).

During this test, children indicated which picture out of four pictures best matches each test word. The original test contained sixty words tested in both languages in two separate sessions on different days. In our study, testing time was reduced by splitting the test into two equivalent parts of thirty words (similar to e.g., Messer, Leseman, Boom, & Mayo, 2010). Two versions of the test were created. For version one, the odd items were assessed in Dutch and the even items in Turkish. Version two, odd items were assessed in Turkish and even items in Dutch. Following the standard protocol of this test, testing was disrupted after three consecutive wrong items. The sum of correct answers resulted in two final scores (maximum = 30), one for each language. Cronbach's alpha values for the different test parts ranged from .89 to .93.

Analyses

Missing data. There were some missing data in the post-test scores. Due to technical

reasons, one child lacked a score for the comprehension test for the L2-condition and for another child both translations tests at the delayed post-test could not be obtained. Second, experimenter errors lead to missing data of enjoyment for five participants. Last, two children refused to participate within the time available. For one child, scores on the on the comprehension test and translation tests of the immediate post-test were missing. The other child only did not participate in the translation tasks at the immediate post-test.

Calculations and data screening. Because participants were taught different target words, learning gains may be influenced by these differences. For example, children could be provided more difficult words in one condition and easier words in the other condition. Therefore, a *difficulty* variable was computed for each participant, to control for differences in target word difficulty between the conditions. All target words were given a value from one to four, based upon their rating in the BAK list (Mulder, Timman, & Verhallen, 2009) where the words were retrieved from. The higher the score, the more difficult the word. One target word (*descend*) was not stated in the BAK list. We decided to categorize this word in the most difficult category, similar to the rating of its antonym (*ascend*). Our pre-test results also showed that *descend* was one of the words that most children did not know. Then, for every participant, the average difficulty was calculated for each condition. Filler words were again excluded from these calculations. The difficulty score of the words in the L2-only condition was extracted from the difficulty score of the words in the L2-L1 condition. This resulted in one difficulty variable for each participant with a positive score meaning that the words provided in the L2-L1 condition were more difficult than the words provided in the L2-only condition, and a negative score indicating

that words in the L2-only condition were more difficult than the words provided in the L2-L1 condition. A paired-samples *t*-test was conducted to check whether difficulty scores differed between the L2-only condition ($M = 2.52$, $SE = 0.61$) and the L2-L1 condition ($M = 2.52$, $SE = 0.69$). This difference was negligible and not significant ($t(66) = -.062$, $p > .05$).

Furthermore, using two different versions of the Diagnostic Test of Bilingualism required a check whether test scores differed between the two versions for the Dutch ($M = 10.15$, $SD = 6.06$ for version one, and $M = 11.26$, $SD = 5.65$ for version two) and Turkish ($M = 15.79$, $SD = 5.07$ for version one, and $M = 13.76$, $SD = 6.67$ for version two) tests. These differences were again small and statistically not significant for respectively the Dutch test ($t(66) = -.79$, $p > .05$) and the Turkish test ($t(66) = 1.41$, $p > .05$).

Analyses. To test whether learning gains varied between conditions and over time, Repeated-Measures Analyses of Variance were performed. First, a basic 2x2 Repeated-Measures analysis was conducted on children's scores in the post-test with both 'condition' and 'time' as within-subjects factors, controlling for 'difficulty' (differences in target word difficulty between the conditions). In a subsequent analysis, Turkish and Dutch vocabulary were added as covariates to explore possible moderation effects. Last, children's preference for either the monolingual or bilingual robot was added as a between-subjects factor to investigate whether their preference affects their learning gains between conditions. All analyses were conducted using SPSS version 24. Listwise deletion was used to deal with missing data in the Repeated-Measures ANOVA, resulting in a sample size of 65 children.

Results

Table 2 displays descriptive statistics for the target word retention scores on the immediate and delayed post-tests for both conditions. Participants showed learning gains across the L2-only and L2-L1 conditions. Scores on the delayed post-test tended to be higher than on the immediate post-test. Additionally, notice that scores on the translation tasks were rather low.

The assumption of normality was checked for all variables by inspecting histograms and z-scores of Skewness and Kurtosis. The assumption of normality was violated for the L1-only immediate post-test scores, being somewhat negatively skewed with a Kurtosis z-score of -1.22 ($SE = .586, p < .05$). The other variables were also somewhat negatively skewed, but did not violate the assumption of normality. Because F -tests in ANOVA are robust to skewed distributions for sample sizes larger than 30 (Field, 2009), it was assumed that one non-normal score distribution would not be problematic. However, for both translations tests all Skewness z-scores of the variables were significant at $p < .001$, clearly indicating a floor effect on these tasks. The translation tasks were clearly too difficult for a majority of the participants. Therefore, we decided not to include the translation test scores in further analyses.

Table 2
Descriptive statistics of learning gains for both conditions on the immediate (post-test 1) and delayed post-tests (post-test 2)

	Range	L2-only condition		L2-L1 condition	
		Post-test 1 <i>M (SD)</i>	Post-test 2 <i>M (SD)</i>	Post-test 1 <i>M (SD)</i>	Post-test 2 <i>M (SD)</i>
Comprehension	0 - 3	1.93 (0.81)	2.06 (0.85)	1.90 (0.67)	2.09 (0.71)
TK → NL translation	0 - 2	0.25 (0.41)	0.37 (0.58)	0.36 (0.57)	0.47 (0.59)
NL → TK translation	0 - 2	0.31 (0.54)	0.30 (0.51)	0.27 (0.56)	0.29 (0.51)

Learning gains in L2-only and L2-L1 condition

First, a Repeated-Measures ANOVA was performed to investigate whether scores on the comprehension task differed between the conditions and over time. In this a 2x2 Repeated-Measures ANOVA, condition and time were within-subject factors and differences in target word difficulty as covariate.

As shown in Table 3, no effect of condition was found, indicating that the learning gains did not vary between the L2-only and the L2-L1 condition. However, a main effect of time was found; scores on the delayed post-test were significantly higher than on the immediate post-test scores, with a medium effect size (Cohen, 1969). No significant interaction effect between condition and time was found.

Table 3

Results of the 2x2 Repeated-Measures ANOVA on condition and time, and controlling for difficulty.

	$F(1,64)$	p	η^2_p
Condition	0.037	.848	.001
Condition * difficulty	1.230	.272	.019
Time	7.211	.009	.103
Time * difficulty	1.447	.234	.022
Condition * time	0.512	.477	.008
Condition * time * difficulty	1.086	.301	.017

Next, a second Repeated-Measures ANOVA was run, in which Dutch and Turkish vocabulary skills, as assessed with the Diagnostic Test of Bilingualism, were added as covariates to explore possible moderation effects on the main effect of condition. Results are presented in Table 4. The difference between the scores on the immediate and delayed comprehension tests remained significant. Furthermore, the results show that Dutch vocabulary skills did not moderate the relation between condition and learning gains but Turkish vocabulary skills did. This moderation effect is discussed in the next paragraph.

Table 4

Results of the Repeated-Measures ANOVA with possible moderators, and controlling for differences in target word difficulty between the conditions.

	$F(1,61)$	p	η^2_p
Condition	2.666	.108	.042
Condition * difficulty	2.632	.110	.041
Condition * Dutch vocabulary	0.183	.670	.003
Condition * Turkish vocabulary	4.731	.034	.072
Time	5.435	.023	.082
Time * difficulty	2.443	.123	.039
Time * Dutch vocabulary	0.837	.364	.014
Time * Turkish vocabulary	1.106	.297	.018
Time * condition	0.447	.507	.007
Condition * time * difficulty	1.198	.278	.019
Condition * time * Dutch vocabulary	0.241	.625	.004
Condition * time * Turkish vocabulary	0.025	.875	.000

Moderation effect of Turkish vocabulary skills

The interaction effect between condition and Turkish vocabulary skills showed the following disordinal trend: the higher children's vocabulary knowledge in Turkish, the more they learned in the L2-only condition compared to the L2-L1 condition. In contrast, the more limited children's Turkish vocabulary knowledge, the better they performed in the L2-L1 condition compared to the L2-only condition.

Aiming to investigate possible explanations for this effect, data was exploratory inspected to look for differences between children performing high or low on Turkish vocabulary knowledge. A weak correlation ($r = .037$, $p = .77$) between Turkish vocabulary score and the sum of target word difficulty across condition showed that this did not account for differences in learning gains between the conditions. A weak positive correlation was found between Turkish and Dutch vocabulary skills ($r = .203$, $p = .10$). Above all, a significant positive correlation between Turkish vocabulary skills and the absolute difference score between Turkish and Dutch vocabulary knowledge was found (r

= .448, $p < .01$). This showed that children performing low on Turkish vocabulary knowledge tend to have close to similar, thus equally low scores on Dutch vocabulary knowledge (i.e. ‘at risk’ bilingualism, according to the categorization of bilingualism profiles by Francot, Blom, Broekhuizen, & Leseman, under review), whereas the vocabulary knowledge gap between the languages tend to be bigger for children with higher Turkish vocabulary scores (i.e. supportive L1 bilingualism and balanced average bilingualism; Francot et al., under review).

Consequently, the absolute difference score between Turkish and Dutch vocabulary knowledge was added to the 2x2 Repeated-Measures ANOVA model, including the Turkish vocabulary moderator and the difficulty control variable. No significant interaction effect of condition and the difference score was found ($F(1, 61) = 0.183$, $p = .670$, $\eta^2_p = .003$). The three-way interaction of the difference score, condition and time was also non-significant ($F(1, 61) = 0.241$, $p = .625$, $\eta^2_p = .004$). The full model is included in Appendix A.

Children’s enjoyment and robot preferences

Self-reported enjoyment was assessed after each condition to investigate possible differences in children’s enjoyment between the conditions. However, the scores showed a ceiling effect ($M = 3.84$, $SD = 0.06$ for the L2-only condition, and $M = 3.80$, $SD = 0.06$ for the L2-L1 condition). Therefore, these scores were not used in further analyses. None of the participants opted for ‘absolutely not enjoyable’ and only five participants (both conditions together) opted for ‘slightly not enjoyable’.

Asking children with which robot they would like to play with again appeared to be more useful to measure children's preference for the L2-only or L1-L2 condition. Results showed that most of the participants, 48 children (71.6%), preferred to play again with the bilingual robot, and only 19 children (28.4%) stated their preference for the monolingual robot.

To investigate whether children's robot preference affected learning gains between the conditions, preference was added as a between-subjects factor in the 2x2 Repeated-Measures ANOVA. Contrary to our hypothesis, children's robot preference did not affect word learning gains between the conditions ($F(1, 63) = 0.993, p = .323, \eta^2_p = .016$). The interaction effect between condition, time and children's robot preference was also not significant ($F(1, 63) = 1.628, p = .207, \eta^2_p = .026$). The results of this analysis are included in Appendix B.

Discussion

The aim of the current study was to investigate whether providing L1 translations in a L2 vocabulary training facilitated L2 word learning compared to providing L2 vocabulary only. A bilingual Turkish-Dutch and a monolingual Dutch social robot were used to teach 67 Turkish-Dutch kindergartners six words for which they knew the Turkish (L1) but not the Dutch (L2) label. The bilingual robot provided Turkish translations of the target words, whereas the monolingual robot used Dutch only.

First, the findings demonstrated that the experimental task using social robots was effective in teaching vocabulary to the Turkish-Dutch kindergartners. Learning gains were found both in the L2-only and the L2-L1 condition, and it was found that children's target

word comprehension significantly increased over time, possibly pointing to so-called sleeper effects on word learning (Axelsson, Williams, & Horst, 2016). More importantly, our main finding concerned the learning gains across conditions. It was expected that learning gains would be higher in the L2-L1 condition compared to the L2-only conditions. However, no significant differences in learning gains were found between the conditions.

There are several possible explanations for the lack of an effect of condition on learning gains. First, it may be that learning gains were influenced by the number of times that participants were exposed to the Dutch target word in each condition. That is, in the L2-L1 condition, there were 8 instances of each target word, whereas there were 10 instances of each target word provided in the L2-only condition. Although we expected this small difference not to negatively impact the L2 learning process, it may be that it, at least partially, explains the lack of an effect in our study.

Another possible explanation for not finding an effect of condition is related to the finding that children's vocabulary knowledge in Turkish moderates the relation between learning gains and condition. This interaction indicated that the children with well-developed Turkish vocabulary knowledge showed higher learning gains in the L2-L1 condition, compared to children with lower levels of Turkish vocabulary knowledge who showed higher learning gains in the L2-only condition. Consequently, it may be that effects have cancelled each other out due to sampling both participants with low and high Turkish vocabulary proficiencies. That is, if we had only sampled participants with either very low or high Turkish vocabulary proficiencies, an effect may have been obtained.

Additionally, the interaction effect is also an interesting finding in itself because of its counterintuitive direction. As described above, further investigation of the data showed that the children scoring low on Turkish vocabulary knowledge were, generally spoken, ‘at risk bilinguals’, meaning that they performed below average on both Turkish and Dutch vocabulary. In contrast, the children with well-developed Turkish vocabulary knowledge could be categorized either as ‘supportive L1 bilinguals’ or ‘balanced average bilinguals’, the more favorable profiles. Supportive L1 bilingualism is characterized by above average scores in both languages, with a strong L1 support at home. Balanced average bilingualism indicates average L1 proficiencies and slightly above average L2 scores. A wide range of studies (e.g., Cummins, 1976; Bee Chin & Wigglesworth, 2007; Bialystok, 2007) show that being bilingual brings several cognitive advantages, such as enhanced metalinguistic awareness and executive control, which subsequently support language learning. Moreover, these studies demonstrated that such benefits would only occur when sufficient language proficiencies are attained in both languages. Therefore, it could be argued that the supportive L1 and balanced average bilinguals in our study did benefit from these cognitive advantages, whereas the at risk bilinguals may not have reached the threshold yet to be able to benefit from these advantages. An explanation for the interaction effect may thus be that the at risk bilinguals required more support, in the form of providing Turkish translation, when learning L2 vocabulary, whereas the children with well-developed vocabulary skills did not need to rely on this extra support.

Finally, our results showed that all children enjoyed working with the robots, and a distinct majority had a clear preference for the bilingual Turkish-Dutch robot. This is in

line with our hypothesis that the acknowledgement of children's cultural identity through using their L1 in education can increase their enjoyment and wellbeing (Holmes, 2008; Pulinx, van Avermaet & Agirdag, 2017). However, contrary to our hypothesis, this preference did not affect children's learning gains. This may be due to the fact that there was only little variation in our data on children's robot preferences, which makes it difficult to discover such an effect. However, it may possibly increase motivation for future interactions with the robot.

Limitations and future research

This study has several limitations. First, because all participants were taught different target words, variety of factors related to the individual target words may have influenced the results. We tried, as much as possible, to reduce the effect of each children learning different target words. That is, all target words were equally divided between the conditions. Therefore, we could argue that if one child was taught more difficult words in the L2-only condition, then another child was probably taught more difficult words in the L2-L1 condition. Moreover, we also controlled for the differences in target word difficulty between the conditions for each participant to alleviate the concerns. However, we were not able to study, for example, whether more difficult words were learned better depending on the condition they appeared in. All in all, because of the enormous linguistic diversity among our target group, this design limitation was inevitable. Another consequence of this diversity was that 28 children were excluded from participation because not enough possible target words could be selected for them. Due to this decreasing sample, the power of this study was not sufficient to find significant small effects.

A further limitation concerns the pre-test, which was used for the selection of children's target words. The test was slightly restricted by assessing only receptive knowledge of the target words. That is, depth of target word knowledge was not measured. Differences in passive and active target word knowledge in Turkish may have influenced the results. For example, deep and active knowledge of the target words may lead to higher learning gains compared to more superficial target word knowledge only, and its consequences may differ between condition.

Last, the novelty effect of should always be kept in mind with robot interventions consisting of only a single session. Short interventions simply offer participants less time to get used to the new technology (Han, Jo, Jones, & Jo, 2008; van den Berghe et al., under review). To illustrate, multiple RALL studies have shown a clear decrease in engagement and learning gains as time went on (e.g., Fernaeus, Håkansson, Jacobsson, & Ljungblas, 2010; Kanda Hirano, Eaton, & Ishiguro, 2004). In contrary, the two- to three-year-old participants in the study of Hsiao, Chang, Lin and Hsu (2012) sustained their interest over an eight-week lasting intervention. As an attempt to tackle possible novelty effects, we included the group introductions in our study, as suggested by previous studies (e.g., Han et al., 2008). However, the time between the group introductions and the one-to-one session substantially varied between the participants due to planning issues. In our study, the novelty effect could not have affected learning gains between conditions, because the presentation of the conditions was counterbalanced. However, the question is if comparable learning gains were found if our experimental design consisted of multiple training sessions. Ideally, future research should therefore contain multiple training

sessions and aiming to achieve sustained engagement by for example technical improvements such as adaptive behavior.

Despite the limitations, this study has contributed to a growing body of literature on L1 use in L2 teaching, while also exploring the use of social robots in this field. Our results demonstrated that, across both conditions, the experimental task using social robots enhanced L2 word learning among Turkish-Dutch kindergartners. This is an important finding because social robots outweigh human teachers in the ability to speak all children's L1, which is a substantial strength of robot-assisted language learning in linguistically diverse classrooms. Our, somewhat counterintuitively, finding that providing L1 translations in a L2 learning context was more beneficial for children with limited L1 vocabulary skills, shows the complexity of effects of L1 use in L2 learning contexts.

Future research should further explore how L1 use could be incorporated into L2 learning programs to optimally support bilingual children, taking into account the differences in language use and proficiencies in both children's L1 and L2. Recall that robots can be purposefully used to provide L1 input in linguistically diverse classrooms, so future research should also focus on other migrant groups than Turkish-Dutch immigrant children. Concluding, integrating multilingual social robots into language learning programs in superdiverse classrooms holds promises for the future, but more research and technological advancements are needed to accomplish this.

References

- Aarts, R., Demir, S., & Vallen, T. (2011). Characteristics of academic language register occurring in caretaker-child interaction: Development and validation of a coding scheme. *Language Learning*, 61(4), 1173- 1221.
doi:10.1111/j.14679922.2011.00664.x
- Agirdag, O. (2017). Het straffen van meertaligheid: de schaamte voorbij [Punishing multilingualism: past all shame]. In O. Agirdag, & E-R. Kambel (Eds.) *Meertaligheid en onderwijs* (pp. 44-52). Amsterdam: Boom Uitgevers.
- Axelsson, E. L., Williams, S. E., & Horst, J. S. (2016). The Effect of Sleep on Children's Word Retention and Generalization. *Frontiers in Psychology*, 7, 1-9.
<https://doi.org/10.3389/fpsyg.2016.01192>
- Bee Chin, N., & Wigglesworth, G. (2007). *Bilingualism: An advanced resource book*. Routledge Taylor & Francis Group.
- Belpaeme, T., Vogt, P., van den Berghe, R., Bergman, K., Göksun, T., De Haas, M., Kumar Pandey, A. (2018). Guidelines for designing social robots as second language tutors. *International Journal of Social Robotics*. [https://doi.org/10.1007/s12369-0180467-van den Berghe, R., Verhagen, J., Oudgenoeg-Paz., van der Ven, S., & Leseman, P. P. M. \(under review\). Social robots for language learning: A review. *Computer & Education*.](https://doi.org/10.1007/s12369-0180467-van den Berghe, R., Verhagen, J., Oudgenoeg-Paz., van der Ven, S., & Leseman, P. P. M. (under review). Social robots for language learning: A review. Computer & Education.)
- [van den Berghe, R., van der Ven, S., Verhagen, J., Oudgenoeg-Paz, O., Papadopoulos, F., & Leseman, P. P. M. \(in preparation\). Investigating the effects of a robot peer on L2 word learning and enjoyment.](https://doi.org/10.1007/s12369-0180467-van den Berghe, R., van der Ven, S., Verhagen, J., Oudgenoeg-Paz, O., Papadopoulos, F., & Leseman, P. P. M. (in preparation). Investigating the effects of a robot peer on L2 word learning and enjoyment.)

- Bialystok, E. (2007). Cognitive effects of bilingualism: How linguistic experience leads to cognitive change. *International Journal of Bilingual Education and Bilingualism*, 10(3), 210-223. <https://doi-org.proxy.library.uu.nl/10.2167/beb441.0>
- Bouvy, C. (2000). Towards the construction of a theory of cross-linguistic transfer. In J. Cenoz & U. Jessner, *English in Europe: The acquisition of a third language* (pp. 143- 156). Multilingual Matters.
- Centraal Bureau voor de Statistiek [CBS]. (2016). *Jaarrapport integratie 2016* [Yearly integrationreport 2016]. Den Haag: Centraal Bureau voor de Statistiek.
- Centraal Bureau voor de Statistiek [CBS]. (2017). Bevolking; kerncijfers [Population; key data]. Retrieved from <http://statline.cbs.nl>
- Celik, M. (2003). Teaching vocabulary through code-mixing. *English Language Teaching*, 57(4), 361-369. [doi:10.1093/elt/57.4.361](https://doi.org/10.1093/elt/57.4.361)
- Cohen, J. (1969). *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press.
- Cole, S. (1998). The use of L1 in communicative English classrooms. *Japan Association for Language Teaching*, 22, 11-14.
- Creese, A., & Blackledge, A. (2008). Flexible bilingualism in heritage language schools. Paper presented at Urban Multilingualism and Intercultural Communication, Antwerp, Belgium.
- Creese, A., & Blackledge, A. (2015). Translanguaging and identity in educational settings. *Annual Review of Applied Linguistics*, 35, 20-35. doi:10.1017/S0267190514000233

- Cummins, J. (1976). The influence of bilingualism on cognitive growth: A synthesis of research findings and explanatory hypotheses. *Working Papers on Bilingualism*, 9, 1- 43.
- Cummins, J. (1981). Bilingualism and minority language children. Ontario: Ontario Institute for Studies in Education.
- Cummins, J. (2005). A proposal for action: Strategies for recognizing heritage language competence as a learning resource within the mainstream classroom. *Modern Language Journal*, 89(4), 585–592.
- Cummins, J. (2008). Teaching for transfer: Challenging the two solitudes assumption in bilingual education. In J. Cummins & N. Hornberger (Eds.), *Encyclopaedia of Language and Education* (pp. 65-75). Boston: Springer. doi:10.1007/978-0-387 30424-3_116
- Demir-Vegter, S., Aarts, R., & Kurvers, J. (2014). Lexical richness in maternal input and vocabulary development of Turkish preschoolers in the Netherlands. *Journal of Psycholinguistic Research*, 43(2), 149-165. doi:10.1007/s10936-013-9245-7
- Duarte, J., & Gogolin, I. (2013). Introduction: Linguistic superdiversity in educational institutions. In J.Duarte & I. Gogolin (Eds.), *Linguistic superdiversity in urban areas: Research approaches* (pp. 1-26). Amsterdam: John Benjamins B.V.
- Ellis, N.C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, 27(2), 164-194.
doi:10.1093/applin/aml015

Extra, G., Aarts, R., van der Avoird, T., Broeder, P., & Yağmur, K. (2002). *De andere talen van Nederland: Thuis en op school* [The other languages of the Netherlands: at home and at school]. Bussum: Coutinho.

Fernaes, Y., Håkansson, M., Jacobsson, M., & Ljungblad, S. (2010). How do you Play with a Robotic Toy Animal ? A long-term study of Pleo. In *Proceedings of the 9th international Conference on Interaction Design and Children* (pp. 39–48).

doi:[10.1145/1810543.1810549](https://doi.org/10.1145/1810543.1810549)

Field (2009). *Discovering statistics using SPSS*. London: SAGE publications.

Francot, R., Blom, W. B. T., Broekhuizen, M. L., & Leseman, P. P. M. (under review).

Changing profiles of bilingualism in early childhood: A person-centred latent class transition approach. *Applied Linguistics*.

Ghorashi, H., & van Tilburg, M. (2006). ‘When is my Dutch good enough?’ Experiences of refugee women with Dutch labour organisations. *Journal of International Migration and Integration*, 7(1), 51-70. doi:10.1007/s12134-006-1002-4

Gogolin, I. (2011). The challenge of super diversity for education in Europe. *Education Inquiry*, 2(2), 239-249. <https://doi.org/10.3402/edui.v2i2.21976>

Gravelle, M. (1996). *Supporting bilingual learners in schools*. London: Trentham Books.

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. <https://doi.org/10.3745/JIPS.2008.4.4.159>

Hartgers, M. (2012). Onderwijs [Education]. In: R. van der Vliet, en J. Ooijevaar en R. van der Bie (Eds.), *Jaarrapport Integratie 2012* (pp. 69–92). Den Haag/Heerlen:

Centraal Bureau voor de Statistiek.

Heller, M. (1999). *Linguistic minorities and modernity: A sociolinguistic ethnography*.

London: Longman.

Holmes, J. (2013). *An introduction to sociolinguistics* (3rd ed.). England: Pearson

Longman.

Hsiao, H. S., Chang, C. S., Lin, C. Y., & Hsu, H. L. (2012). “iRobiQ”: The influence of bidirectional interaction on kindergarteners’ reading motivation, literacy, and behavior. *Interactive Learning Environments*, 23(3), 269–292.

<https://doi.org/10.1080/10494820.2012.745435>

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. <https://doi.org/10.1080/07370024.2004.9667340>

Kanero, J., Geçkin, V., Oranç, C., Mamus, E., Küntay, A. C., & Göksun, T. (2018) Social Robots for Early Language Learning: Current Evidence and Future Directions. *Child Development Perspectives*. doi:10.1111/cdep.12277

Kory Westlund, J., 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* (pp. 7–8).

Le Page, R. B., & Tabouret-Keller, A. (1985). *Acts of identity: Creole-based approaches to language and ethnicity*. England: Cambridge University Press.

Leseman, P. P. M., Henrichs, L. F., Blom, E. & Verhagen, J. (in press). Young mono- and bilingual children’s exposure to academic language as related to language

- development and school achievement. In V. Grøver, P. Ucelli, M. Rowe & E. Lieven (Eds.), *Learning through language*. Cambridge, MA: Cambridge University Press.
- Lewis, G., Jones, B., & Baker, C. (2012a). Translanguaging: Origins and development from school to street and beyond. *Educational Research and Evaluation*, 18(7), 641-654. doi:10.1080/13803611.2012.718488
- 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 (CogSci 2012)* (pp1882-1887).
- Liu, S., Liao, H., & Pratt, J. A. (2009). Impact of media richness and flow on e-learning technology acceptance. *Computers & Education*, 52(3), 599–607. <https://doi.org/10.1016/j.compedu.2008.11.002>
- Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research*, 80(3), 300-335. doi:10.3102/0034654310377087
- Mayo, A. Y., & Leseman, P. P. M. (2008). Off to a Good Start? Vocabulary development and characteristics of early family and classroom experiences of children from native Dutch speaking and bilingual minority families in the Netherlands. *Educational and Child Psychology*, 25(3), 66-78.
- Messer, M. H., Leseman, P. P. M., Boom, J., & Mayo, A. Y. (2010). Phonotactic probability effects in nonword recall and its relationship with vocabulary in monolingual and bilingual preschoolers. *Journal of Experimental Child Psychology*, 105(4), 306–323. doi:10.1016/j.jecp.2009.12.006

- Mulder, F., Timman, Y., & Verhallen, S. (2009). Handreiking bij de Basiswoordenlijst Amsterdamse Kleuters (BAK) [Guidelines for the basic word list for kindergartners in Amsterdam]. Amsterdam: ITTA and UvA.
- OECD. (2015). Immigrant Students at School: Easing the Journey towards Integration, *OECD Reviews of Migrant Education*, Paris: OECD Publishing.
<http://dx.doi.org/10.1787/9789264249509-en>
- Oudgenoeg-Paz, O., Verhagen, J., & Vlaar, R. (2017). Deliverable 1.1: lesson series for three domains. L2tor project European Commission.
- Ministry of Education, PO-Raad, & Lectoreninitiatief Professionalisering Taalonderwijs Nieuwkomers. (2017). Ruimte voor nieuwe talenten: keuzes rond nieuwkomers op de basisschool [Opportunities for new talents: choices concerning newcomers at primary education].
- Probyn, M. (2005). Language and the struggle to learn: The intersection of classroom realities, language policy, and neocolonial and globalisation discourses in South African schools. In: Lin, A. M. Y., Martin, P.W. *Decolonisation, globalisation: Language-in-education policy and practice* (pp153-172). Multilingual Matters.
- Pulinx, R., van Avermaet, P., & Agirdag, O. (2017). Silencing linguistic diversity: The extent, the determinants and consequences of the monolingual beliefs of Flemish teachers. *International Journal of Bilingual Education and Bilingualism*, 20(5), 542- 556. <https://doi.org/10.1080/13670050.2015.1102860>
- Scheele, A. F., Leseman, P. P. M., & Mayo, A. Y. (2010). The home language environment of monolingual and bilingual children and their language proficiency.

- Applied Psycholinguistics*, 31(1), 117-140. doi:10.1017/S0142716409990191
- Snow, C. (1983). Literacy and language: Relationships during the preschool years. *Harvard Educational Review*, 53(2), 165-189.
- Swain, M., & Lapkin, S. (2013). A Vygotskian sociocultural perspective on immersion education: The L1/L2 debate. *Journal of Immersion and Content-based Language Education*, 1(1), 101-129. doi:10.1075/jicb.1.1.05swa
- Vedder, P., & van Geel, M. (2017). Cultural identity development as a developmental resource. In: Cabrera N.J., Leyendecker B. (Eds.), *Handbook on Positive Development of Minority Children and Youth* (pp.123-137). Cham, Switzerland: Springer.
- Verhoeven, L. (2007). Early bilingualism, language transfer, and phonological awareness. *Applied Psycholinguistics*, 28(3), 425–439. doi:10.1017.S0142716407070233
- Verhoeven, L., Narrain, G., Extra, G., Konak, O. A., & Zerrouk, R. (1995). Diagnostische Toets Tweetaligheid (DTT) [Diagnostic Test of Bilingualism]. Arnhem, The Netherlands: CITO.
- Vlaar, R., Verhagen, J., Oudgenoeg-Paz, O., & Leseman, P. P. M. (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. doi:10.475/123 4

Appendix A

Results of the Repeated-Measures ANOVA exploring moderation of absolute difference scores in Dutch and Turkish vocabulary skills, controlling for differences in target word difficulty between the conditions

	$F(1,61)$	p	η^2_p
Condition	2.666	.108	.042
Condition * difference vocabulary skills	0.183	.670	.003
Condition * Turkish vocabulary	1.920	.171	.031
Condition * difficulty	2.632	.110	.041
Time	5.435	.023	.082
Time * difference vocabulary skills	0.837	.364	.014
Time * Turkish vocabulary	2.398	.127	.038
Time * difficulty	2.443	.123	.039
Time * condition	0.447	.507	.007
Condition * time * difference vocabulary skills	0.241	.625	.004
Condition * time * Turkish vocabulary	0.260	.612	.004
Condition * time * difficulty	1.198	.278	.019

Appendix B

Results of the 2x2 Repeated-Measures ANOVA with robot preference as between-factor, Turkish vocabulary knowledge as moderator, and controlling for differences in target word difficulty between the conditions

	<i>F</i> (1,63)	<i>p</i>	η^2_p
Condition	3.120	.082	.049
Condition * Turkish vocabulary	4.212	.044	.065
Condition * difficulty	2.304	.134	.036
Condition * robot preference	0.993	.323	.016
Time	5.628	.021	.084
Time * Turkish vocabulary	1.825	.182	.029
Time * difficulty	2.710	.105	.043
Time * robot preference	1.309	.257	.021
Condition * time	0.032	.858	.001
Condition * time * Turkish vocabulary	0.022	.882	.000
Condition * time * difficulty	0.783	.380	.013
Condition * time * robot preference	1.628	.207	.026