LEARNING TO EAT HEALTHILY WITH A ROBOT

Creating and evaluating a dialogue design for social robots to support children in learning about healthy nutrition by stimulating reflection

Master's Thesis
Human Computer Interaction

M.K.N. van den Berg Bsc
6084184

Internship supervisor: Prof. Dr. M.A. Neerincx
University supervisor / first examiner: Prof. Dr. Ir. J.F.M. Masthoff
Second examiner: Dr. C. van Nimwegen

July 14, 2022

ABSTRACT

With the current shortage of teachers, virtual agents and social robots can provide a solution by supporting pupils in learning independently. They can do so in several manners, where stimulating reflection is considered particularly relevant. With only half of Dutch children meeting nutrition guidelines, health literacy is considered an important learning domain. However, to date it remains unclear how the interaction between agent/robot and child should be designed, especially in those domains. A systematic literature review showed that agents and robots are employed in various domains with various behaviors, but that only a small number of papers focus on reflection and/or health. To design a learning session with a robot that incorporated both, a human-centered design approach was adopted. This involved the creation of personas, scenarios, and reflection strategies, which resulted in a dialogue design for a learning session about healthy nutrition. The reflection strategies were evaluated during a between-subjects experiment (N = 28), consisting of a condition in which a social robot stimulated reflection on learning goals and learning progress, and a control condition, in which these reflection strategies were not employed. The outcome measures consisted of 1) recall of the learning goals, 2) learning outcomes, and 3) motivation to eat healthily. The former two measures were assessed on two occasions: directly after the session (short-term) and one week later (long-term). No significant results were found. However, the control condition showed a stronger decrease in scores over time. Future work is recommended to employ multiple sessions with the robot to stimulate reflection more effectively and to assess long-term effects.

Keywords: Social Robots, Virtual Agents, Healthy Nutrition, Dialogue, Interaction, Human-Centered Design

TABLE OF CONTENTS

1 INTRODUCTION

In The Netherlands, the average number of children in a primary school class is between 23 and 24, a number that has risen over the last years (Rijksoverheid, 2022). In classes of such sizes, it can be difficult for the teacher to give each child the amount of attention that they deserve. Some children may struggle to work independently. They get stuck or are unable to regulate their motivation (Azevedo, Cromley, Moos, Greene, & Winters, 2011). This problem has become more apparent during the COVID-19 pandemic, during which children had to follow education from home. The current shortage of primary school teachers in The Netherlands is another factor that puts extra pressure on this issue (Onderwijsinspectie, n.d.).

Social robots and virtual agents can offer a solution to this problem. They can take various roles, such as tutor, tutee, buddy, motivator, or facilitator. During one-on-one interaction, the robot or agent can track children's progress and mental state and employ personalized interventions. Research has shown that such one-on-one interactions can increase motivation (e.g. Song, Barakova, Markopoulos, & Ham, 2021), improve learning outcomes (e.g. Konijn & Hoorn, 2020), and support self-regulation skills (e.g. Jones & Castellano, 2018).

A social robot or virtual agent can support independent learning in several manners. For instance, it can facilitate the establishment of learning goals and the generation of feedback, the two processes required for self-monitoring (Meyer, Haywood, Sachdev, & Faraday, 2008). For example, the robot or agent can help the learner to explicate the learning goals at the start of the learning session. Multiple studies have shown that providing the learner with specific and attainable goals aids performance compared to a situation where the learner is provided with no goals (Lunenburg, 2011). In addition, the robot or agent can employ a memory of previous interactions with the learner to track their progress towards these learning goals. This can aid the process of reflection, which is also an important component in independent learning (Pintrich, 2000; Zimmerman, 2002).

Social robots and virtual agents can support learning processes in various domains. One that is particularly relevant is health education. Educating children in physical and mental health is beneficial for children's future and contributes to a healthy society (Occupational Medicine & Health Affairs, n.d.). A key contributor to children's health is nutrition (Melanson, 2008). According to research by "het Nederlands Jeugdinstituut" (2022a), only half of Dutch children between four and 12 years old met the nutrition guidelines by "het Voedingscentrum". Less than 47 percent eats vegetables seven days per week. A little over 50 percent eats fruit five days a week. It must be said that the percentages have improved with respect to the year before (2020). However, at the same time an increase was found in the percentage of children in the same age group who are overweight (Nederlands Jeugdinstituut, 2022b). In 2021, 15.5 percent of Dutch children between four and 12 were overweight, whereas this percentage was 13.2 percent in 2020. Research shows that unhealthy nutrition and disrupted eating behaviors can lead to lagging growth, underweight, overweight, intestinal problems, and tooth decay (Florence, 2008; Taras, 2005; Hoyland, 2008; De Wit, 1994, as cited in Nederlands Jeugdinstituut, 2017). Healthy eating habits, as part of a healthy lifestyle, are positively related to cognitive development and educational performance (Florence, 2008; Hoyland, 2008, as cited in Nederlands Jeugdinstituut, 2017). Furthermore, disrupted eating behaviors can pose a heavy burden on the social-emotional wellbeing of a child and their parents (Chatoor, 2004, as cited in Nederlands Jeugdinstituut, 2017).

Health literacy is one of the objectives of the ePartners4all project[1]. This project aims to develop robot buddies or virtual agents, so called ePartners, to enhance children's health and well-being. The outcomes of this project can help prevent health problems in at-risk children and thus create a more resilient society. In turn, this could lead to lower healthcare utilization (Health~Holland, n.d.). Although the current research project is conducted independently, its aim is to provide results that are useful for this project.

Although this area of research sounds promising, there are only a few studies that address this topic. To date, there are no design guidelines for agents or robots that are employed in this field. Therefore, this research project aims to contribute to the dialogue design of social robots and virtual agents to improve health literacy in children between seven and 12 years old. More specifically, this study aims to investigate whether reflection strategies, delivered by a social robot, can increase the effectiveness of children's independent learning processes in the domain of healthy nutrition. This study aims to do so by answering three research questions (RQs), one regarding literature (RQ1), one regarding design (RQ2), and one regarding evaluation (RQ3):

***RQ1**: What is the current state of the art in dialogue design research on virtual agents and social robots that support learning processes of children (7-12 years) through one-on-one interaction?*

***RQ2:** How can a social robot use dialogue to encourage children (7-12 years) to reflect on learning goals and learning progress when learning individually about healthy nutrition?*

***RQ3:** Does the dialogue design for reflection on learning goals and learning progress lead to better reproduction of the learning goals, higher learning outcomes, and more motivation to show healthy behavior?*

These research questions are discussed in Chapter 2, 3, and, 4+5, respectively. Chapter 6 presents the conclusions, limitations, and future work recommendations of this work.

[1] See e.g. https://www.health-holland.com/project/2022/2021/data-driven-blended-care-solution-virtual-buddy-child-health

2 LITERATURE

2.1 Topic, objective and subquestions

A systematic literature review was performed to provide an answer to RQ1: *What is the current state of the art in dialogue design research on virtual agents and social robots that support learning processes of children (7-12 years) through one-on-one interaction?* It was conducted in accordance to the PRISMA 2020 protocol guidelines (see Page et al., 2021). The topic of this review was defined as follows: "Interaction design of virtual agents and social robots to verbally support learning processes of children between seven and 12 years old through one-on-one interaction". The objective was to provide an overview of the current state of the art of the interaction design of virtual agents and social robots that verbally support learning processes of children between seven and 12 years old through one-on-one interaction. To aid this objective, the questions in Table 1 were posed.

Table 1. Subquestions of systematic literature review.

Q1	What type of embodiment is used for the agent or robot?
Q2	What role does the agent or robot fulfill? (i.e., tutor, tutee, or peer[2])?
Q3	In what learning domain is the agent or robot implemented?
Q4	What is the intended effect of the agent or robot?
Q5	What is the specific learning intervention that is facilitated by the agent or robot?

The review consisted of five stages: 1) an identification phase, in which a search query was ran in an online database, 2) a screening phase, during which all abstracts were assessed using exclusion and inclusion criteria, 3) a retrieval phase, during which the full papers were retrieved, 4) an eligibility assessment phase, during which the full texts were screened using the criteria, and finally 5) a backwards retrieval phase, during which snowballing was performed on the selected papers. See Figure 1 below. All phases and the review's results will be discussed in more detail in Section 2.4.

[2] Tutor: instructs and/or provides (corrective) feedback, monitors learning process/progress, more proficient than learner
Peer: an equal companion or opponent, does not teach or guide the learning process, same proficiency level as learner
Tutee: receives instructions and corrective feedback from child, less proficient than child

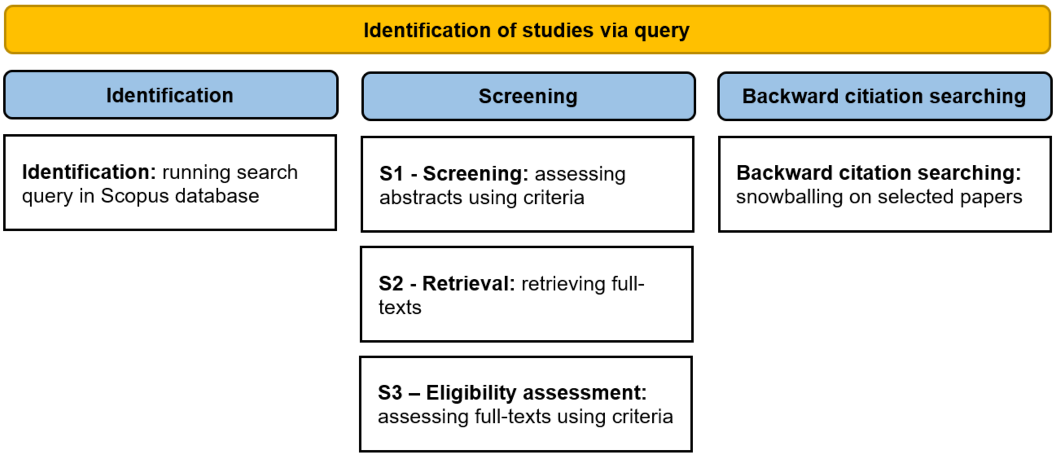


Figure 1. Phases of systematic literature review.

2.2 Information source and search strategy

Scopus[3] was selected as the only information source for this literature review. There were three reasons for this. Firstly, Scopus allows for advanced customizations of queries. Boolean operators and filter terms can be used to specify combinations and conditions. In addition, it offers many filters using automatic detection. Secondly, in contrast to Google Scholar, Scopus allows for downloading the information of all records at once. This is essential for loading all abstracts into Rayyan[4], the tool that was used for this systematic literature review. Finally, Scopus searches a larger database than other considered search engines. To illustrate this: recent research showed that about 99.11 percent of the journals indexed in Web of Science are also indexed in Scopus (Singh, Singh, Karmakar, Leta, & Mayr, 2021). This high percentage was also the reason to not include Web of Science as an additional data source.

A query was composed to retrieve the relevant records from the online Scopus database. After several iterations, the final query was defined as follows:

TITLE-ABS-KEY ((("Virtual agent*" OR "Intelligent agent*" OR "Conversational agent*" OR "Pedagogical agent*" OR "Virtual assistant*" OR "social robot*" OR hri) OR ((companion* OR budd*) AND (technolog* OR virtual OR interacti*)) **AND** (child* OR "primary school*" OR "primary education" OR "elementary school*" OR "elementary education" OR "middle school*") **OR** "Child?robot interaction" OR "Child?agent interaction")
AND (learn* OR education OR tutor* OR coach*))
AND PUBYEAR > 2015 AND DOCTYPE (cp OR ar OR ch) AND LANGUAGE (english)

Logical operators (such as AND) were used to combine or exclude specific terms or characteristics. For example, the OR-operator made sure that each of the documents contained either the word “learn”, “education”, “tutor”, “coach”, or a combination of those words. For some keywords, the asterisk (*) was used to include multiple forms of the word. For instance, interacti* was used to include the terms interactive, interaction and interactions.

The colors of the different terms indicate the relationships within the query. This is used to structure the following explanations:

[3] https://www.scopus.com/
[4] https://www.rayyan.ai/

Medium blue text: Since virtual agents and social robots are both used to support children in learning processes, both were included in the query. It was expected that these two can share the same interaction designs, functionalities, and capabilities.

Several terms were used to describe the concept "virtual agent". A preliminary literature search indicated that many synonyms exist. In the end, the following terms were included: *virtual agent, intelligent agent, conversational agent* (this would also incorporate *embodied conversational agents* and *pedagogical conversational agents* in the query results), *pedagogical agent*, and *virtual assistant*.

Concerning the robots, the following terms were used: *social robot, companion,* and *buddy*. The conscious choice was made to not simply use the term "robot". This was done for two reasons: 1) this term appeared to be too general, including results that discussed education about robotics, 2) this resulted in another 1,000 hits, which were too many to filter through in the given time frame. The term "educational robot" was also considered, but that also resulted in a too many extra hits. It was finally decided that social robot was the best option, because it stresses the human side of interaction and increases the chances of targeting robots that support the learner through dialogue (i.e., verbally).

Light blue text: A preliminary literature search showed that papers with the terms *companion* and *buddy* often addressed living companions or buddies, such as a pet. Therefore, they had to coincide with technology, virtual, or interaction.

Dark blue text: All terms describing a virtual agent or robot should co-occur with at least one of the following terms: *child, primary school, primary education, elementary school, elementary education*, or *middle school*. This was aimed at retrieving records that addressed agents or robots designed for primary school-aged children. The term pupils was also considered. However, it appeared that most results referred to the pupils of the eye. This term was therefore excluded in the end. At a later stage, synonyms of primary school, such as elementary school, were added to make sure the query was inclusive enough.

Purple text: The terms *child-robot interaction* and *child-agent interaction* were included separate from the previously mentioned terms for agent/robot and children/school. The reason was that these two terms already contain both aspects.

Green text: This part made sure that all above terms (agent/robot and children/school) were combined with a term related to learning. More specifically, the following terms were used: *learn, education, tutor,* and *coach*. The term "teach" was not selected to exclude agents or robots that are aimed at transferring knowledge only.

Orange text: Finally, some filters were applied. This ensured that only recent papers were considered (published after 2015), that all papers were written in English, and that all papers were either an article, a conference article, or a book chapter (see C 11 in Table 2). The query was run on December 20, 2021. Hence, the query result was limited to papers published or accepted between January 1st, 2016 and December 20, 2021.

2.3 Exclusion and inclusion criteria

Table 2 shows the exclusion and inclusion criteria used in the systematic literature review. They were composed to ensure that the resulting papers met the objective of this review (see Section 2.1). The criteria were divided over the different phases of the process, which will be elaborated upon in Section 2.4.

Table 2. Exclusion and inclusion criteria.

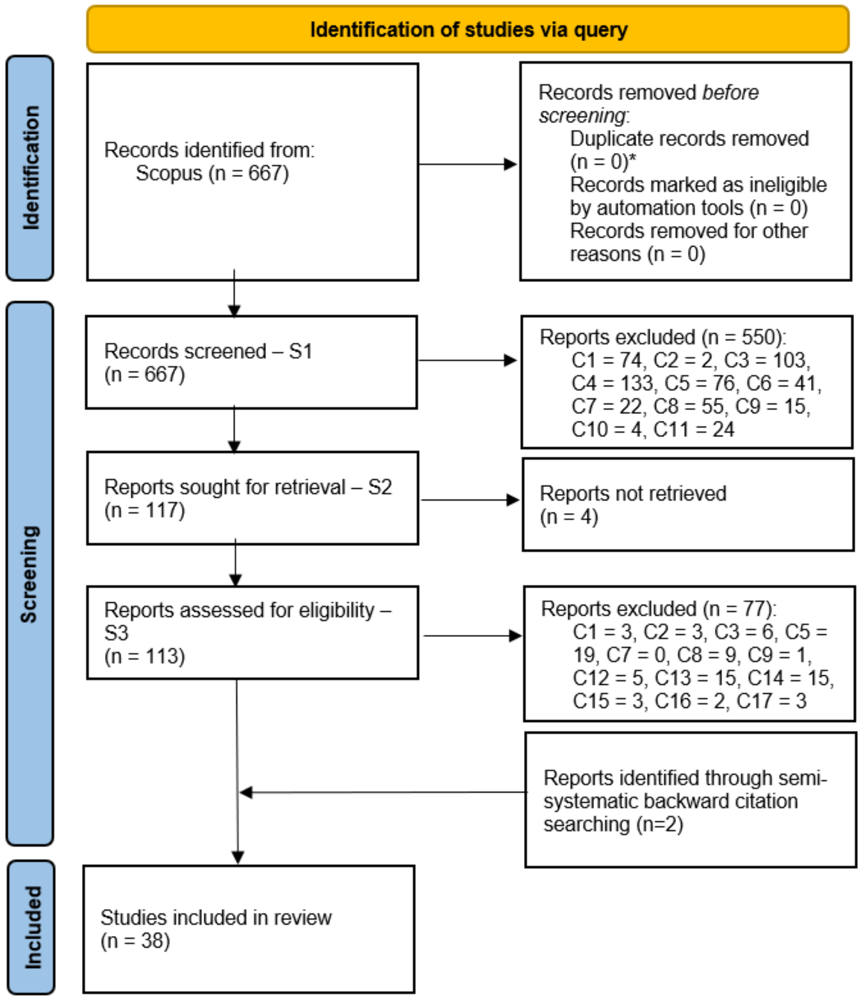
Exclusion criterium for phase S1 - Screening (abstract only)	
C1	A (single) educational virtual agent or social robot is <u>not</u> central in the paper
	Since this is the topic of the literature study, it needs to be the central topic in the acquired papers as well. Examples of papers that were excluded based on this criterium are: papers discussing the role of companion animals for kids and papers about the support of mothers in Third World countries.
C2	The agent or robot lacks embodiment
	This restriction was made because (virtual) embodiment makes it easier for children to see the agent or robot as a social entity (Davison et al., 2021). This is important in the development of a bond between them. For instance, papers that mentioned Amazon Alexa (a voice-based assistant) were excluded.
C3	The agent or robot does <u>not</u> aim to support children in achieving learning goals through 1-on-1 interaction • Note that this excludes robots that replace a teacher lecturing a class (one-to-many) • Note that this also excludes children learning to program a robot
	This criterium ensures that the selected papers discuss a coach or peer that supports children in individual learning processes. It is relevant that the agent or robot can support a child learning at home. Examples of excluded papers are those that discuss a robot teacher lecturing a class (which is not one-on-one) and papers that discuss teaching children how to program a robot. Note that this criterium also excludes papers that do not discuss a learning process, such as playing a simple game with a robot.
C4	The agent or robot is aimed at a specific treatment or special education • Examples include social therapy for children with ASD, rehabilitation exercises for children with physical disabilities, and social robots assisting children with hearing impairments • Does not exclude groups of children that require a general change in lifestyle or attitudes, (such as children at risk for obesity)
	Since this research project is not targeted at a specific group of children, those that are should be excluded. For instance, excluded papers discuss social therapy for children with autism spectrum disorder, rehabilitation exercises for children with physical disabilities, and social robots assisting children with hearing impairments. This criterium does not exclude groups of children that require a general change in lifestyle or attitudes, such as children at risk for obesity.
C5	The age of the target group and/or participants is below seven or above 12 years old • Exceptions can be made when the age range is e.g. 10-14, as long as the majority of the age group and/or the average age is between seven and 12 years.
	This age group was selected, because it includes the ages during which children are able to read and write (in The Netherlands), while excluding the ages where children enter puberty. In The Netherlands, this equivalates the age range of grade 4 until grade 8 of primary school. After grade 8, children enter high school. From this point onwards a differentiation in attainment levels is made, which might make it harder to design an agent or robot for all children.
C6	The paper is purely focused on technical aspects, such as system implementation, sensors' detection accuracy, or digital ecosystem design

	Since the focus of this literature study is on interaction design, papers that only focus on technical aspects are excluded. An example is a paper that presents an improved facial expression detector for a robot.
C7	The paper is purely focused on ethical aspects, such as the acceptance of robots in a classroom
	Since the focus of this literature study is on interaction design, papers that only focus on ethical aspects are excluded. An example is a paper that discusses interviews with parents on their view on robots replacing teachers in the classroom.
C8	The paper does not include an evaluation of the agent or robot, during which children interact with this agent or robot • Pilot studies count as evaluation, and should therefore be included
	It is considered important that an evaluation is present to verify the proposed interaction design of the paper. The validity of this evaluation is strengthened if interaction with children has taken place. In the end, they are the targeted users. Pilot studies can be included, since they provide an indication of the effectiveness of the agent or robot. In addition, it might be hard to distinguish pilot studies from non-pilot studies.
C9	The evaluation only measures the effect of the static appearance of the agent or robot (e.g., voice, facial expression, embodiment style) or the modality of robot expressions (e.g., gestures, paraverbal cues)
	Since the topic of this literature study is interaction design (and not static or predefined design), the appearance and modality of robot expressions are not considered relevant. An example of an excluded paper is one that compares a robot with a human voice to a robot with a robot voice.
C10	The interaction with the agent or robot is clearly aimed at one-time usage • An example is an agent or robot as an information guide at a museum
	The aim of this research project is to contribute to the design of an agent or robot that supports children in long-term learning processes. An agent or robot that is aimed at one-time usage is therefore not relevant. An example is a virtual agent that provides information at an interactive museum exhibit.
C11	The paper is not a (full) article, conference paper or book chapter • Other documents should have been filtered out by the search engine, but reviews still have to be excluded.
	Only journal articles, conference papers and book chapters were selected as appropriate formats. The aim was to find papers that discussed a specific interaction design that the authors based on literature and which they evaluated. For this reason, full books, conference reviews, and literature reviews were excluded. Another reason to exclude the latter was that this literature review was not supposed to be a metareview. In addition, the review papers may be focused on "outdated" papers (i.e., those published before 2015).
Exclusion criteria for phase S2 - Retrieval	
C12	The full text cannot be accessed
	It is important to be able to read the full contents of the paper. Otherwise, relevant information cannot be extracted. If needed and possible, access to the full text was requested via Researchgate.net. In case there was no response within one month, the paper was excluded from the final dataset.
Exclusion criteria for phase S3 – Eligibility Assessment (full-text)	
All of the above	
Inclusion criteria for phase S3 - Eligibility Assessment (full-text)	
C13	In case multiple papers discuss the same research, only keep the publication with the highest quality

	• Take into account which paper is the most recent, the most detailed, the longest, or published as an article (preferably all four) • Some later papers may include a previous study together with a new study. Then exclude the paper that only discusses the previous study. • Small differences may be ignored, such as a version with a pilot evaluation and a version with an extensive evaluation.
	It is redundant to keep multiple papers discussing the same research. An example is a conference paper that proceeded the journal article about the same research.
C14	The paper is at least 5 pages long (including references)
	The level of depth of short papers is considered insufficient. It is assumed that important background information and clarification is missing in case the text is brief.
C15	The agent or robot interacts with the child verbally • This can either be text-based or speech-based • This does not mean that the child has to give verbal input, as long as the robot responds to the child('s actions) in verbal form
	Non-verbal interactions alone are not seen as relevant. They are not expected to be capable of conveying the type of information that this research project aims to provide (i.e., information in the domain of health education). For instance, a study was excluded where the robot indicated information using gaze and movement only.
C16	The paper includes least one study investigating the effect of a specific interaction design (ID), functionality or capability of the agent or robot, or a combination of these IDs, functionalities or capabilities
	As this research project aims to measure the effect of such an interaction design, functionality or capability, it is key that the selected papers have done the same. An example of an excluded paper is one where the virtual agent did not occur in any of the research questions or hypotheses.
C17	The evaluations measures include learning effects (learning in the broadest sense, including understanding, motivation, attitude, performance, creativity, behavior change, time spent studying)
	Similarly, this research project aims to measure the learning effect of the previously mentioned interaction design, functionality, or capability. Therefore, the selected papers should do the same. For example, an evaluation that only measured children's acceptance of an educational robot did not suffice.
Inclusion and exclusion criteria for Backward citation searching (full-text)	
All of the above	

2.4 Selection process

See Figure 2 and the text below for the results of the selection process.



* Two papers were marked as duplicates, but manual inspection by the author resulted in the conclusion that these were separate papers

Figure 2. PRISMA flow diagram.

Identification phase. The final query (see Section 2.2) was run in the Scopus search engine and yielded 667 hits on December 20, 2021. These results were exported and loaded into Rayyan.ai, an online tool for systematic literature reviews (Ouzzani, Hammady, Fedorowicz, & Elmagarmid, 2016).

Screening phase – S1. The first part of the screening phase consisted of reading the abstracts of all 667 papers in Rayyan.ai. Each title was either included or excluded based on the composed exclusion criteria (see Section 2.3). For a large share of the records, multiple exclusion criteria applied. To save time, only one of these criteria was used as a label. This was often the most obvious criterium. In total, the author labeled 116 titles as included and 551 as excluded.

To increase validity, a partial second check was performed by other (semi-)experts in the domain. One of them was a professor in human-computer interaction (rater 1) and the other was a master student in artificial intelligence (rater 2). Both were involved with the topic of this research project. Rater 1 had prior experience with systematic literature reviews and Rayyan.ai, while rater 2 did not. They were handed a short instruction guide on how to use the tool and a list with the exclusion criteria for phase S1.

Since resources were limited, it was not possible to perform a full check on all 667 abstracts. The two raters used a semi-random approach to select the abstracts to rate. Since the author did not know this selection at the time that she performed the initial rating, no bias was present. The aim was to have as many papers checked as possible, but technical and communicational issues made it difficult to avoid that some papers were checked by both raters.

In total, rater 1 rated 281 abstracts and rater 2 rated 50 abstracts. Thirty-seven abstracts were rated by both of them. In total, 20 conflicts were detected. Most conflicts (9x) revolved around criterium 3, *The agent or robot does not aim to support children in achieving learning goals through 1-on-1 interaction,* where the main conflict seemed to be whether a topic involved a learning goal. An example is a study that measures to what extent children copy (= learn?) prosocial behavior from a prosocial robot. In addition, the raters did not always agree whether it was clear if the interaction was one-on-one or not. Another criterium that occurred often (6x) was criterium 8, *The paper does not include an evaluation of the agent or robot, during which children interact with this agent or robot.* It appeared that it was sometimes difficult to judge whether an evaluation was present, based on the content of the abstract only. Furthermore, in three cases the raters did not agree whether the paper targeted children with special needs (criterium 4). Further, in two cases one of the raters excluded the paper based on the age of the children (criterium 5) when the term "preschool" was mentioned. The author had not noticed this. In one case, the rater disagreed with the author on whether the focus was on technological aspects only.

The findings above were used to reconsider the papers that showed a conflict. In case the paper was accepted by the author but not by the raters (6x), the author would check if the paper was excluded in phase S3. It turned out that this was always the case. Moreover, the exclusion criterium selected by the raters in phase S1 was almost always the same as the criterium assigned by the author in phase S3. In case the paper was accepted by the raters but rejected by the author (14x), the full paper was checked to gather more information to decide whether the exclusion criterium was valid or not. If the criterium was indeed valid, the paper was still excluded (10x). In case the exclusion criterium was invalid, it was checked whether any other criterium applied to the abstract, which was twice the case (2x). If not, then the paper was added to the list of papers to be considered for the full-text selection in phase S3 (2x). These two papers were initially excluded by the author based on criterium 3, *The agent or robot does not aim to support children in achieving learning goals through 1-on-1 interaction*, and criterium 7, *The paper is purely focused on ethical aspects, such as the acceptance of robots in a classroom*. Ideally, the author would have done a re-check on all papers excluded based on these criteria. Unfortunately, time constraints prohibited doing both. Therefore, this was only done for criterium 7. This did not result in the inclusion of extra papers.

Retrieval phase – S2 + Eligibility assessment phase – S3. The next steps were performed in a separate Rayyan.ai project. All abstracts that were included in phase S1 were transported to a new file. The full texts were mostly retrieved using the Google Scholar search engine[5]. The TNO link solver often came in handy to retrieve the PDF file from the publisher's website. If the paper could not be accessed via Google or Google Scholar, Researchgate[6] was used. Here, a request for

5 https://scholar.google.com/
6 https://www.researchgate.net/

the full text could be sent to the authors. This was done for five papers, for which one access was granted.

The full texts of all records were scanned to see whether any of the exclusion or inclusion criteria applied (see Section 2.3). Similar to phase S1, a label was used to indicate which criterium applied. Again, to save time, only one of the criteria was used as a label, even if more criteria applied. The author included 38 papers and excluded 80 papers.

Similar to the screening phase (S1), a partial second check was performed by the same two raters. In total, rater 1 rated 64 full-text papers and rater 2 rated seven full-text papers. Three papers were rated by both of them. In total, five conflicts were detected. One of these conflicts appeared because the author applied criterium 13 to exclude a paper, meaning that this was not the only paper describing a specific study. Since the raters did not go through all papers, it is possible that they did not come across any of these cases. Furthermore, rater 1 excluded two papers that were included by the author. In both cases (1x criterium 5 and 1x criterium 15), the author had made a mistake and agreed with the rater. Both mistakes were corrected. In the two remaining cases, rater 2 had made the decision to include the paper, where the author had decided to exclude it. In one case, the conflict resided around the judgment whether the paper involved learning goals or not (criterium 3). The author asked rater 1 for a second opinion, after which they agreed to stay with the author's decision (i.e., to exclude it). In the other case, the author discussed their decision for excluding the paper with the rater in question (i.e., rater 2). In the end, they agreed to stay with the author's decision to exclude this paper. Altogether, the second check resulted in the additional exclusion of two papers in the final selection.

Backward citation searching (snowballing). Due to time constraints, backward citation searching was performed on a part of the final selection. The papers that seemed most relevant for the current research project were selected. These included studies that were aimed at stimulation of (self-)reflective learning processes and/or studies in which a memory of interactions was used. The corresponding 16 papers were highlighted in the Structured Summary table (see Section 2.5).

The backward search was executed in a semi-systematic manner. The references of the 16 papers were scanned. Only titles that referred to a virtual agent or social robot and learning (e.g. "tutor", "teaching", "education") and at least hinted at children as the target audience (e.g. "school", "pupil") were considered. In addition, the filter criteria that were applied by the search engine (see Section 2.2) were also used for scanning the titles for candidates (e.g., only titles published after 2015 were considered). This resulted in 57 eligible titles. These were consequently checked for earlier occurrence in the literature search process. I.e., it was checked whether they were already excluded in phase S1, S2, S3, and whether they were already included in the final selection. This resulted in the exclusion of 19 and 17 papers, respectively. At last, the full texts of the remaining titles were judged based on all 15 exclusion and inclusion criteria (see Section 2.3). This resulted in the exclusion of 19 papers. In the end, this process resulted in the addition of two papers: the paper by Ramachandran, Huang and Scassellati (2017) and the one by Wijnen et al. (2019).

2.5 Data collection process

The author prepared a sheet to summarize the general information and answers to the questions as stated in Table 1. The author scanned all papers to fill in the fields. If the answers were not found, the answer was labeled with "NA". Missing data, such as the average age of the sample, was calculated manually where possible. The resulting tables can be found in Appendix A and Appendix B. The “General information” table (Appendix A) contains basic information, such as the publication type and the sample size of each study. The “Structured summary” table (Appendix B) contains summarized answers to the questions of this literature review (see Table 1).

2.6 Results

2.6.1 General results

In total, 38 papers were selected through the systematic literature review. Six papers were published in 2016, five in 2017, eight in 2018, four in 2019, ten in 2020, and five in 2021. Exactly half of them were journal articles and half of them were conference papers. Seventeen studies were conducted in Europe, ten in the United States of America, six in Asia, and one in South-America. From four papers it could not be derived where the evaluation took place.

On average, the sample size was 46. The smallest sample size was two (Imbernón, Manjarrés, & De La Paz López, 2016) and the largest sample size was 146 (Nielen et al., 2018). The gender distribution was not always published. The largest absolute difference in gender distribution was 38 males and 14 females (Ramachandran, Huan, Gartland, & Scassellati, 2018). The average reported age of participants was 9.44[7]. Only 13 papers stated both the age range and the average age of participants. One paper did not list any data regarding age (Li, Gobert, & Dickler, 2019). Only one paper listed the cultural background of participants (Ramachandran, Litoiu, & Scassellati, 2016). This study took place in the USA where the majority had a Caucasian background (60 percent in one condition and 78.6 percent in the other).

The average number of conditions was 2.16. The majority of the studies (21) consisted of multiple conditions. The highest number of conditions was four. Eleven papers included a condition without the presence of a virtual agent or robot. The majority of the papers (27) employed a between-subjects design. Four papers used a within-subjects design. The longest study period was 10 months. The average number of sessions with the virtual agent or robot was 1.87, but this is without three papers that did not report the number of sessions. The majority of the papers consisted of one session only and the highest number of sessions was an average of 8.3 (Michaelis & Mutlu, 2018)[8]. The average reported session length was 23.08 minutes[9]. The shortest reported length was five minutes and the longest reported length was 60 to 80 minutes.

Twenty-four papers included both a pre-test and a post-test. Two papers only conducted a pre-test and five papers only conducted a post-test. Seven papers used neither of these tests but instead focused on the performance during interaction with the virtual agent or robot. Twenty papers used only quantitative methods, two used only qualitative methods, and 16 papers used a mix of both.

---

[7] In case no average was provided, the average of the age range was taken.
[8] In this study, the number of sessions was not predefined, but relied on the number of times participants would initiate a session with the robot.
[9] This did not include papers that stated the maximum length of the session. In case a range was provided (e.g. 20-30 minutes), the average was taken in the calculation.

2.6.2 Structured summary

The subquestions for the current systematic literature review (see Table 1) were answered by categorizing the answers for each paper. See Appendix B for the results. The categorizations are elaborated upon below.

Table 1. Subquestions of systematic literature review (repetition).

Q1	What type of embodiment is used for the agent or robot?
Q2	What role does the agent or robot fulfill? (i.e., tutor, tutee, or peer[2])?
Q3	In what learning domain is the agent or robot implemented?
Q4	What is the intended effect of the agent or robot?
Q5	What is the specific learning intervention that is facilitated by the agent or robot?

Q1. Concerning embodiment, robots were the subject of the majority of the papers (33x). The NAO robot was encountered most often with 20 occurrences (see Figure 3). The other robot types consisted of the Jibo robot, Pepper robot, Zeno robot, Sota robot, Maki robot, Tega robot, SocibotMini robot, and Julia robot. The robots varied strongly in appearance. For instance, the Julia robot was rather large, bulky, and robot-like, whereas the SocibotMini robot was relatively small, humanlike, and consisted of only a head and torso (see Figure 3). They also varied in terms of moveability. For instance, both the Julia and the SocibotMini robot were relatively static, whereas the NAO robot had a wide variety of possible movements and postures.

Figure 3. Different types of robots. (1) Own picture. (2) From Robotics Today (n.d.). (3) From Song, Barakova, Markopoulos, and Ham (2021).

In the remaining five cases, a virtual agent was used. The embodiments consisted of a bunny (Nielen et al., 2018), a robot-like creature (Ocaña, Morales-Urrita, Pérez-Marín, & Pizarro, 2020), a dog (Ong & Soriano, 2016), and planet earth (Tellols, Lopez-Sanchez, Rodríguez, Almajano, & Puig, 2020). One paper did not present or describe the appearance of the agent (Li, Gobert, & Dickler, 2019).

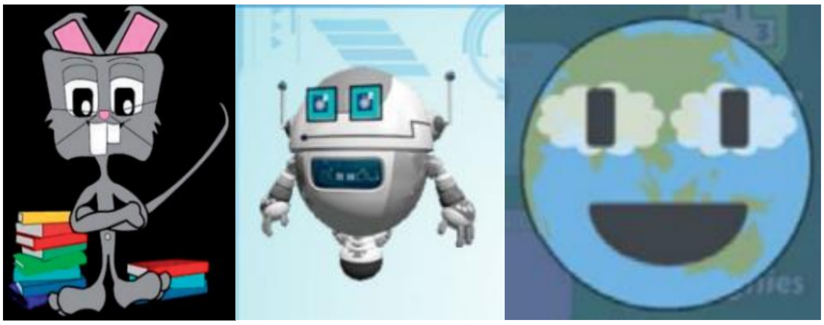


Figure 4. Appearances of virtual agents. (1) From Nielen et al. (2018). (2) From Ocaña, Morales-Urrutia, Pérez-Marín, and Pizarro (2020). (3) From Tellols, Lopez-Sanchez, Rodríguez, Almajano, and Puig (2020).

Q2. For the role of the agents and robots, the most occurring role[2] was that of tutor (21x), followed by peer (7x), followed by tutee (5x). Tutor behaviors included providing instructions (e.g. Konijn & Hoorn, 2020), feedback (e.g. Sandygulova et al., 2020), praising (Guneysu & Anrich, 2017), setting and monitoring goals (Michaelis & Mutlu, 2018), and making suggestions (e.g. to take a break; Ramachandran, Huang, & Scassellati, 2017). Behaviors of peer agents and robots included initiating a conversation (Ong & Soriano, 2016), expressing encouragement (Davison et al., 2021), and playing a game together (Castellano, De Carolis, D'Errico, Macchiarulo, & Rossano, 2020). Tutee agents and robots were taught how to write in a new script (3x) and how to solve ratio problems (2x).

In addition, there were four studies in which the role of tutor and peer were compared. For instance, in the context of a programming game, a study consisted of one condition in which the robot provided instructions and feedback, and one where the robot had to finish the task together with the child. Furthermore, the tone and vocabulary of the robot was based on the robot's role (i.e., tutor or peer; Tazhigaliyeva, Diyas, Brakk, Aimambetov, & Sandygulova, 2016). The four studies provided contradictory results: one showed significantly better results for the tutor role (Tulli et al., 2020), one for the peer role (Tazhigaliyeva et al., 2016), and the results of the other two were inconclusive (Song, Zhang, Barakova, Ham, & Markopoulos, 2020; Song, Barakova, Markopoulos, & Ham, 2021).

Q3. The majority of the learning domains were school-related (30x). These included mathematics (Konijn & Hoorn, 2020), biology (Li, Gobert, & Dickler, 2019), and script learning (Sandygulova et al., 2020). Three papers were in the persuasive domain: two regarded environmental sustainability (Castellano et al., 2020; Tellols, Lopez-Sanchez, Rodríguez, Almajano, & Puig, 2020) and one regarded healthy lifestyle (Ros et al., 2016). The latter aimed to intrinsically motivate learners to adopt a healthy and active lifestyle through a creative dance activity. The dance movements were related to the most important nutritional and physical activity-related guidelines to achieve a healthy lifestyle. The remaining contexts (5x) were categorized as 'Other'. An example is a study in which a Socibotmini robot employed two different roles to measure to what extent they influenced motivation to play a musical instrument (Song, Barakova, Markopoulos, & Ham, 2021).

Q4. The agents and robots had various intended effects. The majority of the studies (21x) aimed to achieve 'Leaning gains'. For instance, two different behavior styles for an autonomously tutoring robot were tested to see if they improved standard school test results for multiplication tables (Konijn & Hoorn, 2020). Other studies were targeted at, inter alia, enhancing creativity (Ali, Devasia, Park, & Breazeal, 2021), attitude change (Castellano et al., 2020), and extended concentration (Maeda, Even, & Kanda, 2019). In some cases the studies had multiple intended effects (17x). One study was a probe study and therefore had no explicit intended effect (Michaelis & Mutlu, 2017).

Papers that were aimed at stimulation of (self-)reflective learning processes were considered particularly relevant for the current research project. In total, four of the titles appeared to have this aim. For instance, a robot tutor supported learners in developing self-regulated learning skills by using an open learner model (Jones, Bull, & Castellano, 2018; Jones & Castellano, 2018). These studies formed an inspiration for the dialogue design of the learning session (see Section 3.3).

Q5. The agents and robots employed a variety of learning interventions. For instance, they used praise (Davison et al., 2021), corrective feedback (Sandygulova et al., 2020), personalized behavior (Baxter, Ashurst, Read, Kennedy, & Beplaeme, 2017), and adaptive scaffolding (Li, Gobert, & Dickler, 2019). Extra attention was paid to papers that used a memory of interactions between the child and the agent or robot, because this can facilitate reflection. This applied to 14 cases. For instance, the learner's data was stored so that the robot could use personalize information and refer back to previous game outcomes (Ahmad, Gao, Alnajjar, Shahid, & Mubin, 2021).

3 DESIGN

The goal of the design phase was to provide an answer to RQ2, as posed in Chapter 1: *How can a social robot use dialogue to encourage children (7-12 years) to reflect on learning goals and learning progress when learning individually about healthy nutrition?* The deliverable of the design phase was a learning session about healthy nutrition, accompanied by a dialogue script for the robot. In designing such a deliverable, a human-centered design approach was taken. The ISO 9241-210 standard for "Ergonomics of human-centered system interaction" was used for this purpose. This standard required explicit understanding of users, tasks and environments (ISO, 2019). This was established by, inter alia, creating personas and scenarios (see Section 3.1). Please note that the personas and scenarios were not yet specialized on healthy nutrition. Instead, they were aimed at health literacy in general, similar to the ePartners4all project.

The selection of a robot over a virtual agent was based on several reasons. The first reason is that physical robots have shown benefits compared to virtual ones, such as more attention (Looije, van der Zalm, Neerincx, & Beun, 2012) and stronger friendship (Sinoo et al., 2018). These findings were both in the context of research with children. Moreover, the systematic literature review (see Section 2) indicated that physical robots are more popular in this field of research than virtual agents. There were also practical reasons. A NAO robot was readily available at TNO. Besides, a number of Dutch primary schools own a NAO robot to teach children how to program them. They use a platform called Robotsindeklas[10]. It is offered by Interactive Robotics[11], a company that provides solutions for social robots in education and healthcare. They also participate in the ePartners4all project, so they could provide a license and support to use their platform for this research project.

3.1 Personas

Three personas were created to understand the target group's needs and goals (Cooper, 1999). They were meant to illustrate the potential struggles that children might have with healthy lifestyles. All three personas can be found in full size in Appendix C.

The first persona is named Michelle (see Figure 5). She is a dreamy eight-year old girl who likes being creative. During creative work, she often loses track of how much she has eaten. She dislikes physical exercise and prefers to sit down with a snack. She struggles with her performance in P.E. classes.

[10] https://portal.robotsindeklas.nl/
[11] https://www.interactive-robotics.com/

Figure 5. Persona 1: Michelle.

The second persona is Bruno (see Figure 6). He is a social 10-year old boy who likes to exercise. Having divorced parents who rarely communicate with each other, there is little consistency in his lifestyle. He is an only child and he sometimes feels alone.

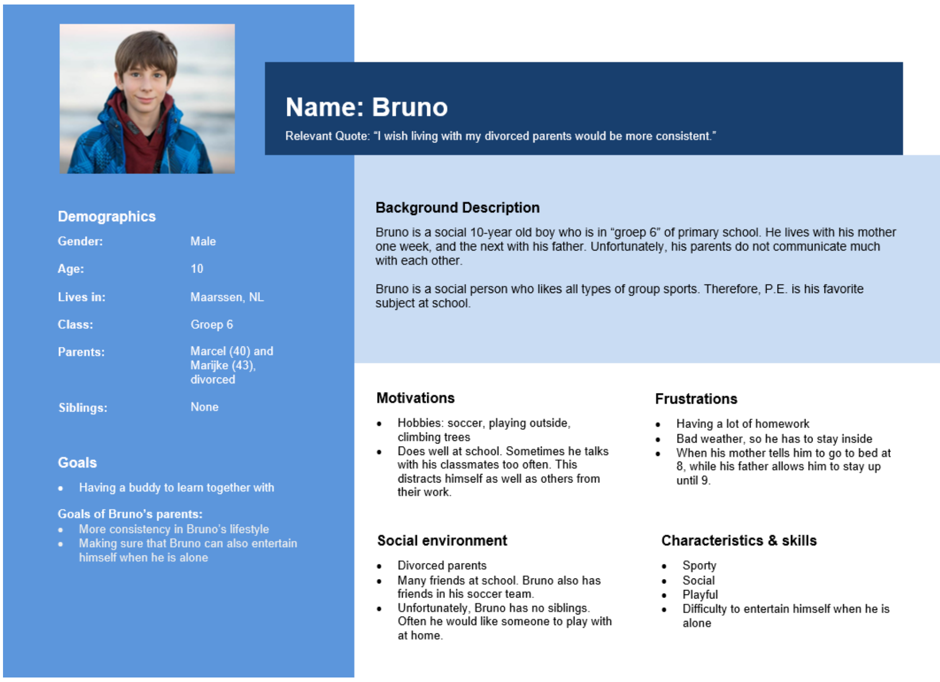


Figure 6. Persona 2: Bruno.

The third persona is called Jamiro. He is an introverted boy who is into gaming. He gets so caught up in gaming that he loses track of time. He has trouble sleeping at night when he games in the evening. During the day, he often feels tired and finds himself daydreaming about videogames.

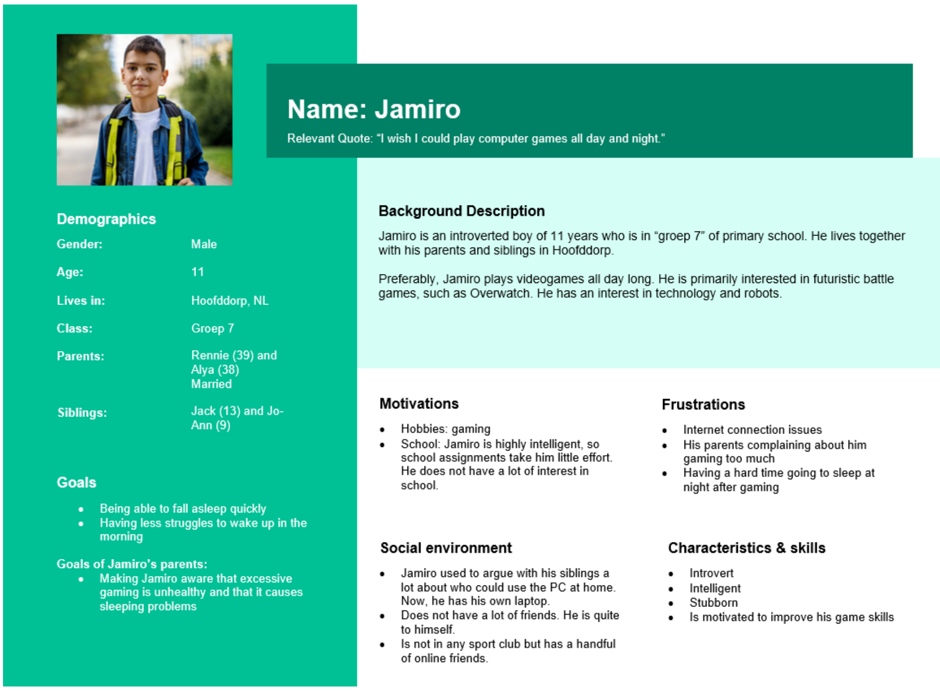


Figure 7. Persona 3: Jamiro.

3.2 Scenarios

Scenarios aim to provide an explicit concrete vision of how some human activity could be supported by technology (Nardi, 1992). Here, scenarios were divided into three types: background scenarios, problem scenarios, and design scenarios. The background scenarios illustrated the current problems and obstacles in the personas' daily lives. The problem scenarios illustrated what issues could occur when the personas learnt about healthy lifestyle without support (of a robot). The design scenarios sketched how the social robot could help the personas to overcome these obstacles and achieve their personal goals (see Section 3.1).

3.2.1 Michelle

Background scenario Michelle. Michelle is a dreamy eight-year old girl who is currently in third grade of an elementary school in Emmen. She has been adopted by her two fathers, together with her biological sister Mathilda. Michelle is a creative girl who loves arts and crafts, drawing and reading. She dislikes physical exercise and prefers to sit down with a book and a snack. As both her dads work full-time, she and her sister have plenty of freedom to do what they like. Even though one primarily works at home, he is often in a call when Michelle and her sister arrive home from school. Usually, they pour themselves a glass of coke and grab a handful of candy. Their parents have bought them crackers as well, but those do not taste as nice. Michelle grabs a small bag of potato chips before she goes to her room. She continues reading the fantasy book that has been on her shelf for a while. She opens the bag of potato chips and finishes it without much awareness. Around the time she reaches chapter 5, she notices that she is thirsty, so she goes downstairs to get a glass of coke again. She bumps into her sister who asks her to come play outside with their friends in the neighborhood. Michelle rejects this invitation, as she dislikes being the slowest player in a game of tag.

Problem scenario Michelle. The next day at school, Michelle's teacher introduces a digital game about healthy and unhealthy habits. Michelle gets a bit anxious, what if the game forbids her to eat candy? She is afraid to be judged by her class mates, as everyone saw her eating two cookies during break time. She knows she is not the slimmest of the class. But her sister eats just as unhealthily as she does without being overweight. It is not fair! People underestimate how difficult it is to stop eating candy. The game starts with information that she was long aware of. It feels like the game does not take into account the difficulty of changing entire patterns. As soon as she finishes the game, she grabs a bar of chocolate from her bag to soothe her feelings.

Design scenario Michelle. The next day at school, Michelle's teacher introduces a digital game about healthy and unhealthy habits. She tells that they will be supported by a robot, called NAO. Michelle gets a bit anxious, she does not like this topic. What if the game forbids her to eat candy? She is afraid to be judged by her class mates. Maybe the robot will judge her as well! But, it looks quite friendly, she thinks. Once she starts the conversation with the robot, she starts to feel at ease. The robot asks about her interests and her personal goals regarding healthy living. She tells him that she wants to become better at P.E.. When the learning goals are presented, she feels like they are doable, because she already has some background knowledge. During the game, her anxiety has almost vanished. Afterwards, the robot asks if she is now more confident in knowing how to reach her goal. She answers with "yes".

3.2.2 Bruno

Background scenario Bruno. Bruno is a social 10-year old boy who is fourth grade of elementary school. He lives with his mother one week, and the next with his father. Unfortunately, his parents do not communicate much with each other. In the weeks that he lives with his mother, he walks home from school, because she lives so nearby. After having drunk a glass of water and having eaten a banana, he sees some of his friends playing outside and notifies his mom that he is going outside. At dinner time, his mother calls him to come home. They have spaghetti as main course and kiwi as dessert. Bruno is disappointed, at his father's he have had ice cream or chocolate mousse every evening. He should have known better, as his mother tries to keep the amount of candy at her house at a minimum. After dinner, Bruno has a short look at his homework and then browses through YouTube. At eight pm, his mother tells him to shut down his laptop and go to sleep. Bruno grunts, his father always allows him to stay up until ten. However, he also knows that he wakes up tired if he goes to bed at that time.

Problem scenario Bruno. The next day at school, his teacher introduces a digital game about healthy and unhealthy habits. Bruno is curious about this game. When he starts playing, he does not immediately understand the relevance of the game. He likes the game, but is not very focused on the information. After a couple of minutes, he gets distracted and starts a conversation with his friend sitting opposite of him. His teacher notices this and asks them whether they have finished the game already. They have not, so the teacher tells them to continue with this. Bruno has lost his interest and motivation and does not manage to fully finish the game. He wish he could have played this game together with a friend. That would have made it less boring.

Design scenario Bruno. The next day at school, his teacher introduces a digital game about healthy and unhealthy habits He announces that a robot will help them. Bruno gets excited, he is curious what the robot will talk about with him. He greets the robot enthusiastically. The robot asks Bruno about his interests and he responds elaborately. The robot seems to listen to him well!

Bruno plays the game attentively, since he expects the robot to talk about it as soon as he finishes. And indeed, the robots asks him whether he feels confident in reaching the learning goals. Bruno feels that this is indeed the case and answers with "yes". Bruno looks forward to talking with the robot again.

3.2.3 Jamiro

Background scenario Jamiro: Jamiro is an introverted boy of 11 years who is in fifth grade of elementary school. He lives together with his parents and siblings in Hoofddorp. Preferably, Jamiro plays videogames all day long. He is primarily interested in futuristic battle games, such as Overwatch. Normally, as soon as he arrives home after school, he goes to his room to start playing. He stops at five pm and goes downstairs, because his parents do not allow him to play after that. After dinner, he watches some television, and after some time he goes upstairs to his room again. He secretly starts his laptop again and plays Overwatch. As usual, he loses track of time and is startled when his mother knocks on the door to say goodbye. He quickly puts his pajamas on and lays down in his bed. His mother kisses him goodbye and leaves. However, he cannot sleep yet. After one hour in bed, he still has not fallen asleep. He also notices a bit of neck ache. Finally he starts dreaming. The next morning, his alarm goes off at the usual time. Jamiro has a lot of trouble waking up and still notices some ache in his neck.

Problem scenario Jamiro. That afternoon at school the teacher introduces a digital game about healthy and unhealthy habits. To Jamiro, this seems like a lame game. No cool characters, weapons, or animations.. boring! All his classmates enthusiastically start to play the game, but Jamiro tries to go through it as quickly as possible just to get it over with. He does not pay a lot of attention to the content. Why would he need to live more healthily? He eats enough vegetables, his parents make sure of that. At some point in the game something about playing videogames is mentioned. He already knew that gaming is not necessarily healthy. It makes sense that his parents disallow him to play after five pm for a reason. Jamiro finally finishes the game, without much awareness of what he has learnt and why.

Design scenario Jamiro. That afternoon at school the teacher introduces a digital game about healthy and unhealthy habits. She tells that they will be supported by a robot. Jamiro starts getting excited, an actual robot? The robot starts by asking Jamiro what his motivation is to play the game. He cannot think of anything immediately. The learning goals of the game are displayed. There is something about healthy sleeping patterns, which draws Jamiro's attention. Jamiro starts playing the game and the robot points at his progress towards this goal. At the end of the game, Jamiro notices that he has reached the learning goal. He feels happy about this, especially since the robot compliments him about this.

3.3 Dialogue design

The overall structure of the dialogue can be found in Figure 8 below. It consisted of four components: getting acquainted, introducing the topic, the learning tasks, and rounding up. It was based on the personas and scenarios, design patterns (Ligthart et al., 2019), and reflection principles and strategies. All will be explained in the following sections. The full dialogue script can be found in Appendix D.

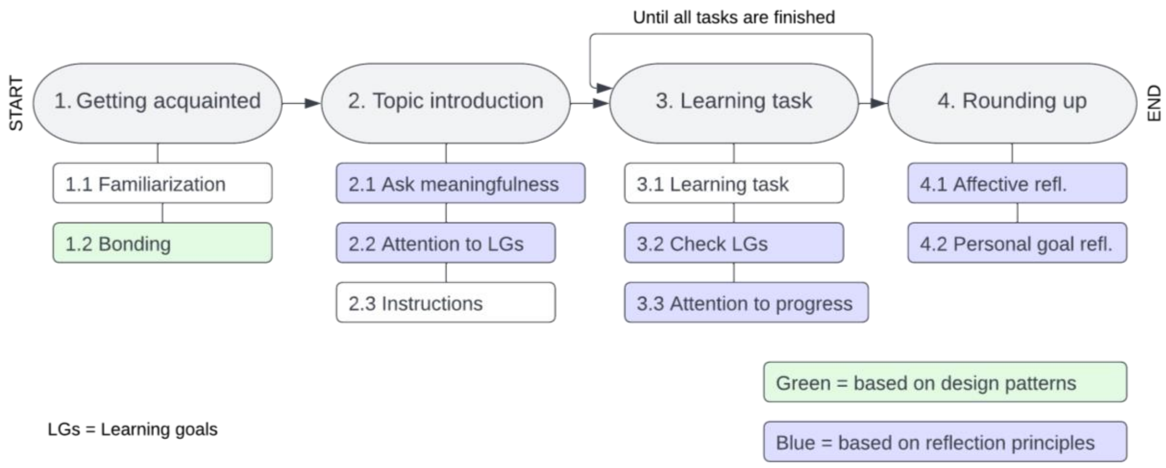


Figure 8. Dialogue structure.

3.3.1 Design patterns and general requirements

It was important to ensure that bonding between child and robot was facilitated. This would help to align expectations, put the learner at ease, practice talking with the robot, and counter a novelty effect. Ligthart et al. (2019) designed a collection of interaction design patterns, called structured dyadic interaction design, for this purpose. These design patterns aimed to have the robot and child getting acquainted by reciprocally disclosing personal information to each other. The example in their paper was slightly adapted and used during the “Bonding” phase of the conversation (see element 1.2 in Figure 8). Applying personalization to the conversation required recording the child’s personal answers. This resulted in a semi-structured conversation, meaning that the sentences were predefined but that there was room for personalization based on the learner’s responses (see Figure 9).

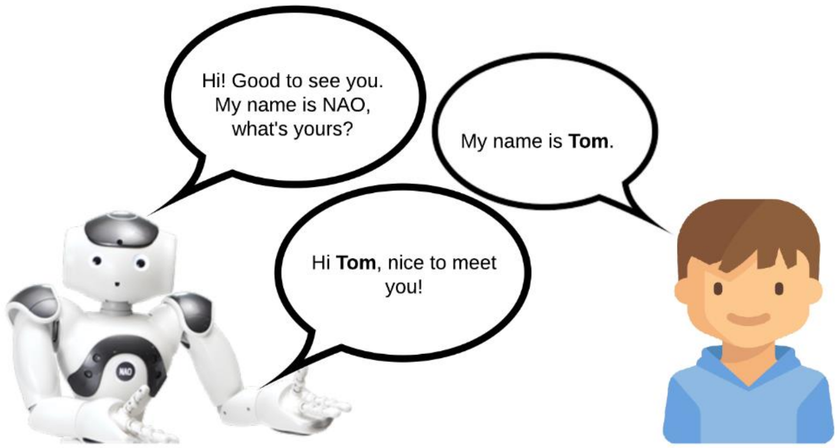


Figure 9. Semi-structured conversation between child and robot.

Furthermore, the insights gathered from the personas and scenarios led to four requirements for the dialogue design. These were incorporated throughout the entire dialogue. See Table 3 below.

Table 3. Requirements from personas and scenarios.

#	Requirement	Based on…
R1	The robot should be encouraging	Michelle & Jamiro
R2	The robot should be understanding	Michelle
R3	The robot should help children to focus on the content	Bruno
R4	The robot should pay attention to child's personal goals	All

R1 was implemented through utterances such as "Very good!" and "Good to hear!" when the child responded positively. R2 was implemented by refuting negative answers with "That's okay, I'll help you" and "Are you sure? I think you've learnt a lot today. You can be proud of yourself!". R3 was implemented through the robots movements (explained below) and R4 will be discussed in Section 3.3.2.

To make the dialogue as natural as possible, movements were incorporated in its design. The robot should speak animatedly, using head movements and small hand gestures. At the start of the session, the robot should wave to greet the learner. At the end of the session, the robot should applaud for the learner's efforts. To direct the attention to the learning content on the screen at the right times (R3), the robot should crouch shortly before a learning task started. The robot should stay crouched until the task was finished. After that it should stand up again.

3.3.2 Reflection principles and strategies

To facilitate reflection on learning goals and learning progress, a list of reflection principles was composed. These principles consisted of elements that could be used in a conversation to encourage reflection. They were each accompanied by a reflection strategy that the robot could use to facilitate this principle. The reflection principles, strategies, and examples can be found in Table 4. It also presents where these reflection strategies can be found in the dialogue structure of Figure 8.

Table 4. Reflection principles and strategies.

#	Reflection principle (RP)	Reflection strategy by robot	Example	# in Figure 8
1	Awareness of the learning goals	Drawing attention on the current learning goal(s)	"Do you see the learning goals on the screen?"	2.2
2	Establishment of meaningfulness of the goal	Asking the child why they think the goal is meaningful	"Why do you think that eating healthily is important?"	2.1, 4.2
3	Reflection on achievement of learning goals	Verifying whether the child thinks they have achieved the learning goal	"Do you think that you now know which beverage is the healthiest?"	3.2
4	Attention to progress in learning process	Drawing attention on progress towards overall learning goal	"Do you see the progress we are making?"	3.3

5	Affective appraisal of achievement of learning goals	Asking the child how they feel about their achievements	"How do you feel about what we learnt today?"	4.1

Reflection principle 1 aimed at directing the learner's attention to the learning goals. Multiple studies have shown that providing the learner with specific and attainable goals aids performance compared to a situation where the learner is provided with no goals (Lunenburg, 2011). Learning goals were an important element in the PAL project, which stands for Personal Assistant for healthy Lifestyle (Neerincx et al., 2019). Together with multiple other parties in The Netherlands and Italy, TNO contributed to the design of a personalized system for child's diabetes self-management training. The system allowed for personal goal-setting, which supported the "autonomy" aspect of the Self-Determination Theory (Ryan & Deci, 2017). Neerincx et al. (2019) argued that personalizing the learning objectives and providing explanations improved the transfer of responsibility to the child. This led to the establishment of Reflection principle 2.

Reflection principle 3 focused on the learner's aim to master a learning topic. This was based on a study by Jones, Bull, and Castellano (2018), in which a social robot supported self-regulated learning in children. They describe mastering a topic as having covered all of the content and being confident in correctly answering the content. For instance, the robot would ask: "Do you think you have mastered this activity?". This aligned with having completed a learning goal.

Reflection principle 4 was based on the PAL project. In their system, they used an achievement dashboard to show, amongst others, the learning goals and learner's progress towards them. In addition, it was one of the robot's functions to provide feedback on the learner's learning progress (Neerincx et al., 2019). They based this on Coninx et al. (2016), who stated that, "to pursue learning and therapeutic goals through child-robot interaction, it is important to ensure the child remains engaged in the relationship and that the child experiences progress in achieving educational goals" (Neerincx et al., 2019, p.13).

Finally, the fifth reflection principle aimed to form an association with an emotion to the learning session. This was initially meant as something to refer back to in future sessions, especially when the emotion was positive. For instance: "last time we learnt about … and you felt happy about it". In the PAL project, emotions were also recorded over time to, inter alia, perform mood-matching between the robot and child (Neerincx et al., 2019). Unfortunately, due to time constraints, it was only possible to incorporate one session in the experimental design. Still, it was hoped that positive affective appraisal would result in a positive effect on children's motivation to eat healthier. For instance, a study by Duncan (1993) indicated that teenagers who experienced positive levels of affect in physical education were more motivated to participate in physical activity outside of school.

3.3.3 Learning content

Finally, to deliver a prototype of the learning session for evaluation, the content of the learning session was selected. These formed the learning tasks mentioned in Figure 8 (element 3.1). The topic of the learning content was narrowed down to healthy nutrition. As mentioned in Chapter 1, nutrition is a key contributor to children's health (Melanson, 2008). The learning material that was looked for fell in the category "edutainment": a combination between education and entertainment (Anikina & Yakimenko, 2015). This included educational videos and digital games, which were deemed suitable for out-of-class independent learning (Swatevacharkul, 2021). In the end, one video and one game[12] were selected. Four learning goals were composed based on its content: three for the video and one for the game (see Figure 12). They were pilot tested before beforehand (see Appendix L).

The educational video used as part of the experiment is a YouTube video called "Duurzaam & Gezond eten met de Schijf van Vijf | Voedingscentrum"[13] (see Figure 10). The video was created by "het Voedingscentrum"[14], a well-known Dutch organization that provides guidelines for healthy nutrition. The video explains how the "Schijf van Vijf"[15] can be used to make healthy and sustainable nutrition choices. The video did not appear to have a specific target audience in mind, but the level of explanation and visualization was deemed appropriate for children. The duration of the original video is two minutes and 38 seconds. As the current experiment focuses on healthy nutrition and not on sustainable nutrition, a snippet of 39 seconds was cut from the video, resulting in a video of one minute and 59 seconds.

Figure 10. Screenshot of the video[13].

The educational game used for the experiment is the game "Gezond of Ongezond" (Healthy or Unhealthy) on the website Kijkopgezond.nl[16] (see Figure 11). In this game, the user has to categorize food products as healthy or unhealthy. In turn, they receive feedback on their decision. The website was created approximately six years ago by the Dutch company Netrex[17]. This is a social organization that aims to aid the inclusion of vulnerable groups in society. An interview with its owner, Jacques de Wit, was conducted to gather relevant background information about the game. The game was created to educate groups with intellectual disability about healthy living in an informal manner. This group of people is generally likely to be overweight, to have a slow metabolism, eat too much, and exercise too little. The game did not target a specific age group. The website was created in collaboration with Steffie, LFB, Zorg in Beweging, Ieder(in) and experts from sheltered employment. "Het Voedingscentrum" also played a role in the information provision about healthy nutrition. The game was funded by several Dutch foundations. As of

[12] It can be argued that this particular "game" does not fulfill the criteria for being an actual game. It could be considered as a gamified test, since a score was recorded and feedback was provided in the form of sounds. For convenience purposes, the term "game" is used throughout this Thesis.
[13] https://www.youtube.com/watch?v=2pKJJPBLn-0
[14] https://www.voedingscentrum.nl/nl.aspx
[15] The "Schijf van Vijf" is a visualization that recommends which food categories should be consumed every day, and in what proportions.
[16] https://www.kijkopgezond.nl/spel_gezond/
[17] https://www.netrex.nl/

today, the website receives a few hundred unique visitors per day. Netrex is currently working on a new version of the game, which is expected to be finished before the end of 2022.

Figure 11. Screenshot of the game[16].

A PowerPoint page served as a visual aid during the learning session (see Figure 12). It contained the links to the educational video and game. This page also presented the learning goals that belonged to them. If the learner clicked on one of the white boxes, a checkmark would appear. The checkmark would disappear if the box was clicked again.

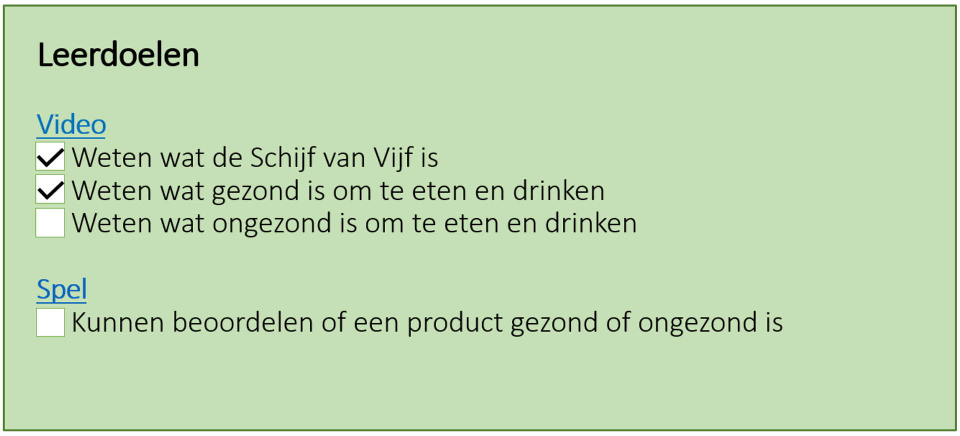


Figure 12. Learning goals screen.

4 METHOD

The current chapter presents the design of an experiment that evaluated the dialogue session presented in Chapter 3 and Appendix D. This served RQ3: *Does the dialogue design for reflection on learning goals and learning progress lead to better reproduction of the learning goals, higher learning outcomes, and more motivation to show healthy behavior?* This research question is accompanied by three hypotheses, based on work in a similar project (PAL; Neerincx et al., 2019). When the social robot employs reflection strategies, the child….

H1: can reproduce the learning goals better;
H2: reaches higher learner outcomes;
H3: is more motivated to show the healthy behavior that they learnt about…
… compared to a social robot that does not use reflection strategies.

4.1 Participants

4.1.1 Sampling procedure

Convenience sampling was used to gather the sample for the current experiment. Two primary schools were approached that were connected with Interactive Robotics[11], a party that also participated in the ePartners4all project. These schools had purchased a license to use the Robotsindeklas[10] platform and owned a NAO robot. It was therefore expected that the pupils in these schools had worked with the robot before. This would rule out a novelty effect in the results, thereby making the results more valid. I.e., the initial enthusiasm for the robot would be reduced and the children would have more knowledge of the robot's capabilities.

Both primary schools were introduced to the researcher by an email from Interactive Robotics, after which the researcher sent them an email with general information about the experiment. They were asked to let the researcher know if there were any classes between grade 2 and 5 that were willing to participate. In the end, two classes of the school Harbour Bilingual in Rotterdam[18] (grade 3 and 4) expressed an interest and the researcher contacted the teachers to make the arrangements. Due to limited resources including time no attempt was made to approach more schools for participation. The sample size was therefore limited to the participants in these two classes.

In principle, all pupils in these classes were eligible to participate in the study. The only inclusion criterium was that participants were able to speak and read Dutch. In case a child would not be able to read and converse (in Dutch) without specific adaptations, for instance because they were hearing or visually impaired, they would still have been allowed to participate in the experiment. However, their data would have been excluded from the analysis.

4.1.2 Participant characteristics

The sample was limited to the class members that were willing to participate and had informed consent from their parents. They also had to be present in class on one of the experiment days. In total, 38 pupils participated, of which 28 in the original conditions (see Section 4.4.2). See Table 5 for the exact numbers. To safeguard the privacy of this vulnerable group (i.e., underage

[18] https://www.harbourbilingual.nl/

children), no characteristics other than gender were recorded, with the exception of optional photographs for illustrative purposes (see Section 4.2.1). Based on the grade that they were in, the age of all participants was estimated between eight and 10 years old. See Table 6 for an overview of the distribution per condition. The division of participants over the conditions will be discussed in Section 4.3.

Table 5. Sample numbers.

	Grade 3 / "Groep 5"	Grade 4 / "Groep 6"	Total
Number of pupils	23	26	49
Received consent	21	18	39
Consent & present	20	18	38
Individual session[19]	14	14	28
Group session[19]	6	4	10

Table 6. Gender distribution.

Grade 3 / "Groep 5"		Grade 4 / "Groep 6"	
Experimental	Control	Experimental	Control
3 male, 4 female	3 male, 4 female	2 male, 5 female	3 male, 4 female
6 male, 8 female		5 male, 9 female	
11 male, 17 female			

4.2 Materials

See Section 3.3.3 for the video, game, and learning goal screen.

4.2.1 Information forms and consent form

Information forms were created for both the pupils and their parents. See Appendix F and Appendix G, respectively. The former was based on a standard template provided by the CCMO, "Centrale Commissie Mensgebonden Onderzoek", for children under 12 years[20]. The latter was based on a template for participant information in intervention research provided by TNO. The age group of the participants (under 12 years) required the parents to give consent for the participation of their child. See Appendix H for the consent form. TNO's template for parental consent in intervention research was used for this end. Two extra options were added, namely those for taking a recognizable picture or unrecognizable picture of the participant. Neither of these options were required to be selected in order for a child to be able to participate in the study, but limited which children could be photographed and how.

4.2.2 Robot control

The dialogue designs for the experimental and control condition can be found in Appendix D and Appendix E, respectively. More information on the design of the dialogues can be found in Chapter 3. The dialogue was fully controlled by the researcher in a semi Wizard-of-Oz-based setting. I.e., the researcher would control the robot, while being (visibly) present in the same room. This included listening to the participant's answers, personalizing the script if necessary, and sending the speech commands to the robot at the right time. An HTML-page with JavaScript embedded

[19] See Section 4.4.2.
[20] https://www.ccmo.nl/onderzoekers/publicaties/formulieren/2019/10/14/e1-e2-model-bespreekblad-voor-kinderen-tot-12-jaar

was used for this purpose (see Figure 13 for a screenshot for the control condition and see Appendix I for the accompanying code).

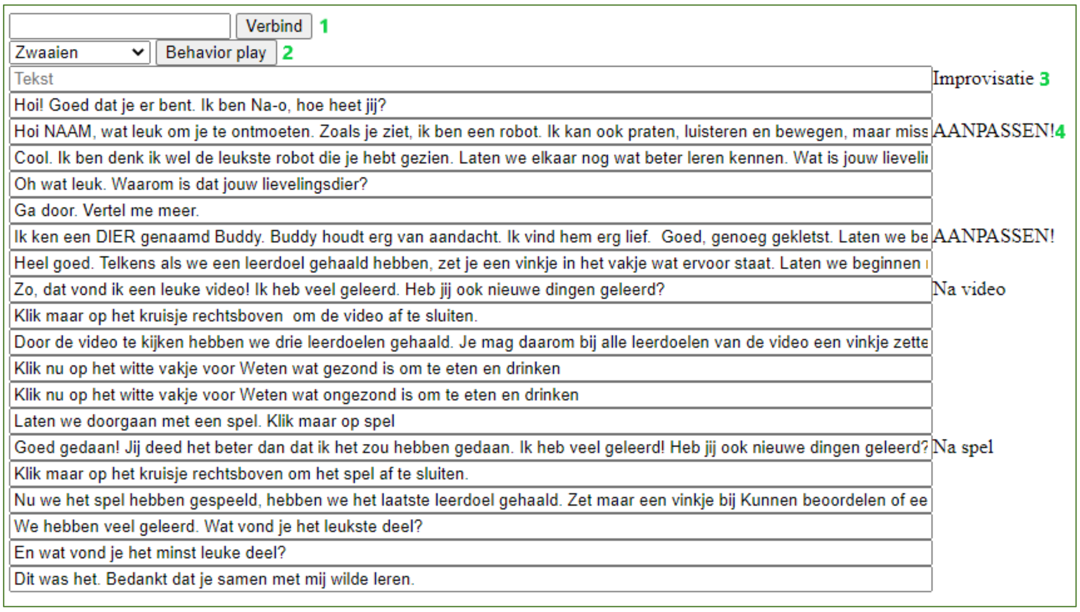


Figure 13. WoZ HTML page.

The first element (1) on the page was a text entry box with a button that said “connect”. Here the robot could be connected to the Robotsindeklas platform by entering its realm. A realm can be seen as the robot’s closed environment, comparable to the robot’s address. It was disclosed on the Robotsindeklas website once the NAO robot was turned on and connected to internet. The “wamp.2.msgpack” protocol from the Autobahn library[21] was used to establish the connection. The second element (2) consisted of a drop-down menu with selected robot behaviors and a “behavior play” button. Robot behaviors included waving, crouching, standing up, and clapping. Once a behavior was selected and the button was pressed, the playBehavior function was used to make the robot perform this action. The third web page element (3) was a blank text box that was meant for the rare occasions that required improvised dialogue, for instance if the participant asked the robot a question before or after the session. The text was sent as a speech command to the robot by pressing enter. In the session.call() functions that were used for speech, the parameter “rie.dialogue.say_animated” would make the robot speak more vividly by using small gestures while talking. The parameter “rie.dialogue.config.native_voice” was meant to give the robot a softer, more human voice compared to its default voice. The remaining part of the page consisted of text boxes in which the dialogue sentences were already entered (see Figure 13). Again, the speech commands could be sent to the robot by clicking the text box and hitting enter. The text in the boxes could be edited before sending (4). This way, some words could be personalized based on the responses by the participant.

[21] https://github.com/crossbario/autobahn-js

4.2.3 Questionnaires

The two post-session questionnaires were created using Survalyzer[22]. This is the standard survey tool used by TNO. The tool did not allow for many customizations to the layout of the questionnaires. It was decided to keep the appearance clean and simple to not distract the participants. The full questionnaires can be found Appendix J and Appendix K.

Questionnaire 1 was filled in directly after the session with the robot. It consisted of four pages (see Appendix J). On the first page, the researcher could enter the participant number. This was used as an identifier for the pseudonymized data. On the second page, participants were asked to fill in as many learning goals as they could remember (see Figure 14, for the learning goals see Section 3.3.3 and Figure 12). This was meant to assess short-term recall of the learning goals.


Schrijf hier het eerste leerdoel op
Als het niet meer weet, laat het vakje dan leeg
Weten wat ....

Schrijf hier het tweede leerdoel op
Als het niet meer weet, laat het vakje dan leeg
Weten wat ....

Schrijf hier het derde leerdoel op
Als het niet meer weet, laat het vakje dan leeg
Weten wat ....

Schrijf hier het vierde leerdoel op
Als het niet meer weet, laat het vakje dan leeg
Kunnen beoordelen of ....

Figure 14. Questionnaire 1 page 2: learning goals.

The third page contained test questions about the content of the video and game (see Figure 15). A mix of multiple choice, single choice, and open questions was used. The goal was to have questions where the answer was not simply correct or incorrect, but to facilitate a range of possible scores. This would allow for more varied test scores that would measure the performance of both conditions more precisely. The questions were composed by the researcher and tested during a pilot with four children in the same age group as the participants (see Appendix L).

[22] https://www.survalyzer.com/


1. Waar zijn voedingsstoffen goed voor?
Er zijn **2** goede antwoorden

- [ ] Dunner worden
- [ ] Fit blijven
- [ ] Groeien
- [ ] Minder gaatjes in je tanden
- [ ] Slimmer worden

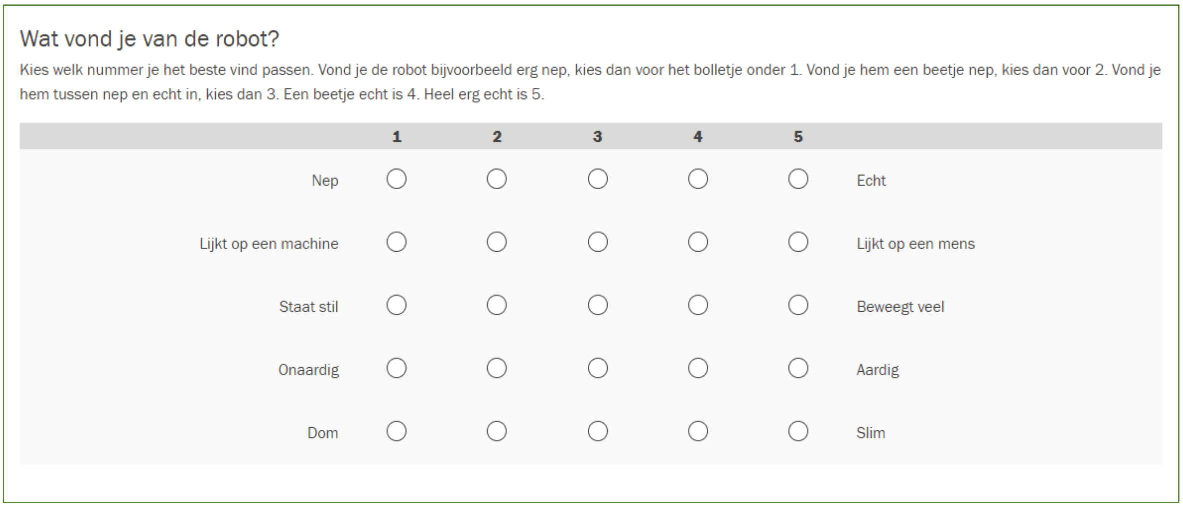
2. Wat staat er **wel** in de Schijf van Vijf?
Er zijn **3** goede antwoorden

- [ ] Aardappels
- [ ] Cola
- [ ] Koekjes
- [ ] Thee
- [ ] Vlees

Figure 15. Questionnaire 1 page 3: test questions.

The fourth page of this questionnaire contained an adapted version of the Godspeed questionnaire, which is widely used to evaluate the perception of social interactions with robots (Bartneck, Croft, Kulić, & Zoghbi, 2009; see Figure 16). Together with domain expert Mark Neerincx, who has multiple publications in the field of child-robot interaction, we selected the most essential and suitable items. This resulted in a selection of five items out of the 24 original items: fake-natural, stagnant-lively, machinelike-humanlike, unintelligent-intelligent, and unkind-kind. These belonged to four out of the five different subquestionnaires of the original Godspeed questionnaire. These were Anthropomorphism (2x), Animacy, Likeability, and Perceived Intelligence, respectively. The fifth subquestionnaire, Perceived Safety, was not deemed relevant in this context, so no items were selected from it. There were three reasons for the strong reduction in items: 1) the limited number items that were expected to be easy to understand for children, 2) the short attention span of children in this age group, and 3) the limited amount of time of the experiment. This adapted version of the Godspeed questionnaire was also used during an experiment by PhD student Anouk Neerincx[23], which took place prior to the current experiment. This allowed for pilot-testing the understandability of the questionnaire for children. The researcher observed that the vast majority of participants in this experiment were able to fill in the questionnaire without help.

[23] To date, the results of this study have not been published yet.



Wat vond je van de robot?

Kies welk nummer je het beste vind passen. Vond je de robot bijvoorbeeld erg nep, kies dan voor het bolletje onder 1. Vond je hem een beetje nep, kies dan voor 2. Vond je hem tussen nep en echt in, kies dan 3. Een beetje echt is 4. Heel erg echt is 5.

	1	2	3	4	5	
Nep	◯	◯	◯	◯	◯	Echt
Lijkt op een machine	◯	◯	◯	◯	◯	Lijkt op een mens
Staat stil	◯	◯	◯	◯	◯	Beweegt veel
Onaardig	◯	◯	◯	◯	◯	Aardig
Dom	◯	◯	◯	◯	◯	Slim

Figure 16. Questionnaire 1 page 4: robot assessment.

Questionnaire 2 was filled in approximately one week after the session with the robot. It was the same as questionnaire 1, but without the robot assessment part on page 4 (see Appendix K). Additionally, the first page did not state "the researcher will fill in a code here", but instead stated "the teacher will fill in a code here".

4.2.4 Food choice

After the session, participants could choose a small treat as a thank-you for participating. They could choose between a healthy box of raisins or an unhealthy box of chocolate Smarties (see Figure 17).

Figure 17. Food choice[24].

4.2.5 Apparatus

The robot used in the experiment was a NAO V6 on which the Interactive Robotics software was installed. The laptop used by the participants was a HP ZBook 15 G6. It was accompanied by a

[24] Source left image: https://www.jumbo.com/producten/jumbo-snoep-rozijntjes-14-stuks442213ZK
Source right image: https://www.jumbo.com/producten/smarties-mini-melk-chocolade-uitdeelzak-465254ZK

Maxxter gaming mouse with model number LB-KM-MOGAU01-BBL. The laptop used by the researcher to control the robot was a Dell Latitude 7280. Audio was recorded using the Samsung Voicerecorder app on a Samsung Galaxy A40 smartphone. Pictures were taken by a Huawei MAR-LX1B and a Xiaomi Mi 10 smartphone. The software OpenShot Video Editor v2.6.1. was used to shorten the educational video by "het Voedingscentrum".

4.2.6 Overview of outcome variables

See Table 7 below for an overview of all primary, secondary and explorative outcome variables.

Table 7: Measurement scheme.

Outcome variable	Type	Timing	Material
Retention of the learning goals	Primary – H1	After the session	Questionnaire 1+2 page 2: participant writes down as many learning goals as they remember
Retention of the educational content	Primary – H2	After the session	Questionnaire 1+2 page 3: six questions about the content of the videos and game
Motivation to eat healthily	Primary – H3	After the session	Choice for a healthy or unhealthy treat (+ the motivation behind this choice)
Assessment of the robot (agency)	Secondary	After the session	Questionnaire 1 page 4: adapted version of Godspeed Questionnaire
Responses to the robot	Explorative	During the session	Audio recordings + notes

4.3 Study design

An experiment with a between-subjects design was conducted to measure the effect of the reflection strategies, as delivered by the robot. In the experimental condition, five reflection strategies were incorporated in the dialogue between robot and child (see Appendix D). These were based on prior work in this domain, see Section 3.3.2. In the control condition, these strategies were left out. This resulted in a slightly shorter session, with less questions asked by the robot and less responses expected from the participant (see Appendix E).

As the sample consisted of two primary school classes, it was assumed that all participants were approximately the same age and generally possessed the same background knowledge. It could have therefore sufficed to do a random division over the two conditions. However, it was deemed useful to strive for an equal balance between the groups in terms of gender and school performance. To avoid recording too much personal data of the pupils (gender, age, and attainment level), the teachers of the classes were asked to make two equal groups of children. They were not aware that these groups participated in different conditions. On the days that the experiment took place, the teacher was asked to consecutively send a participant assigned to the first group (experimental condition) and a participant assigned to the second group (control condition) to the experiment room. This was done to outbalance time effects.

4.4 Procedure

4.4.1 Experiment set-up

The experiment took place in Harbour Bilingual's school building in Rotterdam[18]. A large office on the top floor was used for the experiments. This was a relatively quiet area, where only small groups of children or teachers would walk by every now and then.

The researcher and participant were seated at separate desks (see Figure 18 and Figure 19). The NAO robot and a laptop with mouse (see Section 4.2.5) were placed at the head of the table, turned towards the participant's seat. The research assistant would sit somewhere else at this table.

As the researcher was positioned behind the participant, she could see their actions on the laptop screen. Due to resource constraints, it was not possible to build a full WoZ set-up, where the researcher would be invisible. This would require the use of a camera, microphone, a screening wall and, ideally, technical staff.

Figure 18. Experiment set-up, back view.

Figure 19. Experiment set-up, front view.

4.4.2 Protocol

In the week before the experiment took place, the researcher paid a visit to the participating classes. She guided a short group conversation about science and robots and provided practical information about the experiment. Pupils could ask questions about the research and the robot. Additionally, she handed out the information sheets and consent forms. She told the pupils that they could hand in the completed consent forms to their teacher on one of the following days, if they and their parents agreed for them to participate.

The next week, the researcher and a research assistant[25] conducted experiments on five consecutive days. In total, 32 sessions took place. All sessions followed a script that was composed by the researcher. See Appendix M for the document with all practical details for the research assistants. The procedure was thoroughly tested prior to the experiment. See Appendix L for more information on the three pilot studies that took place.

Before each session, the research assistant would pick up the participant from their classroom. On the way to the experiment room, the research assistant would start a casual conversation with the participant, asking questions such as "are you looking forward to the session?" and "what do you expect from the robot?". Meanwhile, the researcher would prepare the laptop screen and WoZ connection and start the audio recording.

Once the participant had arrived, they were greeted by the researcher and asked to sit down behind the laptop. The researcher would tell the participant that once they waved to the robot, that it would "wake up". This was meant to give the participants the sense that the robot behaved autonomously. Once the participant started waving, the researcher would start conducting the dialogue using the WoZ page. During the dialogue, the research assistant would check if consent was provided to take a recognizable or unrecognizable picture. If this was the case, they would make one or two pictures randomly during the session. Participants were helped when needed (e.g., controlling the laptop). Once the dialogue was finished, the researcher would compliment the participant and say that they now would answer a number of questions. The researcher would walk up to the participant laptop to open the questionnaire and enter the corresponding participant number. The audio recording was stopped and saved. The participant was given as much time as needed to fill in the questionnaire. If they had questions, the researcher and research assistant would help them out. However, they did not provide any answers to the test questions.

Once the participant had finished, the researcher would tell them that she wanted to give the whole class a treat and that they could already pick what they wanted. They were not handed the treat right away to ensure that participants were not influenced by each other's choice. She would pull out a bag with Smarties boxes and a bag with raisins boxes. The participant's decision was noted down and the researcher thanked them for their participation. The research assistant would walk the participant back to their classroom. On the way, they would ask them to not tell other classmates about the session (because "this could spoil the experience for the subsequent participants"). Once the participant arrived back in the classroom, the research assistant would guide the next participant to the experiment room. The entire session took approximately 20 minutes. See Figure 20 for an impression of the sessions.

---

[25] A research assistant was present for practical and ethical reasons. Practical reasons included saving time by distributing the tasks. Ethical reasons included safeguarding the integrity of how the researcher treated the underage participants.

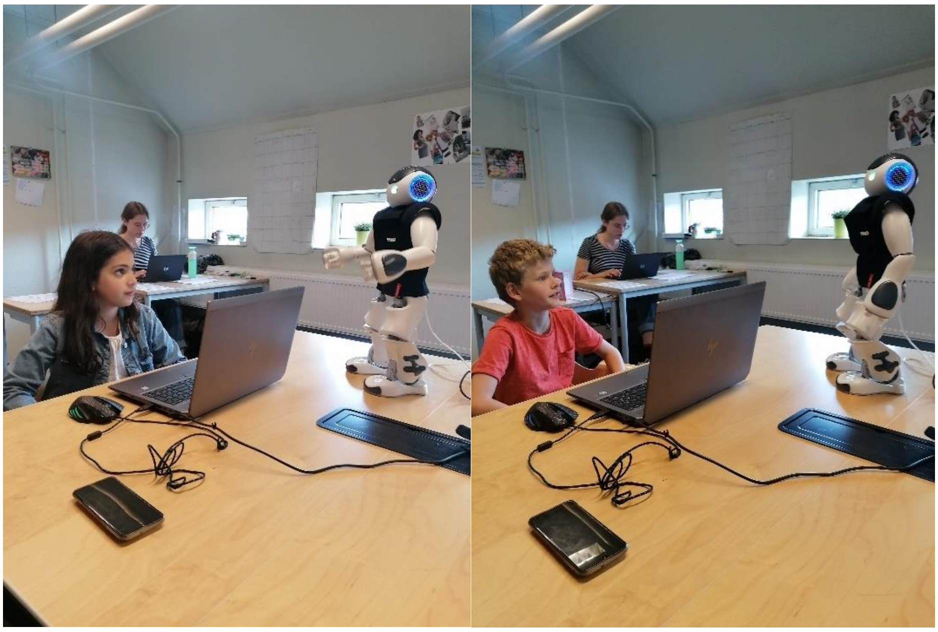


Figure 20. Impression of the experiment.

Due to unforeseen changes in the classes' schedules, there was too little time to conduct a "normal" session for each of the pupils with parental consent. Since the researcher did not want to disappoint them, she facilitated group sessions for those that were left at the end (six in grade 3 and four in grade 4). It should be noted that in the case of grade 3, these participants were generally the lower performing pupils of the class, according to their teacher. During their sessions, the participants conversed with the robot as a group. The script of the control version was used, because that one was shorter. Participants played the game as a team. They did not fill in questionnaire 1, because there was only one laptop available. However, they did fill in questionnaire 2.

Approximately one week later, the participants completed questionnaire 2 in class. The goal of this questionnaire was to measure long-term retention of the learning goals and long-term learning outcomes (see Appendix K). The part about robot assessment was therefore left out. The researcher was not present and therefore the teachers were responsible for the execution. They were instructed to hand every participant a Chromebook and direct participants to the corresponding link. They were sent a list with the participants' names and accompanying participant numbers. This way they could fill in the correct participant number on the first page (see Figure 21). The teachers were told that participants were not allowed to discuss their answers with each other. In addition, they were not allowed to help them out with the answers. Unfortunately, two participants in the experimental condition were not present when questionnaire 2 was made in class.


De juf vult hier een code voor je in

Volgende

Figure 21. First page of questionnaire 2.

4.4.3 Safety and handling of adverse events

No adverse events or negative side effects were expected beforehand. The child's state was closely monitored by the researcher and the research assistant. The child was never left alone. Teachers were welcome to visit the room at all times. In case one of the participants would have shown any signs of distress, the session would have been ended instantly and the teacher would have been informed. In case a complaint or concern had been expressed about the way the experiment is conducted, it would have been recorded and both the internship supervisor and the research manager would have been informed.

4.4.4 Data preprocessing and analysis

First, the questionnaire data was downloaded in Excel format. All irrelevant columns in the dataset were deleted, such as the start date. The Godspeed questionnaire values of participant R8 were removed (see Section 5.1). Columns were renamed using abbreviations and numbering (e.g. "LG1" for the column with the first learning goal). An extra column was added named "Condition". This was meant to facilitate the comparative analyses. All data underwent a sanity check. Missing values were not replaced. Obvious outliers were removed if present. This was done once, when participant R7 did not provide answers to any question in questionnaire 2.

Consecutively, the test scores were calculated. The number of recalled learning goals was counted for each participant. The order of the learning goals (e.g. naming the learning goal on the top of the page first) was disregarded. No points were deducted in case a child had mentioned a learning goal that was incorrect. The scores of the short test were calculated by accumulating the number of correct answers. For some questions, multiple points could be scored. See Table 8 for all possible scores per question. No points were deducted for incorrect answers. The preprocessed data was then imported into SPSS Statistics v28.

Table 8. Possible scores.

Question	Possible scores
Q1	0, 1, 2
Q2	0, 1, 2, 3
Q3	0, 1
Q4	0, 1
Q5	0, 1, 2
Q6	0, 1
Total	0, 1 ,2, 3, 4, 5, 6, 7, 8, 9, 10

A descriptive statistics analysis was performed on all quantitative outcomes. This included calculating the mean, standard deviations, and summarizing the data in figures. The Godspeed questionnaire series scores are usually analyzed by calculating the mean scores for each of its subquestionnaires (Anthropomorphism, Animacy, Likeability, Perceived intelligence, and Perceived safety; Bartneck, Croft, & Kulić, 2008). Since a strongly reduced version of this questionnaire series was used (see Section 4.2.3) the mean score was calculated for each item.

The data file was split based on the two conditions. This allowed for the comparison of the groups. A Shapiro-Wilk test for normality was conducted to determine whether an independent-samples t-test or a Mann-Whitney U-test should be used to compare the conditions. One of these two tests was performed on the number of learning goals remembered correctly and the scores on the content questions. An alpha level of .05 was used for all statistical tests in this study.

For the qualitative data, no specific data analysis technique was used. The frequencies of the treat selected (healthy or unhealthy) and children's motivation behind this choice was compared between conditions without further statistical analysis. Children's responses to the robot were analyzed for remarkable and/or recurring patterns without looking for anything specific.

5 RESULTS AND DISCUSSION

This chapter presents the results of RQ3: *does the dialogue design for reflection on learning goals and learning progress lead to better reproduction of the learning goals, higher learning outcomes, and more motivation to show healthy behavior?* See Chapter 4 for the hypotheses and the design of the experiment.

5.1 Abnormalities

In some sessions the experiment did not fully go as planned.

Delay in speech command. During three sessions, it occurred that a speech command was delayed. The robot would stay silent after the researcher had sent the command, after which she would send the command again. This resulted in the two speech commands (same text) being executed shortly followed by one another. This sounded somewhat like an echo. In these cases the researcher told the child that this is something that could happen with robots. The researcher judged whether the participant had still understood what the robot had said and based on that either continued with the dialogue or repeated the speech command.

Connection lost. During three sessions, the connection between the robot and the Interactive Robotics platform was lost. The robot would indicate this by saying: "*I have lost my connection*". A few seconds of silence would pass and after that the robot would say: "*I am trying to connect to the server*", shortly followed by the idle behavior that occurs when the robot has connected to the platform. This idle behavior would consist of saying "*Hi, what are we going to do today?*" and movements such as leaning on one leg, looking up, and sneezing. The researcher would execute the procedure to connect the robot to the WoZ web page again. This would take a few seconds, after which the dialogue proceeded. Similar to the "echo" case, the researcher would make a comment such as "*it is still a robot, after all*".

Environmental noise. The experiment room was generally quiet. Sometimes a teacher or a pupil would walk by, but this did not seem to distract participants. From the second day onwards the windows of the room were open, because it was rather warm and humid in the attic. This would sometimes cause noise from cars, classes playing outside, and thunder storms. Three sessions could not take place at the described experiment room. The researcher received a late notice that it would be used by others on Friday morning. These morning sessions therefore took place in a small corridor near the classroom (see Figure 22). The location was rather noisy, since children were moving in and out of the classroom every now and then. The researcher and assistant signed to those children to be quiet. There was no interference with other children. Classmates who walked in the corridor and saw the NAO were the ones that had already participated in the experiment or that did not have permission to participate (to the extent to which the researcher could observe this).

Figure 22. Session in corridor.

Not fully completed. It was important to take into account the class's schedule when conducting the sessions. One session was about to exceed the start of the school's break. In this case, the researcher decided to skip the robot assessment part of the questionnaire to save time. After the participant had finished the questions about the content, she would let the research assistant accompany them back to the class, while she entered random values in this last part of the questionnaire. This data is left out in the analysis.

5.2 General observations

Conversation flow. In general, the conversations between the robot and participants went smoothly. There was only one participant who tried to change the conversation flow and ask the robot questions, such as "*How old are you?".* In this case the researcher would have the robot repeat the question that was originally posed. Eventually, the participant seemed to accept the conversation flow and stopped asking these questions.

Understanding. Overall, participants seemed to understand what the robot was saying and feel comfortable responding. However, they did not always understand when an answer was expected. It appeared that they did not always identify the questions as questions. This mainly occurred with the question that was posed after the video and game: "I have learnt a lot. Have you also learnt new things?" Participants who stayed silent were encouraged by the researcher to respond to the robot. In general, participants responded with "yes". There were two cases where a participant indicated that they did not learn new things from the video. Instead of answering "yes" or "no", one participant started to list the things that they had learnt during the video or game. In this case, the researcher could deduce from the participant's answer that they had reached the learning goal. However, an autonomous robot might not be able to respond adequately in such a situation (see Section 6.3). Participants sometimes struggled to provide a response when the robot said "go on, tell me more" during the introduction phase. Sometimes they indicated to the researcher or assistant that they did not know what (more) to say. The researcher or assistant would ask them to tell the robot more about their favorite animal (as originally meant). In other cases participants started telling things that were not related to their favorite animal. The researcher did not deviate from the consequent dialogue when this occurred.

Mastery of learning goals. In the experimental condition, the robot would ask for each learning goal if the child thought that they had mastered it. Two participants indicated that they did not feel confident in having mastered the last learning goal (*being able to assess whether a food* *product is healthy or unhealthy*). They also scored below average during the game. The other responses to the robot were all “yes”, which could either be because they felt that this was the case, or because they felt that was the desired answer.

Game. Participants did not always seem to know the food products presented in the game. Two participants asked the researcher or assistant to explain some of the items[26]. When one of them made an incorrect choice, they commented “well, that is because I did not know what it was!”. This participant ended the game with one correct answer only and indicated that this was due to this fact. Other than that, participants performed well in the game and scored 5 out of 6 points on average.

Group condition. The most important difference between the unplanned group condition (see Section 4.4.2) and the “regular” individual conditions, was that the groups expressed more negative sentiment during the conversation. I.e., at the end of the session, when the robot asked them what part of the session they liked the least, they were quite vocal about this. In contrast, in the individual conditions, the majority answered that there was nothing they did not like. This seems to be in line with studies on the differences between in robots interacting with groups and with individuals. These indicate that people feel more comfortable engaging socially with robots from the safety of a group as opposed to being alone (Sebo, Stoll, Scasselati, & Jung, 2020).

5.3 Descriptive statistics

The descriptive statistics for all participants are presented in Table 9 below. The results per condition are discussed in the following sections. Please note that “short-term” refers to the results acquired through questionnaire 1, which was filled in directly after the learning session (see Section 4.4.2). “Long-term” refers to data acquired through questionnaire 2, which was filled in approximately one week after the experiment.

Table 9. General descriptive statistics.

Variable	N	Minimum	Maximum	Mode	Mean	SD
Short-term recall of learning goals	28	0	4	3	2.39	.994
Long-term recall of learning goals incl. group condition	33	0	4	0&2	1.7	1.357
Long-term recall of learning goals excl. group condition	24	0	4	2	2.04	1.301
Short-term learning outcomes	28	6	10	8&9	8.21	1.166
Long-term learning outcomes incl. group condition	33	4	10	7	7.88	1.317
Long-term learning outcomes excl. group condition	24	4	10	9	8.13	1.361

[26] The items included sausage rolls (Dutch: saucijzenbroodjes) and marmite.

5.4 H1: Recall of learning goals

5.4.1 Short-term

The figures below illustrate the short-term recall of learning goals in both conditions. Figure 23 shows the recall scores of the experimental condition and Figure 24 shows the recall scores of the control condition.

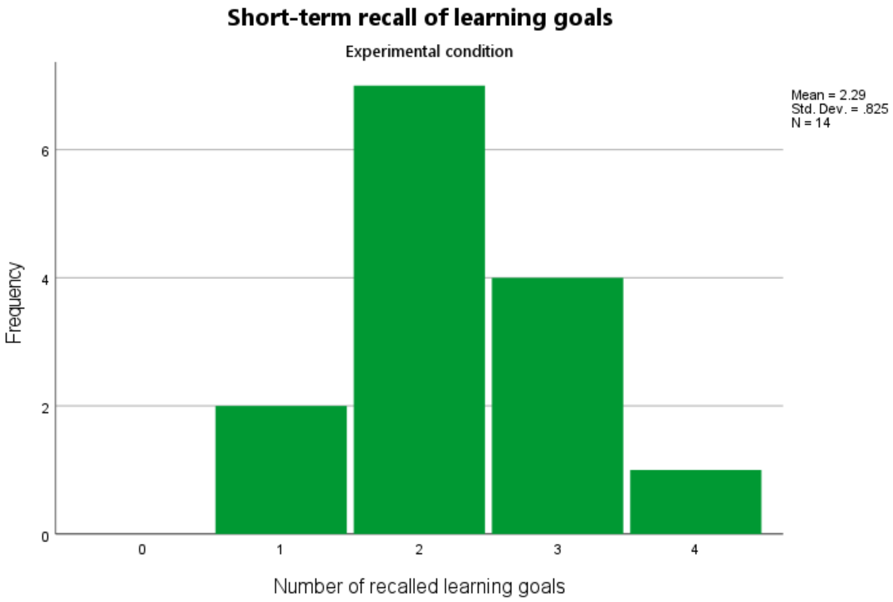


Figure 23. Short-term recall of learning goals – experimental condition.

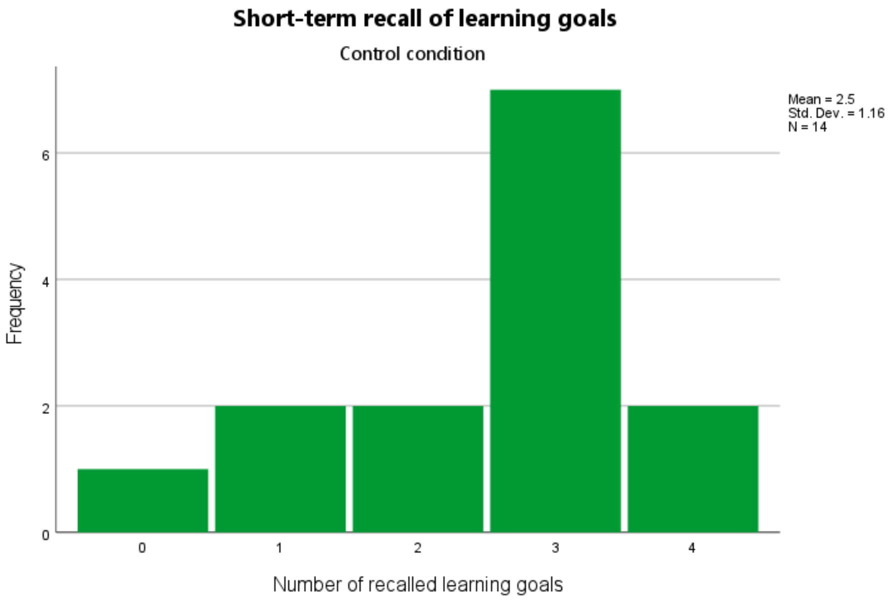


Figure 24. Short-term recall of learning goals – control condition.

To select the correct statistical test to compare these groups, it was determined whether the data was normally distributed. A Shapiro-Wilk test showed no significant departure from normality for recall of learning goals in the experimental condition, $W(.878)$, $p = .055$, opposite to the control condition, $W(.865)$, $p = .036$. Therefore, the analysis proceeded with the non-parametric Mann-

Whitney U-test. This revealed no significant difference (U = 79, $N_{experimental}$ = 14 , $N_{control}$ = 14, p = .358) between the experimental condition (*mean rank* = 13.14) and the control condition (*mean rank* = 15.86). See Appendix N for the full SPSS output tables.

5.4.2 Long-term

The figures above illustrate the long-term recall of learning goals of both conditions. Figure 25 shows the recall scores of the experimental condition and Figure 26 shows the recall scores of the control condition.

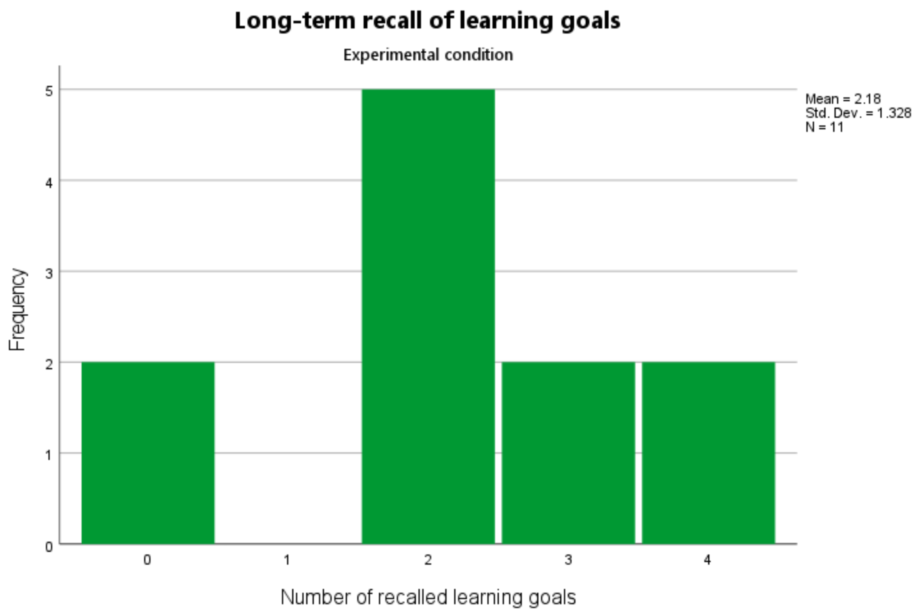


Figure 25. Long-term recall of learning goals – experimental condition.

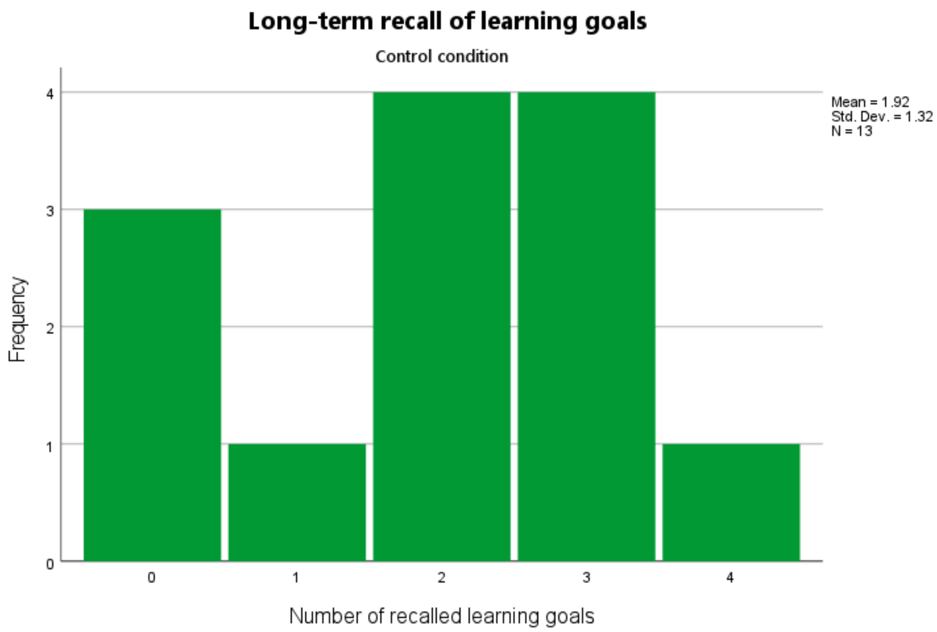


Figure 26. Long-term recall of learning goals – control condition.

A Shapiro-Wilk test showed no significant departure from normality for both conditions, $W(.878)$, $p = .099$ for the experimental condition, and $W(.890)$, $p = .098$ for the control condition. Therefore, the analysis proceeded with the parametric independent-samples t-test. This showed no significant difference in the scores between the experimental condition (M = 2.18, SD = .1.328) and the control condition (M = 1.92, SD = 1.320), $t(22) = -.477$, $p = .834$. See Appendix N for the full SPSS output tables.

As mentioned in Section 4.4.2, there was an additional, unplanned condition, namely the group condition. It consisted of one group of four children and three pairs that participated in the learning session together. The conversation structure was the same as the control condition. Participants did not fill in questionnaire 1 directly afterwards (short-term outcomes), but they did participate in questionnaire 2 (long-term outcomes). See Figure 27 for its scores in recall of the learning goals. See Figure 28 for a comparison with the “individual” conditions (i.e., the experimental and the control condition).

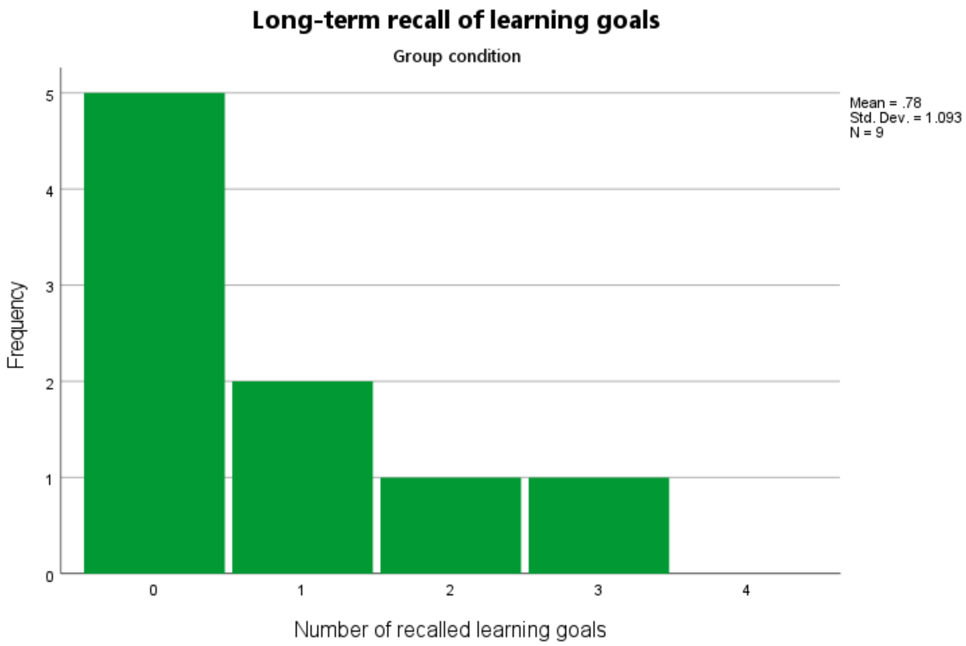


Figure 27. Long-term recall of learning goals – group condition.

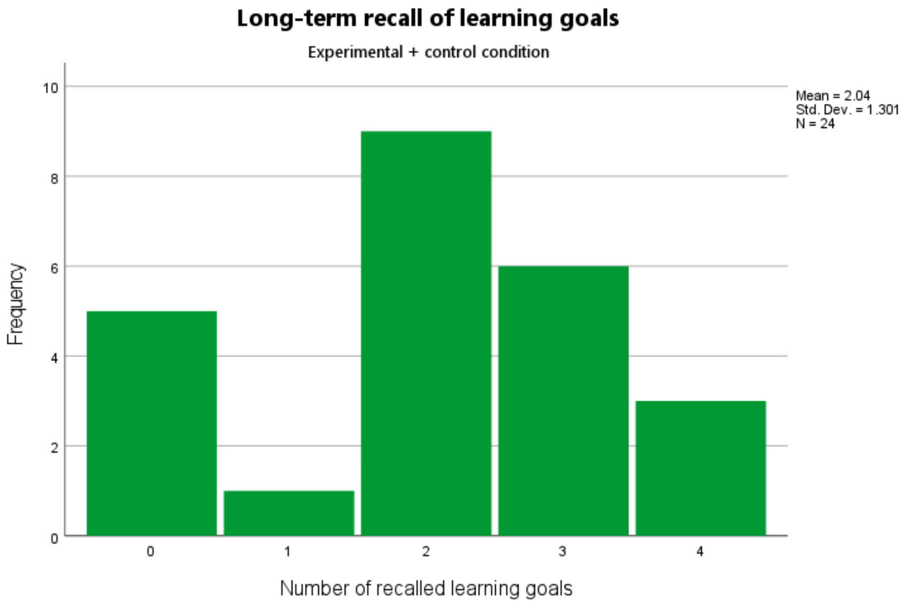


Figure 28. Long-term recall of learning goals – individual conditions.

As these figures indicate a difference between these two groups, it was investigated whether they differed significantly. A Shapiro-Wilk test showed a significant departure from normality for recall of learning goals in both the group condition, $W(.767)$, p = .009, and the individual conditions, $W(.878)$, p = .008. Therefore, the analysis proceeded with a Mann-Whitney U-test. This revealed a significant difference ($U = 52$, $N_{group} = 9$, $N_{individual} = 24$, $p = .019$) between the group condition (*mean rank* = 10.78) and the individual conditions (*mean rank* = 19.33). See Appendix N for the full SPSS output tables.

5.4.3 Discussion

The results above did not show a significant difference between the experimental and control condition regarding the recall of learning goals. Therefore, this provided no indication of the effect of the reflection strategies used by the robot on remembering learning goals. This could have various reasons. For instance, maybe the posed questions did not suffice to actually encourage reflection. Maybe participants automatically answered any question by the robot with "yes" without actually considering if they feel like having mastered the learning goal. Another reason could be that the visual encoding of the learning goals overpowered any verbal encoding that was the result of reflection. Since the learning goals were displayed on the laptop screen throughout most of the session (see Figure 12), this could have been a strong influence.

However, when the short-term and long-term results were compared with each other, we see that the scores of the control condition have decreased more strongly over time (-0.58) than the experimental condition (-0.11). It should be noted however that three participants in the experimental condition and one participant in the control condition were missing in the long-term results. When their scores were removed from the short-term scores, the averages change to $M_{experimental}$ = 2.45 and $M_{control}$ = 2.46. This results in a larger decrease for the experimental condition (-0.27 instead of -0.11) and slightly smaller decrease for the control condition (-0.54 instead of -0.58). Still, the scores of the control condition decreased more over time than the scores of the experimental condition[27]. However, it should be noted that the control condition had better short-term results, so for them there was more content to forget over time.

In addition, a significant difference was found between the group condition and the two individual conditions. This could indicate several things. The difference could have been caused by the fact that the individual conditions filled in the learning goals twice. Participants in the group condition only did this one week later. By writing down the learning goals directly after the learning session, there could have been a stronger encoding effect.

A second possible explanation is that the group conditions consisted of participants with lower class performance scores. In fact, the teacher of one of the classes explained that she had put the best performing pupils at the top of her list and the lower performing ones at the bottom. As a result, the latter participated at the end, when the group conditions took place.

A third possible explanation is a reduced focus on the robot and content due to the presence of other children in the group. In another child-robot study, children listened to two robots playing out interactive narratives either individually or in groups of three. Learning and recall scores for children in the groups were shown to be worse than those of the individual children (Leite et al., 2015; Leite et al., 2017). Sebo, Stoll, Scassellati, and Jung (2020) posed that “a likely explanation for this result is that the children directed less attention to the robots when they were in groups of three because they were also attending to one another, and thus did not retain as much of the information the robots were trying to convey” (p. 176:3).

5.5 H2: Learning outcomes

5.5.1 Short-term

The figures below illustrate the short-term learning outcomes. Figure 29 shows the test scores of the experimental condition and Figure 30 shows the test scores of the control condition.

[27] It was considered to statistically compare the decrease in score per participant over the two conditions. However, since this was such a small decrease and this was not part of the original data analysis plan, no further statistical analysis was performed. Nevertheless, such comparisons are recommended for future work, see Section 6.4).

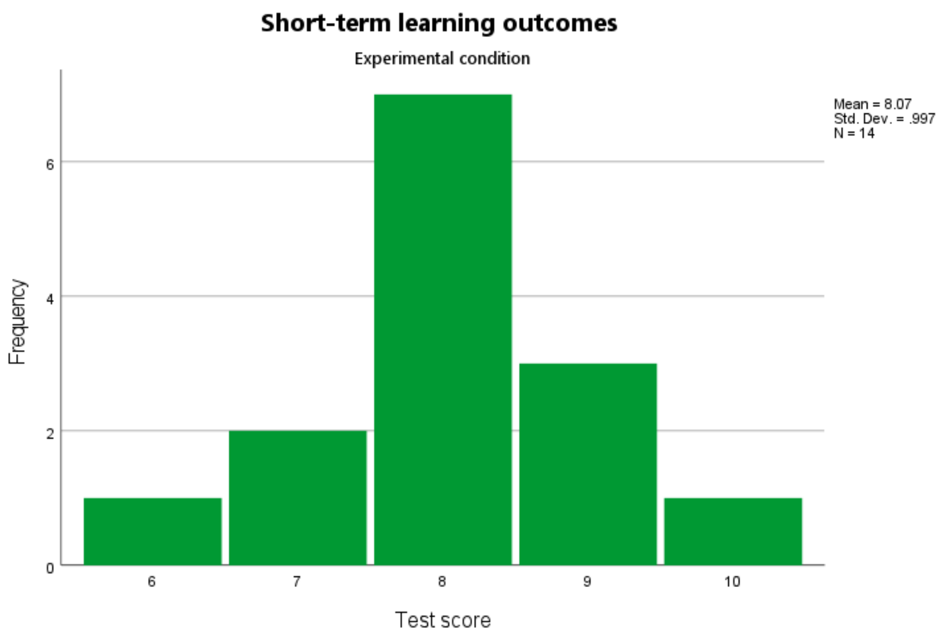


Figure 29. Short-term learning outcomes – experimental condition.

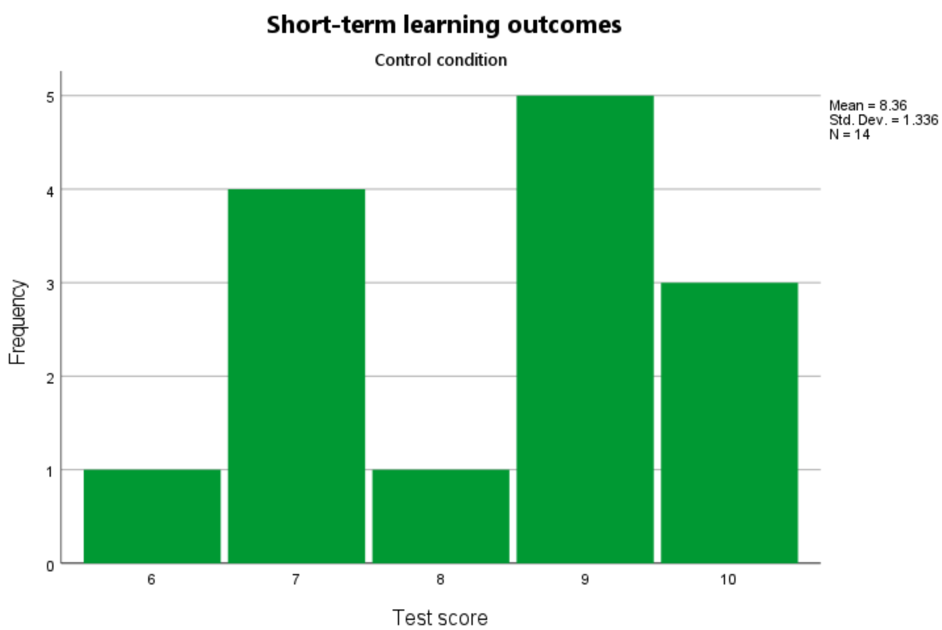


Figure 30. Short-term learning outcomes – control condition.

Again, it was determined whether the data were normally distributed. A Shapiro-Wilk test showed no significant departure from normality for learning outcomes in both conditions, $W_{experimental}$(.913), p = .174, and $W_{control}$(.878), p = .054. Therefore, the analysis proceeded with an independent-samples t-test. This showed no significant difference in the scores between the experimental condition (M = 8.07, SD = .997) and the control condition (M = 8.36, SD = .357), t(26) = .641, p = .055. See Appendix N for the full SPSS output tables.

5.5.2 Long-term

The figures below illustrate the long-term learning outcomes. Figure 31 shows the test scores of the experimental condition and Figure 32 shows the test scores of the control condition.

Figure 31. Long-term learning outcomes – experimental condition.

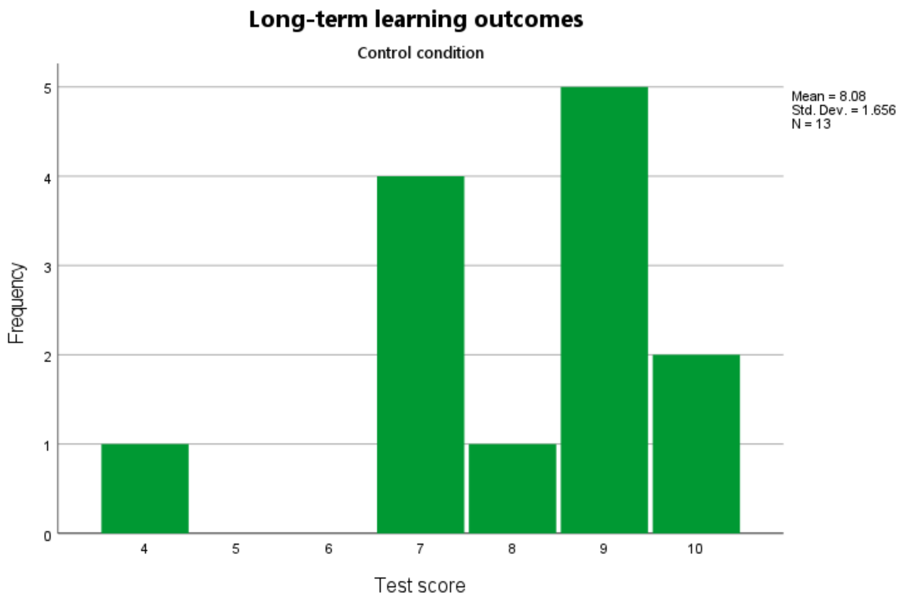


Figure 32. Long-term learning outcomes – control condition.

A Shapiro-Wilk test showed no significant departure from normality for recall of learning goals in the experimental condition, $W(.896)$, p = .165, opposite to the control condition, $W(.857)$, p = .035. Therefore, the analysis proceeded with a Mann-Whitney U-test. This revealed no significant difference (U = 68.5 $N_{\text{experimental}}$ = 11 , N_{control} = 13, p = .857) between the experimental condition (*mean rank* = 12.23) and the control condition (*mean rank* = 12.73). See Appendix N for the full SPSS output tables.

Similar to the recall of the learning goals, there is also data from the group condition. See Figure 33 for its scores in long-term recall of the learning goals. See Figure 34 for a comparison with the "individual" conditions (i.e., the experimental and the control condition).

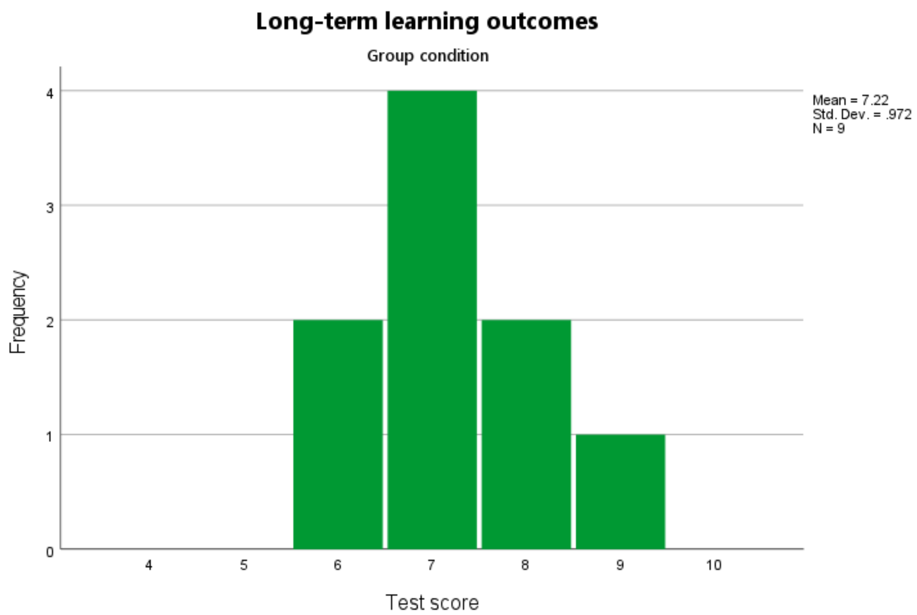


Figure 33. Long-term learning outcomes – group condition.

Figure 34. Long-term learning outcomes – "individual" conditions.

As these figures indicate a difference between these two groups, it was investigated whether they differed significantly. A Shapiro-Wilk test showed no significant departure from normality for learning outcomes in the group condition, $W(.903)$, $p = .273$, but did show significant departure from normality in the individual conditions, $W(.871)$, $p = .005$. Therefore, the analysis proceeded with a Mann-Whitney U-test. This revealed a significant difference ($U = 58$, $N_{group} = 9$, $N_{individual}$ =

24, p = .036) between the group condition (*mean rank* = 11.44) and the individual conditions (*mean rank* = 19.08). See Appendix N for the full SPSS output tables.

5.5.3 Discussion

Similar to the recall of the learning goals, the statistical tests did not show a significant difference between the experimental and control condition regarding learning outcomes. Therefore, there is no indication of the effect of the reflection strategies used by the robot on test performance. These results are therefore not in line with Jones and Castellano (2018) and Jones, Bull, and Castellano (2018), who did show higher learning gains when a robot tutor supported self-regulated learning. However, they included more differences between the conditions and used a more extensive study design. For instance, they used an open learner model and a variety of learning tasks, which learners could select themselves.

When the short-term and long-term results are compared with each other, we see that the scores of the control condition have decreased (-0.28), whereas the scores of the experimental condition have improved (+0.11). Again, it should be noted however that three participants in the experimental condition and one participant in the control condition were missing in the long-term results. When their scores were removed from the short-term scores, the averages change to $M_{experimental}$ = 8.09 and $M_{control}$ = 8.31. This results in a decrease instead of an improvement for the experimental condition (-0.09 instead of +0.11) and slightly smaller decrease for the control condition (-0.23 instead of -0.28). Still, the scores of the control condition decreased more over time than the scores of the experimental condition[27]. Similar to the recall of learning goals, it should be noted that the control condition had better short-term results, so for them there was more content to forget over time.

Again, a significant difference was found between the group condition and the two individual conditions. This could have been caused by the same aspects that were mentioned in the discussion of the results of the recall of learning goals (see Section 5.4.3).

5.6 Results H3: Motivation to eat healthily

See Table 10 below for an overview of the number of times a type of treat was selected per condition. As can be seen, only five out of 28 participants chose the healthy option. Three of those were in the experimental condition and two were in the control condition. They mainly provided reasons such as being vegan or lactose intolerant rather than it being healthier. One participant based their choice on their judgment that raisins are “healthy and tasty”. The remaining 23 participants selected the unhealthy treat. In one of those cases, the participant did add that “I know that raisins are healthier…” before making their choice.

Table 10. Treat choices.

Condition	Selected healthy treat	Selected unhealthy treat
Control	2	12
Experimental	3	11

Discussion: These results indicate that children were unaffected by the presented information about healthy nutrition, irrespective of the condition. This aligns with previous studies, which have found that although children's nutritional knowledge is good, they might not employ it when selecting snacks (Dias & Agante, 2011). Previous research has provided contradictory findings as to whether games can stimulate children to eat healthier. Elementary school children (7-8 years) who played an advergame with healthy food items were more likely to select these items afterwards, compared to children who played the advergame with unhealthy food items (Dias & Agante, 2011). However, primary school children (8-12 years) who played a game that promoted a healthy lifestyle did not make different food choices than children who did not play the game (Folkvord, Haga, & Theben, 2021).

When the treats were distributed in class, it was not noted down whether participants actually made the same choice that they had communicated to the researcher after the learning session. Social pressure could have been present here. Whether that would have led to more children choosing the unhealthy option (following the majority) or healthy option (doing what is desired) remains unclear.

Another uncertainty that remains is to what extent participants understood in which aspects the treats were relatively (un)healthy. Although they were taught during the session that chocolate was an unhealthy snack, it remains unclear to what extent this information translated to the presented choice.

5.7 Robot assessment

The figures below present the robot assessment scores per item.

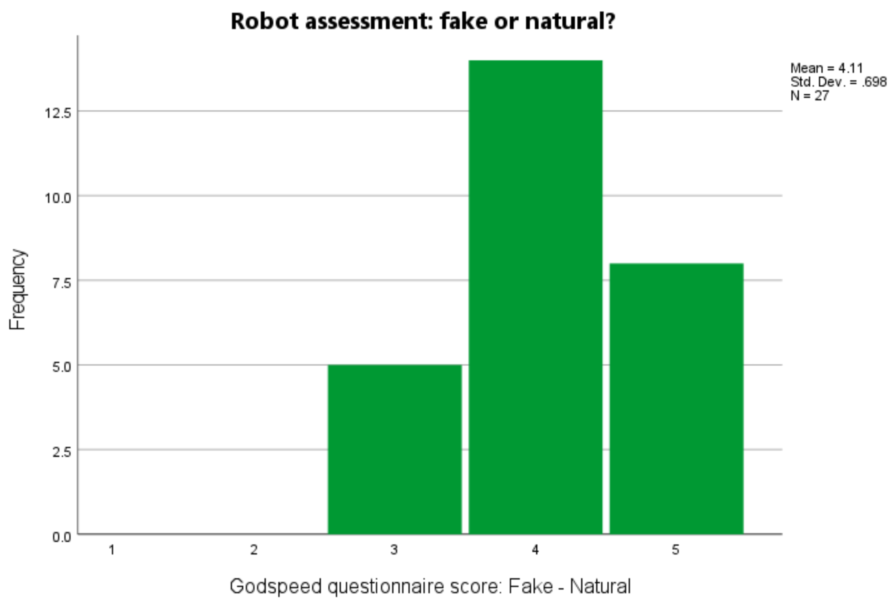


Figure 35. Robot assessment – fake or natural?

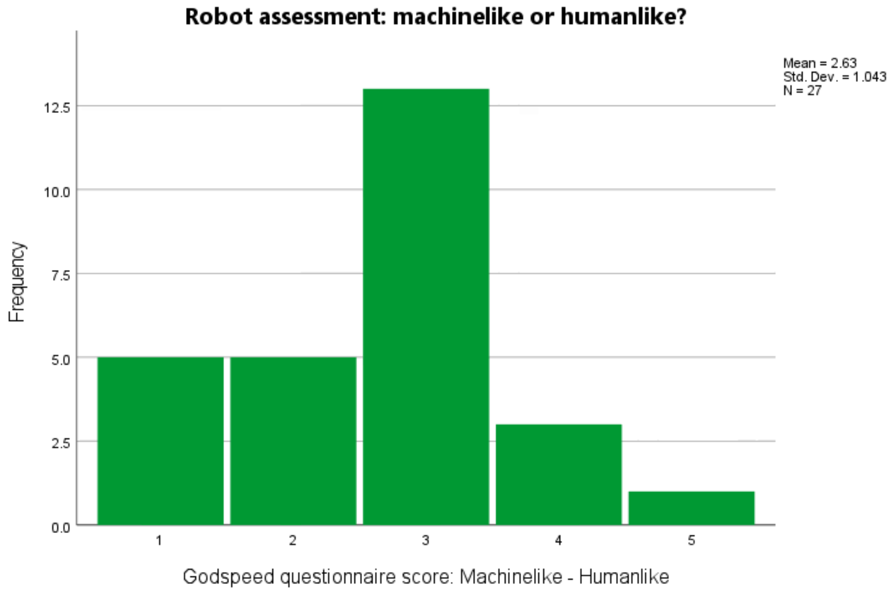


Figure 36. Robot assessment – machinelike or humanlike?

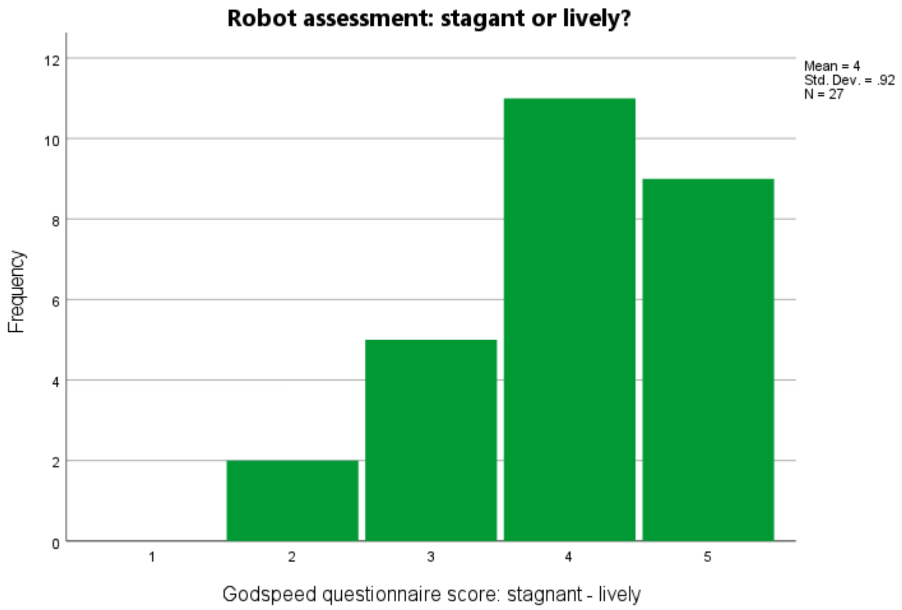


Figure 37. Robot assessment – stagnant or lively?

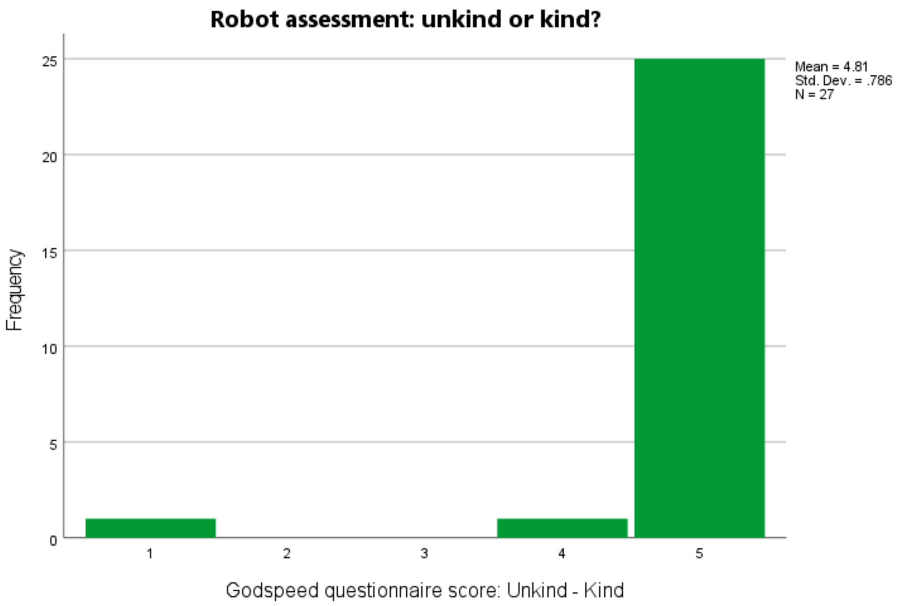


Figure 38. Robot assessment – unkind or kind?

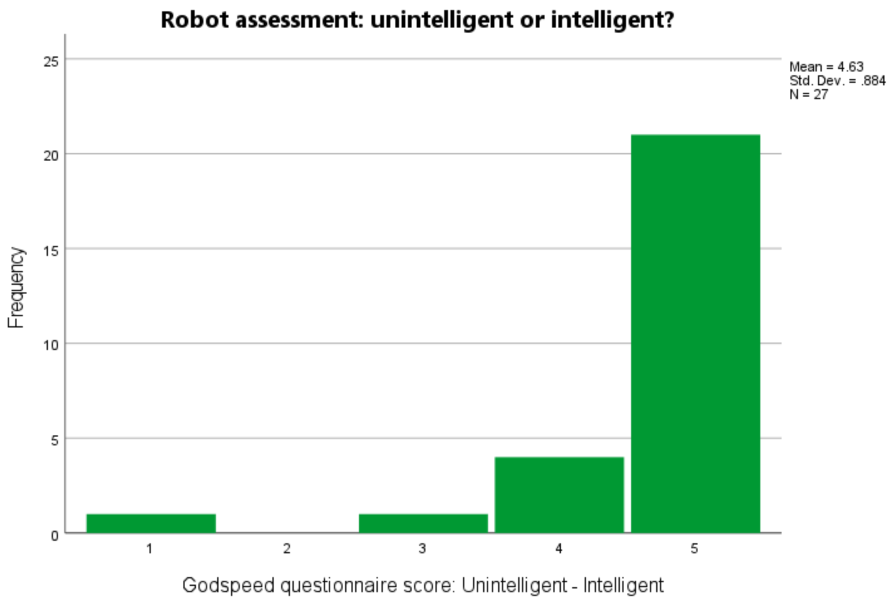


Figure 39. Robot assessment – unintelligent or intelligent?

Table 11 presents the average robot assessment scores per condition. The highest difference between the two conditions is 0.71, where the experimental condition rated the robot as more humanlike than the control condition[28].

[28] Statistical analysis of the differences between the conditions was considered. However, this would require a Bonferroni correction, leaving very little room for significant output. Besides, these comparisons were not part of the goals of this research project. It is therefore a recommendation for future work to research the influence of robot behavior on robot assessment in this particular domain.

Table 11. Average robot assessment score per condition.

	Experimental	Control	Total
Fake - natural	4.08	4.14	4.11
Machinelike - humanlike	3.00	2.29	2.63
Stagnant - lively	4.00	4.00	4.00
Unkind - kind	5.00	4.64	4.81
Unintelligent - intelligent	4.77	4.50	4.63

5.7.1 Discussion

Participants seemed to agree most on finding the robot kind and intelligent. As can be seen in Figure 38 and Figure 39, there was one participant who considered the robot unkind and unintelligent. The corresponding audio recording was played to find an explanation for this. For instance, this could have been one of the sessions where the robot had lost its connection and uttered “I am trying to reconnect”. However, no such abnormality was found.

Overall, participants seemed to assign human characteristics to the robot. It was assessed as considerably natural (4.11), lively (4), kind (4.81), and intelligent (4.63). However, the robot was seen as only slightly more humanlike than machinelike (2.63). For all of these human characteristics, the experimental condition scored higher than the control condition[28]. This could indicate that the reflection strategies made the robot more humanlike. However, it must be noted that for kindness and intelligence, this was caused by a single participant who was the only one to assign a score of 1 to both items (see Section 5.7).

Higher human likeness is generally seen as positive, as a common assumption is that humans prefer to interact with machines in the same way that they interact with other people (Fong, Nourbakhsh, & Dautenhahn, 2003). Anthropomorphism is implemented in robots to increase acceptance of robots and facilitate interaction (Fink, 2012). In robotics, anthropomorphic design refers to three parts: a robot’s shape, behavior, and interaction with the human (DiSalvo, Gemperle, Forlizzi, & Kiesler, 2002). It remains unclear to which extent each of these factors formed participants’ assessments of the robot. For instance, the robot’s interactions with the participant (including the reflection strategies) could have been the main determinant to judge it as machinelike or humanlike, but it could have also been the robot’s shape or gestures.

6 CONCLUSIONS, LIMITATIONS AND FUTURE WORK

6.1 Conclusions

This research project aimed to contribute to the dialogue design of social robots and virtual agents to improve health literacy in children between seven and 12 years old. More specifically, this study aimed to investigate whether reflection strategies, delivered by a social robot, can increase the effectiveness of children's independent learning processes in the domain of healthy nutrition. This study aimed to do so by answering three research questions, one regarding literature (RQ1), one regarding design (RQ2), and one regarding evaluation (RQ3). The main findings and conclusions are discussed per research question below.

6.1.1 Research question 1

RQ1*: What is the current state of the art in dialogue design research on virtual agents and social robots that support learning processes of children (7-12 years) through one-on-one interaction?*

The first research question was answered by conducting a systematic literature review. This resulted in the selection of 38 papers that were published between 2015 and 2021. The results were gathered through five subquestions, which regarded embodiment (Q1), role (Q2), learning domain (Q3), intended effect (Q4), and interventions (Q5). The conclusions are presented in the same order.

Overall, the results presented virtual agents and social robots with various embodiments and various roles, that were implemented in various learning domains, with a variety of intended effects, accomplished by a variety of intervention strategies. The results also revealed a number of limitations. Some of the embodiment styles appeared to be rather static, potentially making the conversations less realistic and impactful. Additionally, authors did not always substantiate why a particular agent/robot role was chosen, especially when it came to the tutor role. Admittedly, the role of the robot in this research project was also most similar to a tutor, and this was mainly a result of practical reasons (e.g., it was convenient to have the robot guide the learner through pre-existing material). However, it remains desirable to decide on the role based on a critical assessment of its effectiveness for the intended outcomes.

In addition, only a small number of studies discussed the applicability of virtual agents and robots in a persuasive domain, such as health, while this is clearly an important topic (see Chapter 1). Furthermore, only four papers seemed to study the stimulation of (self-)reflection processes, while literature points out that this is considered an important component of independent learning (Pintrich, 2000; Zimmerman, 2002). Moreover, a minority of the papers (14x) used a memory of interactions between the child and the agent or robot, while this is an important facilitator of (self-)reflection.

6.1.2 Research question 2

RQ2: *How can a social robot use dialogue to encourage children (7-12 years) to reflect on learning goals and learning progress when learning individually about healthy nutrition?*

The second research question was answered using a human-centered design approach (ISO, 2019), which involved creating personas and scenarios, using design patterns, and implementing reflection strategies. The final deliverable of this research question was a learning session about healthy nutrition, accompanied by a dialogue script for the robot. This session consisted of four phases. The first phase was the "Getting acquainted" phase. The purpose of this phase was to align expectations, make the child at ease, practice talking with the robot, and counter a novelty effect. The interaction design patterns by Ligthart et al. (2019) were used to facilitate bonding between child and robot. The next phase was the "Introduction of the topic" phase. During this phase, the learning goals were introduced and the learner was asked for their personal learning goal. The third phase, "Learning tasks", consisted of the learning activities of the session. The category "edutainment" was deemed appropriate for the selection of these activities. In the context of the current research project, these consisted of a short video and game about healthy nutrition. In between these activities, reflection on learning progress was facilitated. The fourth and final phase, "Rounding up", included a reflection on the session as a whole, on the personal learning goal, and affective appraisal of the learning session.

Throughout the entire session, reflection was encouraged by incorporating a set of reflection strategies in the robot dialogue. These were composed specifically for the current research project. They were mainly based on previous work in the PAL project (Neerincx et al., 2019) and similar studies (Jones, Bull, & Castellano, 2018). The effectiveness of these reflection strategies was consecutively tested during a between-subjects experiment (RQ3).

Besides dialogue, specific robot movements were incorporated in the session design to make the interaction more natural and to draw attention at the right times.

6.1.3 Research question 3

RQ3: *Does the dialogue design for reflection on learning goals and learning progress lead to better reproduction of the learning goals (H1), higher learning outcomes (H2), and more motivation to show healthy behavior (H3)?*

The third research question was answered by conducting a between-subjects experiment. In total, 38 primary school children in third and fourth grade of elementary school participated in a learning session with a NAO robot. In the experimental condition, the proposed dialogue design with reflection strategies was used (see Chapter 3). In the control condition, these reflection strategies were not present. The results and conclusions are summarized below.

H1+H2. Comparisons between the two conditions did not show significant differences, both on short term and long term. Therefore, these two hypotheses were not accepted. This provided no indication of an effect of reflection strategies employed by the robot on recall of learning goals and learning outcomes. This is not in line with prior work by Neerincx et al. (2019), Jones and Castellano (2018), and Jones, Bull, and Castellano (2018). However, when short-term and long-term results were compared with each other, it showed that the scores of the control condition had decreased slightly more over time than the experimental condition. This could be an indication of deeper processing caused by the reflection strategies. Furthermore, a significant difference was found between the group condition and the two individual conditions. This finding could be attributed to several factors, including the fact that the individual conditions filled in the learning goals twice (short-term and long-term), while the group condition only filled it in one week later (long-term).

H3. The results showed that the majority (23 out of 28) of participants favored the unhealthy option, even though they had just learnt about healthy nutrition. They mainly provided reasons such as being vegan or lactose intolerant rather than it being healthier. This implies that children (at least in this age category) are not likely to be influenced by certain knowledge, even when this information was presented by a robot. However, it should be noted that it is not fully known to what extent participants understood in which aspects the treats were relatively (un)healthy.

Robot assessment. Overall, participants assigned human characteristics to the robot. It was assessed as considerably natural (4.11), lively (4), kind (4.81), and intelligent (4.63). However, the robot was seen as only slightly more humanlike than machinelike (2.63). For four of these human characteristics, the experimental condition provided higher averages than the control condition. This could indicate that the reflection strategies made the robot more humanlike. However, statistical analysis was not included in the current data analysis plan[28].

6.2 Contributions

Altogether, this study is in the first place an enrichment of available literature on social robots and virtual agents used for educative purposes. As existing literature mostly focused on learning outcomes only, and not so much on the learner's awareness of their learning process, this study provided insights that were not demonstrated before, for instance regarding the awareness of learning goals. The focus on healthy nutrition as the learning domain made this a unique contribution to this field as well, as most previous research focused on "classical" school subjects, such as mathematics. Thus, it can be said that this study is a first step in studying the effects of robot dialogue on reflective learning processes in the field of health literacy. Although there were no significant results, the evaluation showed that children were motivated to learn with the robot, that they were comfortable talking with the robot, and that they found the robot considerably real, lively, kind, and intelligent. Moreover, their learning performance was good: on average, participants in the individual conditions recalled more than half of the learning goals. In addition, all but one participant in the individual conditions scored at least 6 points in the test questions.

Besides these contributions, the study has also resulted in a number of practical recommendations for similar experiments. The first one is to conduct pilot tests with participants in the right age category. This showed that some of the test questions were too easy, so they were changed to provide more useful measurements. The second one is to distribute the informed consent forms for parents on time and to not expect all parents to provide consent. Two days after handing out these forms, only nine out of 26 forms were returned. After the teacher sent a reminder mail to the parents, nine more forms were handed in before the end of the week. The third recommendation is to have participants initiate the conversation with the robot. In this case, the researcher told them to wave to the robot so that it would "wake up". This helped to make the robot come across as autonomous. The fourth recommendation is to pre-enter as much text as possible in a WoZ set-up. At the time of the first pilot test, the WoZ page (see Section 4.2.2) only contained one text entry box. During this pilot, the text from the script was copied and pasted into the WoZ page one-by-one. This resulted in hiccups in the dialogue that disrupted the conversation flow. The fifth recommendation is to take into account that younger children might not be very skilled with working with a laptop. Apparently, this particular primary school (Harbour Bilingual) trained pupils to work with Chromebooks during forth grade. It was noticed that a number of pupils in third grade indeed struggled to work with the touchpad and mouse.

6.3 Limitations

Literature review. The systematic literature review has three limitations. The first limitation is that only the Scopus database was searched. Although there were reasons to choose this database (see Section 2.2), it might have limited access to relevant papers. The second limitation is that the researcher's decisions in phase S1 and S3 were only partially checked by an external rater (see Section 2.4). Therefore, this process could have been subject to the researcher's subjectivity. The third limitation is that the researcher was the only person involved in the data collection process.

This meant that there was no second check on the data entered in the two literature tables (see Appendix A and Appendix B).

Sample. The experiment also has several limitations. The first limitations regards its sample. Due to limited resources, the sample was limited to 38 participants, or 28 when the improvised group sessions are disregarded. These participants all went to the same primary school in Rotterdam, thereby limiting the generalizability of the results to schools in other areas. Furthermore, individual differences between participants could have affected the validity of the outcomes. An attempt was made to balance these out by having the teacher make two groups that were equal in terms of gender and class performance. However, there were other factors could have interfered too, such as motivation and prior knowledge about the topic outside of the classroom.

Methodology. The experiment consisted of a relatively short learning session (approximately 10 minutes) with two short learning activities (a video and a game). Therefore, there was a small amount of content that could be learnt. As a results, the repertoire of test questions was limited. It could be argued that a longer session with more content and more test questions would result in a more accurate measurement of the effects of the intervention.

Regarding the motivation to eat healthily, the selection of treats was limited to raisins and Smarties. It could have been the case that participants particularly disliked raisins, and that if the healthy options would have been something else, they would have picked it. On the other hand, if the unhealthy option would have been dairy-free, then maybe that would have resulted in more participants picking that option, as participants who chose for raisins indicated that they avoided dairy.

Furthermore, the study design was limited to two conditions that both involved a robot. There are more conditions to think of that could put the results into a more complete perspective. For instance, the robot could have been replaced by a teacher or virtual agent. Or, the participant could have received written instructions only. This could provide a better indication of the effectiveness of the robot's guidance.

Transfer. Finally, the experimental set-up was fully WoZ-based and the robot could not converse autonomously. This would require developing a model, ontology, and implementation of interaction designs. Speech detection would be required to store and use the user's input. However, speech detection in children faces multiple challenges, such as lower accuracy in pronunciation compared to adults (Tian, Tang, Jiang, Tsutsui, & Miyanaga, 2018). Therefore, the proposed dialogue and set-up are not easily transferred to a setting without human supervision and control. This would require thorough examination of potential answers and the establishment of so-called "repair mechanisms" (Ligthart et al., 2019). In addition, this requires thorough ethical considerations of what data is stored, where, how and for how long.

6.4 Future work recommendations

The systematic literature review (see Chapter 2) showed that a limited number of relevant studies focused on self-reflection in learning processes, while this is considered an important component of independent learning (Pintrich, 2000; Zimmerman, 2002). Therefore, more research should be done to determine how to implement stimulation of reflection in child-robot interactions.

Furthermore, a minority of the papers employed a study design with multiple sessions, so long-term effects remain largely unknown. Unfortunately, time constraints limited the current experiment to one session only, but future research projects with a wider timeframe are recommended to employ multiple sessions. Especially for reflection strategies, which are expected to be more effective on the long term, this seems a relevant approach. Having multiple sessions with the robot allows for reflecting on learning progress on previous occasions. For instance, a conversational agent that employed multiple references to shared experiences within a session, increased motivation to engage in diet-related behavior change (Saravanan, Tsfasman, Neerincx, & Oertel, 2022). Moreover, the results of the current experiment showed that the scores of the control condition decreased more over time than the scores of the experimental condition. This could be an indication of the effectiveness of the reflection strategies on the long term, but this is something that needs further investigation.

Finally, the results provided suggestions for research on robot assessment. Since overall, the experimental condition showed higher scores for human likeness, which is generally seen as positive in robotics, it could be interesting to study this in isolation. If such research would indeed confirm a positive correlation between the reflection strategies and human likeness, it would be interesting to see which aspect causes this relationship, e.g., the robot talking more, the content of the questions, the personalization of robot responses, etc.

REFERENCES

Ahmad, M. I., Gao, Y., Alnajjar, F., Shahid, S., & Mubin, O. (2021). Emotion and memory model for social robots: a reinforcement learning based behaviour selection. *Behaviour & Information Technology*, 1-27.

Ali, S., Devasia, N., Park, H. W., & Breazeal, C. (2021). Social robots as creativity eliciting agents. *Frontiers in Robotics and AI*, *8*.

Anikina, O. V., & Yakimenko, E. V. (2015). Edutainment as a modern technology of education. *Procedia-Social and Behavioral Sciences, 166*, 475-479.

Azevedo, R., Cromley, J. G., Moos, D. C., Greene, J. A., & Winters, F. I. (2011). Adaptive content and process scaffolding: A key to facilitating students' self-regulated learning with hypermedia. *Psychological Test and Assessment Modeling*, *53*(1), 106.

Bartneck, C., Croft, E., & Kulić, D. (2008). Measuring the anthropomorphism, animacy, likeability, perceived intelligence and perceived safety of robots.

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, *1*(1), 71-81.

Baxter, P., Ashurst, E., Read, R., Kennedy, J., & Belpaeme, T. (2017). Robot education peers in a situated primary school study: Personalisation promotes child learning. *PloS one*, *12*(5), e0178126.

Castellano, G., De Carolis, B., D'Errico, F., Macchiarulo, N., & Rossano, V. (2021). PeppeRecycle: improving children's attitude toward recycling by playing with a social robot. *International Journal of Social Robotics*, *13*(1), 97-111.

Coninx, A., Baxter, P., Oleari, E., Bellini, S., Bierman, B., Henkemans, O. B., et al. (2016). Towards long-term social child-robot interaction: using multi-activity switching to engage young users. *J. Hum. Robot Interact. 5,* 32–67. doi: 10.5898/JHRI.5.1.Coninx

Cooper, A. (1999). *The Inmates Are Running the Asylum.* Indianapolis, IN: Sams.

Davison, D. P., Wijnen, F. M., Charisi, V., van der Meij, J., Reidsma, D., & Evers, V. (2021). Words of encouragement: how praise delivered by a social robot changes children's mindset for learning. *Journal on multimodal user interfaces*, *15*(1), 61-76.

Dias, M., & Agante, L. (2011). Can advergames boost children's healthier eating habits? A comparison between healthy and non-healthy food. *Journal of Consumer Behaviour, 10*(3), 152-160.

DiSalvo, C. F., Gemperle, F., Forlizzi, J., & Kiesler, S. (2002). All robots are not created equal: the design and perception of humanoid robot heads. In *Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques* (pp. 321-326).

Duncan, S. C. (1993). The role of cognitive appraisal and friendship provisions in adolescents' affect and motivation toward activity in physical education. *Research quarterly for exercise and sport, 64*(3), 314-323.

Fink, J. (2012). Anthropomorphism and human likeness in the design of robots and human-robot interaction. In *International conference on social robotics* (pp. 199-208). Springer, Berlin, Heidelberg.

Folkvord, F., Haga, G., & Theben, A. (2021). The Effect of a Serious Health Game on Children's Eating Behavior: Cluster-Randomized Controlled Trial. *JMIR Serious Games, 9*(3), e23050.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and autonomous systems, 42*(3-4), 143-166.

Guneysu, A., & Arnrich, B. (2017). Socially assistive child-robot interaction in physical exercise coaching. In *2017 26th IEEE international symposium on robot and human interactive communication (RO-MAN)* (pp. 670-675). IEEE.

Health~Holland (n.d.). *Data-driven blended care solution with virtual buddy for child health.* Retrieved March 23, 2022, from Data-driven blended care solution with virtual buddy for child health | Health~Holland (health-holland.com)

Imbernón, L. E., Manjarrés, A., & De La Paz López, F. (2016). ARTIE: An integrated environment for the development of affective robot tutors*. Frontiers in computational neuroscience, 10,* 77.

International Organization for Standardization (2019). Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems (ISO Standard No. 9241-210:2019). Retrieved, June 22 2022, from https://www.iso.org/standard/77520.html

Jones, A., Bull, S., & Castellano, G. (2018). "I know that now, i'm going to learn this next" promoting self-regulated learning with a robotic tutor. *International Journal of Social Robotics, 10*(4), 439-454.

Jones, A., & Castellano, G. (2018). Adaptive robotic tutors that support self-regulated learning: A longer-term investigation with primary school children. *International Journal of Social Robotics*, *10*(3), 357-370.

Konijn, E. A., & Hoorn, J. F. (2020). Robot tutor and pupils' educational ability: Teaching the times tables. *Computers & Education*, *157*, 103970.

Latham, G. P. (2003). Goal setting: A five-step approach to behavior change. *Organizational Dynamics, 32*(3), 309-318.

Leite, I., McCoy, M., Lohani, M., Ullman, D., Salomons, N., Stokes, C., ... & Scassellati, B. (2015). Emotional storytelling in the classroom: Individual versus group interaction between children and robots. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 75-82).

Leite, I., McCoy, M., Lohani, M., Ullman, D., Salomons, N., Stokes, C., ... & Scassellati, B. (2017). Narratives with robots: The impact of interaction context and individual differences on story recall and emotional understanding. *Frontiers in Robotics and AI, 4*, 29.

Li, H., Gobert, J., & Dickler, R. (2019). Testing the robustness of inquiry practices once scaffolding is removed. In *International Conference on Intelligent Tutoring Systems* (pp. 204-213). Springer, Cham.

Ligthart, M., Fernhout, T., Neerincx, M. A., van Bindsbergen, K. L., Grootenhuis, M. A., & Hindriks, K. V. (2019, May). A child and a robot getting acquainted-interaction design for eliciting self-disclosure. In *Proceedings of the 18th international conference on autonomous agents and multiagent systems* (pp. 61-70).

Looije, R., van der Zalm, A., Neerincx, M. A., & Beun, R. J. (2012). Help, I need some body the effect of embodiment on playful learning. In *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication (pp. 718-724).* IEEE.

Lunenburg, F. C. (2011). Goal-setting theory of motivation. *International journal of management, business, and administration*, *15*(1), 1-6.

Maeda, R., Even, J., & Kanda, T. (2019). Can a social robot encourage children's self-study? In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 1236-1242). IEEE.

Melanson, K. J. (2008). Nutrition review: lifestyle approaches to promoting healthy eating for children. *American journal of lifestyle medicine, 2*(1), 26-29.

Meyer, B., Haywood, N., Sachdev, D., & Faraday, S. (2008). Independent learning. Learning and Skills Network, Research Report DCSF-RR051. Retrieved March 31, 2022, from https://www.associationforpsychologyteachers.com/uploads/4/5/6/6/4566919/independence_learning_lit_review.pdf

Michaelis, J. E., & Mutlu, B. (2017). Someone to read with: Design of and experiences with an in-home learning companion robot for reading. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 301-312).

Michaelis, J. E., & Mutlu, B. (2018). Reading socially: Transforming the in-home reading experience with a learning-companion robot. *Science Robotics*, *3*(21), eaat5999.

Nardi, B. A. (1992). The use of scenarios in design. *ACM SIGCHI Bulletin, 24*(4), 13-14.

Nederlands Jeugdinstituut (2017). Gevolgen van ongezonde voeding en verstoord eetgedrag. Retrieved March 25, 2022, from https://www.ncj.nl/richtlijnen/alle-richtlijnen/richtlijn/?richtlijn=4&rlpag=559

Nederlands Jeugdinstituut (2022a). Cijfers over eten. Retrieved March 24, 2022, from https://www.nji.nl/cijfers/eten

Nederlands Jeugdinstituut (2022b). Cijfers over overgewicht. Retrieved March 24, 2022, from https://www.nji.nl/cijfers/overgewicht

Neerincx, M. A., Van Vught, W., Blanson Henkemans, O., Oleari, E., Broekens, J., Peters, R., ... & Bierman, B. (2019). Socio-cognitive engineering of a robotic partner for child's diabetes self-management. *Frontiers in Robotics and AI, 6*, 118.

Nielen, T. M., Smith, G. G., Sikkema-de Jong, M. T., Drobisz, J., van Horne, B., & Bus, A. G. (2018). Digital guidance for susceptible readers: Effects on fifth graders' reading motivation and incidental vocabulary learning. *Journal of Educational Computing Research*, *56*(1), 48-73.

Ocaña, J. M., Morales-Urrutia, E. K., Pérez-Marín, D., & Pizarro, C. (2020). Can a learning companion be used to continue teaching programming to children even during the COVID-19 pandemic? *Ieee Access*, *8*, 157840-157861.

Occupational Medicine & Health Affairs (n.d.). *Child Health Education.* Retrieved March 21, 2022, from https://www.omicsonline.org/scholarly/child-health-education-journals-articles-ppts-list.php

Onderwijsinspectie (n.d.). Lerarentekort en kansenongelijkheid. Retrieved April 4, 2022, from https://www.onderwijsinspectie.nl/onderwerpen/staat-van-het-onderwijs/trends-in-het-onderwijs/primair-onderwijs/lerarentekort-en-kansenongelijkheid

Ong, E. C. J., & Soriano, Z. C. (2015). A conversational agent to shift students' affect state. In *Principles and Practice of Multi-Agent Systems* (pp. 86-97). Springer, Cham.

Pintrich (2000). The role of goal orientation in self-regulated learning. In M Boekaerts, PR Pintrich and M Zeidner (eds.), *Handbook of self-regulation* (pp 451- 502). San Diego: Academic Press.

Ramachandran, A., Huang, C. M., Gartland, E., & Scassellati, B. (2018). Thinking aloud with a tutoring robot to enhance learning. In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction* (pp. 59-68).

Ramachandran, A., Huang, C. M., & Scassellati, B. (2017). Give me a break! Personalized timing strategies to promote learning in robot-child tutoring. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 146-155).

Ramachandran, A., Litoiu, A., & Scassellati, B. (2016). Shaping productive help-seeking behavior during robot-child tutoring interactions. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 247-254). IEEE.

Rijksoverheid (n.d.). *Hoe is de groep van mijn kind op de basisschool samengesteld?* Retrieved January 31, 2022, from https://www.rijksoverheid.nl/onderwerpen/basisonderwijs/vraag-en-antwoord/hoe-zijn-de-groepen-in-het-basisonderwijs-bo-samengesteld

Robotics Today (n.d). *JULIA*. Retrieved July 4, 2022, from https://www.roboticstoday.com/robots/julia

Ros, R., Oleari, E., Pozzi, C., Sacchitelli, F., Baranzini, D., Bagherzadhalimi, A., ... & Demiris, Y. (2016). A motivational approach to support healthy habits in long-term child–robot interaction. *International Journal of Social Robotics*, *8*(5), 599-617.

Ryan, R. M., and Deci, E. L. (2017). *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness*. New York, NY: Guilford Publications.

Sandygulova, A., Johal, W., Zhexenova, Z., Tleubayev, B., Zhanatkyzy, A., Turarova, A., ... & Dillenbourg, P. (2020). Cowriting kazakh: learning a new script with a robot. In *2020 15th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 113-120). IEEE.

Saravanan, A., Tsfasman, M., Neerincx, M.A., and Oertel, C. (2022). Giving Social Robots a Conversational Memory for Motivational Experience Sharing. *31st IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN).*

Sebo, S., Stoll, B., Scassellati, B., & Jung, M. F. (2020). Robots in groups and teams: a literature review. *Proceedings of the ACM on Human-Computer Interaction, 4*(CSCW2), 1-36.

Sinoo, C., van Der Pal, S., Henkemans, O. A. B., Keizer, A., Bierman, B. P., Looije, R., & Neerincx, M. A. (2018). Friendship with a robot: Children's perception of similarity between a robot's physical and virtual embodiment that supports diabetes self-management. *Patient education and counseling, 101*(7), 1248-1255.

Song, H., Zhang, Z., Barakova, E. I., Ham, J., & Markopoulos, P. (2020). Robot role design for implementing social facilitation theory in musical instruments practicing. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 253-260).

Song, H., Barakova, E. I., Markopoulos, P., & Ham, J. (2021). Personalizing HRI in Musical Instrument Practicing: The Influence of Robot Roles (Evaluative Versus Nonevaluative) on the Child's Motivation for Children in Different Learning Stages. *Frontiers in Robotics and AI*, 282.

Swatevacharkul, R. (2021). L2 Motivational Self System: A Case of Thai EFL Senior High School Students. 3L: *Southeast Asian Journal of English Language Studies, 27*(4).

Tazhigaliyeva, N., Diyas, Y., Brakk, D., Aimambetov, Y., & Sandygulova, A. (2016). Learning with or from the robot: exploring robot roles in educational context with children. In *International Conference on Social Robotics* (pp. 650-659). Springer, Cham.

Tellols, D., Lopez-Sanchez, M., Rodríguez, I., Almajano, P., & Puig, A. (2020). Enhancing sentient embodied conversational agents with machine learning. *Pattern Recognition Letters*, *129*, 317-323.

Tian, Y., Tang, J., Jiang, X., Tsutsui, H., & Miyanaga, Y. (2018). Accuracy on Children's Speech Recognition under Noisy Circumstances. In *2018 18th International Symposium on Communications and Information Technologies (ISCIT)* (pp. 101-104). IEEE.

Tulli, S., Couto, M., Vasco, M., Yadollahi, E., Melo, F., & Paiva, A. (2020, November). Explainable Agency by Revealing Suboptimality in Child-Robot Learning Scenarios. In *International Conference on Social Robotics* (pp. 23-35). Springer, Cham.

Zimmerman (2002). Becoming a self-regulated learner: an overview. *Theory into Practice, 41*(2), 64-72.

APPENDIX

Appendix A Systematic literature review: General information table

Part A (1/2)

Authors	Year	Title	Publication type	Country[1]
Ahmad, Gao, Alnajjar, Shahid, Mubin	2021	Emotion and memory model for social robots: a reinforcement learning based behaviour selection	Journal article	NA
Ahmad, Mubin	2018	Emotion and memory model to promote mathematics learning - An exploratory long-term study	Conference paper	NA
Ali, Devasia, Park, Breazeal	2021	Social Robots as Creativity Eliciting Agents	Journal article	US
Baxter, Ashurst, Read, Kennedy, Beplaeme	2017	Robot education peers in a situated primary school study: Personalisation promotes child learning	Journal article	UK
Castellano, De Carolis, D'Errico, Macchiarulo, Rossano	2021	PeppeRecycle: Improving Children's Attitude Toward Recycling by Playing with a Social Robot	Journal article	Italy
Chandra, Dillenbourg, Paiva	2020	Children Teach Handwriting to a Social Robot with Different Learning Competencies	Journal article	Portugal
Davison, Wijnen, Charisi, van der Meij, Reidsma, Evers	2021	Words of encouragement: how praise delivered by a social robot changes children's mindset for learning	Journal article	NL
De Carolis, D'Errico, Macchiarula, Palestra	2020	Socially Inclusive Robots: Learning Culture-Related Gestures by Playing with Pepper	Conference paper	Italy
Guneysu, Arnrich	2017	Socially assistive child-robot interaction in physical exercise coaching	Conference paper	NA
Imbernon Cuadrado, Manjarres Riesco, De La Paz Lopez	2016	ARTIE: An Integrated Environment for the Development of Affective Robot Tutors	Journal article	NA
Jones, Bull, Castellano	2018	"I Know That Now, I'm Going to Learn This Next" Promoting Self-regulated Learning with a Robotic Tutor	Journal article	UK
Jones, Castellano	2018	Adaptive Robotic Tutors that Support Self-Regulated Learning: A Longer-Term Investigation with Primary School Children	Journal article	UK
Kennedy, Baxter, Senft, Belpaeme	2016	Social robot tutoring for child second language learning	Conference paper	UK
Konijn, Hoorn	2020	Robot tutor and pupils' educational ability: Teaching the times tables	Journal article	NL
Li, Gobert, Dickler	2019	Testing the robustness of inquiry practices once scaffolding is removed	Conference paper	US
Lubold, Walker, Pon-Barry, Flores, Ogan	2018	Using iterative design to create efficacy-building social experiences with a teachable robot	Conference paper	US
Lubold, Walker, Pon-Barry, Ogan	2018	Automated pitch convergence improves learning in a social, teachable robot for middle school mathematics	Conference paper	US
Maeda, Even, Kanda	2019	Can a Social Robot Encourage Children's Self-Study?	Conference paper	Japan
Michaelis, Mutlu	2017	Someone to read with: Design of and experiences with an in-home learning companion robot for reading	Conference paper	US
Michaelis, Mutlu	2018	Reading socially: Transforming the in-home reading experience with a learning-companion robot	Journal article	US
Michaelis, Mutlu	2019	Supporting interest in science learning with a social robot	Conference paper	US
Nielen, Smith, Sikkema-de Jong, Drobisz, van Horne, Bus	2018	Digital Guidance for Susceptible Readers: Effects on Fifth Graders' Reading Motivation and Incidental Vocabulary Learning	Journal article	NL
Ocana, Morales-Urrutia, Perez-Marin, Pizarro	2020	Can a learning companion be used to continue teaching programming to children even during the COVID-19 pandemic?	Journal article	Ecuador
Ong, Soriano	2016	A conversational agent to shift students' affect state	Conference paper	Filipines
Park, Rosenberg-Kima, Rosenberg, Gordon, Breazeal	2017	Growing Growth Mindset with a Social Robot Peer	Conference paper	Austria?
Ramachandran, Huan, Gartland, Scassellati	2018	Thinking Aloud with a Tutoring Robot to Enhance Learning	Conference paper	US?
Ramachandran, Litoiu, Scassellati	2016	Shaping productive help-seeking behavior during robot-child tutoring interactions	Conference paper	US?

Part A (2/2)

N[2]	Gender	Age range + M	Cultural backgr	Nr of conditions[3]	Betw/Within	Study period[4]	Nr of sessions[5]	Length of session[6]	Pre/post test[7]	Quant/qual?[8]
24	12M 12F	10-12, 10.69	NA	1	Within	4 weeks	4	24 minutes	Pre+post	Quant
20	10M 10F	8-9, 8.4	NA	2	Between	2 weeks	3	11 minutes	Post each sess	Quant
43	24M 19F	6-10, 8.11	NA	2	Between	NA	1	NA	Pre+post	Both
59	24M 35F	7-8, NA	NA	2	Between	2 weeks	M = 2?	5 minutes	Pre+post	Both
51	26M 25F	7-9, NA	NA	1	Neither	NA	1	NA	Pre	Both
25	12M 13F	7-9, 7.92-8.08	NA	2	Between	1 month	4	13-15 minutes	Pre+post	Both
44	26M 18F	6-10, 7.0-7.1	NA	2 (+1 baseline)	Between	Several months	2	5-15 minutes	Pre+post	Quant
10	NA	6-10, NA	NA	1	Neither	NA	1	NA	Neither	Quant
19	NA	NA, 8.4	NA	1	Neither	NA	1	NA	Neither	Quant
2	1M 1F	8-11, 9.5	NA	1	Neither	NA	1	NA	Neither	Both
80	46M 34F	10-12, NA	NA	4	Between	NA	1	Approx 11 min	Pre+post	Quant
24	14M 10F	10-12, NA	NA	2	Between	1 month	4	Approx 20 min	Pre+post	Quant
67	30M, 37F	NA, 8.8	NA	3, one control	Between	8 days	1	M = 11 min	Pre+post	Quant
86	45M, 41F	8-10, 8.76	NA	2	Between	3 weeks	3	5 minutes	Pre+post	Quant
107	NA	NA	*	2	Within	170 days	4	NA	Neither	Quant
14	10M, 4F	10-13, 11	NA	3	Between	NA	1	12-35 minutes	Pre+post	Both
69	30-33M, 3	NA, ± 11.25	NA	3	Between	NA	1	30 minutes	Pre+post	Quant
22	16M, 6F	9, 9	NA	2, one control	Within	NA	1	Max 25 min	Neither	Quant
8	5M, 3F	11-12, 11.6	NA	1	Neither	NA	1	Approx 30 min	Pre	Qual
24	10M, 14F	10-12, NA	NA	2, one control	Between	13 days	M = 8,3	M = 20.72 min	Pre+post	Both
58	31-36M, 2	10-12, ± 10.9	NA	2	Between	NA	1	NA	Pre+post	Quant
146	70M, 76F	10-12, 11.10	NA	3, two control	Between	2 months	NA	NA	Pre+post	Quant
137	Almost 50	10-12, NA	NA	1	Neither	10 months	NA	Approx 1 hour	Pre+post	Both
17	NA	8-10, NA	NA	1	Neither	NA	1	NA	Neither	Qual
40	23M, 17F	5-9, 6.75	NA	2	Between	NA	1	NA	Pre+post	Quant
52	38M, 14F	NA, 11.21	NA	4, two control	Between	NA	1	Approx 30 min	Pre+post (+ delayed post)	Quant
29	16M, 13F	NA, 10.68-10.9	**	2, one control	Between	Approx 2 weeks	4	NA	Pre+post	Quant

* Entire school: 39.9% white, 20.6 Hispanic, 23.5% Asian. 11% black, 5.7% >= 2 races

** 7.1-13.3% Asian, 60-78.6% Caucasian, 0-13.3% Hispanic, 6.7-7.1% more than 1 ethnicity, 6.7%-7.1% did not report

Part B (1/2)

Authors	Year	Title	Publication type	Country[1]
Ramachandran, Litoiu, Scassellati	2016	Shaping productive help-seeking behavior during robot-child tutoring interactions	Conference paper	US?
Ramachandran, Huan, Scassellati	2017	Give Me a Break! Personalized Timing Strategies to Promote Learning in Robot-Child Tutoring	Conference paper	US?
Ros et al.	2016	A Motivational Approach to Support Healthy Habits in Long-term Child–Robot Interaction	Journal article	Italy
Sandygulova et al.	2020	CoWriting Kazakh: Learning a new script with a robot	Conference paper	Kazakhstan
Song, Barakova, Markopoulos, Ham	2021	Personalizing HRI in Musical Instrument Practicing: The Influence of Robot Roles (Evaluative Versus Nonevaluative) on the Child's Motivation for Children in Different Learning Stages	Journal article	NL
Song, Zhang, Barakova, Ham, Markopoulos	2020	Robot role design for implementing social facilitation theory in musical instruments practicing	Conference paper	NL
Tazhigaliyeva, Diyas, Brakk, Aimambetov, Sandygulova	2016	Learning with or from the robot: Exploring robot roles in educational context with children	Conference paper	Kazakhstan
Tellols, Lopez-Sanchez, Rodriguez, Almajano, Puig	2020	Enhancing sentient embodied conversational agents with machine learning	Journal article	Spain
Tulli, Couto, Vasco, Yadollahi, Melo, Paiva	2020	Explainable Agency by Revealing Suboptimality in Child-Robot Learning Scenarios	Conference paper	Portugal
Wijnen et al.	2019	Now We're Talking: Learning by Explaining Your Reasoning to a Social Robot	Journal article	NL
Yueh, Lin, Wang, Fu	2020	Reading with robot and human companions in library literacy activities: A comparison study	Journal article	Taiwan
Zhexenova et al.	2020	A Comparison of Social Robot to Tablet and Teacher in a New Script Learning Context	Journal article	Kazakhstan

Part B (2/2)

N[2]	Gender	Age range + M	Cultural backgr	Nr of conditions[3]	Betw/Within	Study period[4]	Nr of sessions[5]	Length of session[6]	Pre/post test[7]	Quant/qual?[8]
29	16M, 13F	NA, 10.68-10.9	**	2, one control	Between	Approx 2 weeks	4	NA	Pre+post	Quant
38	25M, 13F	NA, 8.53	NA	3	Between	NA	1	40 min	Pre+post	Quant
84	41M, 43F	9-11, NA	NA	4, two control	Between	3 weeks	3	Max 30 min	Pre+post	Both
67	35M, 32F	8-11, NA	NA	2	Between	NA	1	NA	Pre+post	Both
31	16M, 15F	9-12, NA	NA	3, one control	Within	NA	1	10 min	Post	Both
20	10M, 10F	5-12, 9.85	NA	2	Between	NA	1	15 min	Post	Both
25	14M, 11F	9-10, NA	NA	2	Between	NA	1	NA	Pre+post	Quant
45	NA	10-12, NA	NA	2 versions	Between/neith	8 days	NA	NA	Neither	Both
32	17M, 15F	NA, 7.03	NA	2	Between	NA	1	NA	Pre+post	Quant
46	19M, 27F	NA, 8.1-8.2	NA	2, one control	Between	NA	1	Approx 20 min	Post	Both
36	20M, 16F	NA (3rd grade)	NA	3, two control	Between	NA	1	60-80 min	Post	Both
62	31M, 31F	7-9, NA	NA	3, two control	Between	NA	1	Approx 20-30 min	Pre+post	Both

Footnotes
1 Country in which the evaluation is conducted
2 Number of participants that actually participated (e.g. without dropouts)
3 Control condition means that no agent or robot was present
4 Total period of study
5 Per participant, together with agent/robot
6 With the robot present
7 Of learning outcomes (in the broadest sense)
8 A survey is seen as a quantative measure

Figure 40. Footnotes for general information table.

Appendix B Systematic literature review: Structured summary table

Part A (1/2)

Authors	Year	Title	1. Embodiment	2. Role	3. Learning domain	4. Intended effect
Ahmad, Gao, Alnajjar, Shahid, Mubin	2021	Emotion and memory model for social robots: a reinforcement learning based behaviour selection	NAO robot	Tutor + peer	Vocabulary learning	Social engagement + Learning gains
Ahmad, Mubin	2018	Emotion and memory model to promote mathematics learning - An exploratory long-term study	NAO robot	Tutor	Mathematics	Social engagement + Learning gains
Ali, Devasia, Park, Breazeal	2021	Social Robots as Creativity Eliciting Agents	Jibo robot	Peer	Creativity	Creativity
Baxter, Ashurst, Read, Kennedy, Beplaeme	2017	Robot education peers in a situated primary school study: Personalisation promotes child learning	NAO robot	Peer	Multiplication tables + history	Learning gains
Castellano, De Carolis, D'Errico, Macchiarulo, Rossano	2021	PeppeRecycle: Improving Children's Attitude Toward Recycling by Playing with a Social Robot	Pepper robot	Peer	Waste recycling	Attitude change
Chandra, Dillenbourg, Paiva	2020	Children Teach Handwriting to a Social Robot with Different Learning Competencies	NAO robot	Tutee	Handwriting skills	Learning gains
Davison, Wijnen, Charisi, van der Meij, Reidsma, Evers	2021	Words of encouragement: how praise delivered by a social robot changes children's mindset for learning	Zeno robot	Peer	Physics	Growth mindset
De Carolis, D'Errico, Macchiarula, Palestra	2020	Socially Inclusive Robots: Learning Culture-Related Gestures by Playing with Pepper	Pepper robot	Tutor	Culture-related gesture learning	Learning gains
Guneysu, Arnrich	2017	Socially assistive child-robot interaction in physical exercise coaching	NAO robot	Tutor	Physical exercise	Motivation + Learning gains
Imbernon Cuadrado, Manjarres Riesco, De La Paz Lopez	2016	ARTIE: An Integrated Environment for the Development of Affective Robot Tutors	NAO robot	Tutor	General (but evaluation: programming)	Engagement + motivation
Jones, Bull, Castellano	2018	"I Know That Now, I'm Going to Learn This Next" Promoting Self-regulated Learning with a Robotic Tutor	NAO robot	Tutor	Geography	Self-regulation skills + Learning gains
Jones, Castellano	2018	Adaptive Robotic Tutors that Support Self-Regulated Learning: A Longer-Term Investigation with Primary School Children	NAO robot	Tutor	Geography	Self-regulation skills + Learning gains
Kennedy, Baxter, Senft, Belpaeme	2016	Social robot tutoring for child second language learning	NAO robot	Tutor	Second language tutoring	Learning gains
Konijn, Hoorn	2020	Robot tutor and pupils' educational ability: Teaching the times tables	NAO robot	Tutor	Multiplication tables	Learning gains
Li, Gobert, Dickler	2019	Testing the robustness of inquiry practices once scaffolding is removed	Virtual agent	Tutor	Biology	Inquiry skills
Lubold, Walker, Pon-Barry, Flores, Ogan	2018	Using iterative design to create efficacy-building social experiences with a teachable robot	NAO robot	Tutee	Mathematics	Social experience + self-efficacy (-> learning + motivation)
Lubold, Walker, Pon-Barry, Ogan	2018	Automated pitch convergence improves learning in a social, teachable robot for middle school mathematics	NAO robot	Tutee	Mathematics	Rapport + Learning gains
Maeda, Even, Kanda	2019	Can a Social Robot Encourage Children's Self-Study?	Sota robot	Tutor	Mathematics + literacy	Extended concentration
Michaelis, Mutlu	2019	Someone to read with: Design of and experiences with an in-home learning companion robot for reading	Customized Maki robot	Tutor	Science education	Situational interest + learning
Michaelis, Mutlu	2018	Reading socially: Transforming the in-home reading experience with a learning-companion robot	Customized Maki robot	Tutor	Reading	Reading time + motivation + comprehension
Michaelis, Mutlu	2017	Supporting interest in science learning with a social robot	Customized Maki robot	Tutor	Reading	N/A (probe study)

Part A (2/2)

5. Intervention (how to reach int. eff.)	6. > 1 session?	7. >1 condition?
(Improved) emotion and memory model of users' emotional events, facereading	Yes	No
Emotion and memory model of users' emotional events, facereading	Yes	Yes
Creativity demonstration and creativity scaffolding	No	Yes
Personalized vs non-personalized behavior	Yes	Yes
Recycling game against robot	No	No
Non-learning vs continuous learning vs personalized learning tutee	No	Yes
Praise	Yes	Yes
Showing gestures and providing feedback	No	No
Praise + corrective feedback	No	No
Emotion detection model + behavior model for robot	No	No
Open learner model, personalized and adaptive scaffolding	No	Yes
Open learner model, personalized and adaptive scaffolding	Yes	Yes
'Low' and 'high' verbal availability	No	Yes
Neutral vs social behavior (differed per proficiency level)	Yes	Yes
Adaptive scaffolding	No (not with agent)	Yes
Dialogue design (degree to which the robot is understanding and learning)	No	Yes
Multi-feature entrainment + social dialogue	No	Yes
Behavioral model triggering specific questions	No	Yes
Social adept behavior	No	Yes
Setting + monitoring goals, recommendations, interest, refer to prev. interactions, etc	Yes	Yes
Setting + monitoring goals, recommendations, interest, refer to prev. interactions, etc	Yes	No

Part B (1/2)

Authors	Year	Title	1. Embodiment	2. Role	3. Learning domain	4. Intended effect
Nielen, Smith, Sikkema-de Jong, Drobisz, van Horne, Bus	2018	Digital Guidance for Susceptible Readers: Effects on Fifth Graders' Reading Motivation and Incidental Vocabulary Learning	Virtual agent	Tutor	Reading	Motivation + Learning gains
Ocana, Morales-Urrutia, Perez-Marin, Pizarro	2020	Can a learning companion be used to continue teaching programming to children even during the COVID-19 pandemic?	Virtual agent	Tutor	Programming	Learning gains
Ong, Soriano	2016	A conversational agent to shift students' affect state	Virtual agent	Peer	Reading (comprehension)	Positive affect
Park, Rosenberg-Kima, Rosenberg, Gordon, Breazeal	2017	Growing Growth Mindset with a Social Robot Peer	Tega robot	Peer	Problem-solving (puzzles)	Growth mindset
Ramachandran, Huan, Gartland, Scassellati	2018	Thinking Aloud with a Tutoring Robot to Enhance Learning	NAO robot	Tutor	Mathematics	Engagement + Learning gains
Ramachandran, Litoiu, Scassellati	2016	Shaping productive help-seeking behavior during robot-child tutoring interactions	NAO robot	Tutor	Mathematics	Effective help-seeking behavior + Learning gains
Ramachadran, Huang, Scassellati	2017	Give Me a Break! Personalized Timing Strategies to Promote Learning in Robot-Child Tutoring	NAO robot	Tutor	Mathematics	Learning gains
Ros et al.	2016	A Motivational Approach to Support Healthy Habits in Long-term Child–Robot Interaction	NAO robot	Tutor	Healthy lifestyle	Engagement + motivation
Sandygulova et al.	2020	CoWriting Kazakh: Learning a new script with a robot	NAO robot	Tutee	Script learning	Learning gains
Song, Barakova, Markopoulos, Ham	2021	Personalizing HRI in Musical Instrument Practicing: The Influence of Robot Roles (Evaluative Versus Nonevaluative) on the Child's Motivation for Children in Different Learning Stages	SocibotMini robot	Tutor OR peer	Musical instrument practicing	Motivation
Song, Zhang, Barakova, Ham, Markopoulos	2020	Robot role design for implementing social facilitation theory in musical instruments practicing	SocibotMini robot	Tutor OR peer	Musical instrument practicing	Motivation
Tazhigaliyeva, Diyas, Brakk, Aimambetov, Sandygulova	2016	Learning with or from the robot: Exploring robot roles in educational context with children	NAO robot	Tutor OR peer	Programming	Engagement + Learning gains
Tellols, Lopez-Sanchez, Rodriguez, Almajano, Puig	2020	Enhancing sentient embodied conversational agents with machine learning	Virtual agent	Tutor	Environmental sustainability	Attitude/behavior change
Tulli, Couto, Vasco, Yadollahi, Melo, Paiva	2020	Explainable Agency by Revealing Suboptimality in Child-Robot Learning Scenarios	NAO robot	Tutor OR peer	Logical + mathematical thinking	Lower pereived difficulty + higher efficiency
Wijnen et al.	2019	Now We're Talking: Learning by Explaining Your Reasoning to a Social Robot	Zeno robot	Tutor	Physics	Explanatory behavior -> learning gains
Yueh, Lin, Wang, Fu	2020	Reading with robot and human companions in library literacy activities: A comparison study	Julia robot	Peer	Reading	Social interaction + reading support
Zhexenova et al.	2020	A Comparison of Social Robot to Tablet and Teacher in a New Script Learning Context	NAO robot	Tutee	Script learning	Learning gains

Legend

Particularly relevant for the current research project (see Thesis)			
NAO robot	Tutor	School subjects	Stimulates (self-)reflection of learning processes
Other type of robot	Tutee	Persuasive topics	
Virtual agent	Peer / mix	Other	

Part B (2/2)

5. Intervention (how to reach int. eff.)	6. > 1 session?	7. >1 condition?
Providing information + encouragement	Yes	Yes
Empathy, personalization, feedback + more	Yes	No
Tekst-based dialogue based on learner's affective state	No	No
Behaviors for growth mindset vs neutral mindset	No	Yes
Think-aloud support (remind, prompt, reflect)	No	Yes
Adaptive strategies to counter suboptimal help-seeking behavior	Yes	Yes
Personalized break strategies	No	Yes
Role-switching mechanism	Yes	Yes
Corrective feedback	No	Yes
Evaluative vs nonevaluative robot (differed per proficiency level)	No	Yes
Evaluative vs nonevaluative robot	No	Yes
Tutor vs peer role	No	Yes
Modules including conversations, emotions, advice, reminders	Yes	Yes
Robot explains their decisions	No	Yes
Robot provides feedback and asks questions	No	Yes
Conversing, using gestures and showing emotions	No	Yes
Teaching the robot	No	Yes

Appendix C Personas

Figure 41. Persona - Michelle, large.

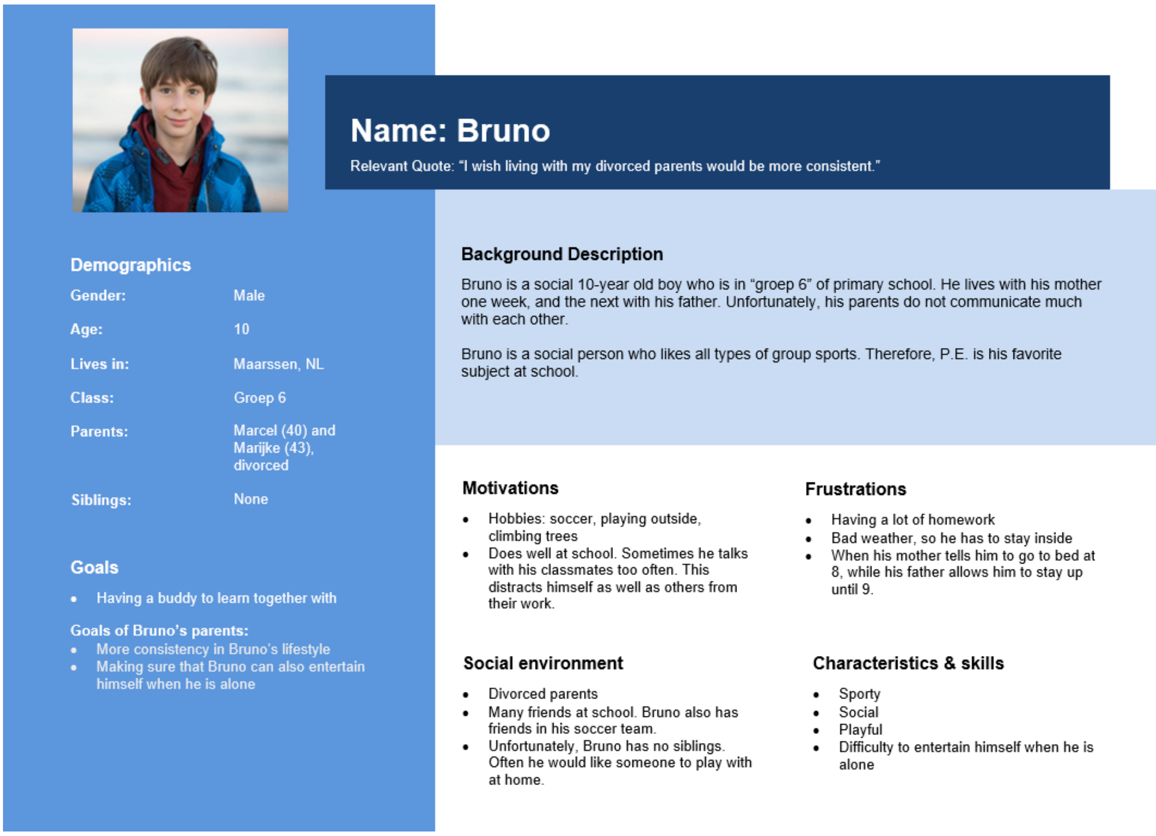


Figure 42. Persona - Bruno, large.

Name: Jamiro

Relevant Quote: "I wish I could play computer games all day and night."

Demographics

Gender:	Male
Age:	11
Lives in:	Hoofddorp, NL
Class:	Groep 7
Parents:	Rennie (39) and Alya (38) Married
Siblings:	Jack (13) and Jo-Ann (9)

Goals

- Being able to fall asleep quickly
- Having less struggles to wake up in the morning

Goals of Jamiro's parents:

- Making Jamiro aware that excessive gaming is unhealthy and that it causes sleeping problems

Background Description

Jamiro is an introverted boy of 11 years who is in "groep 7" of primary school. He lives together with his parents and siblings in Hoofddorp.

Preferably, Jamiro plays videogames all day long. He is primarily interested in futuristic battle games, such as Overwatch. He has an interest in technology and robots.

Motivations

- Hobbies: gaming
- School: Jamiro is highly intelligent, so school assignments take him little effort. He does not have a lot of interest in school.

Frustrations

- Internet connection issues
- His parents complaining about him gaming too much
- Having a hard time going to sleep at night after gaming

Social environment

- Jamiro used to argue with his siblings a lot about who could use the PC at home. Now, he has his own laptop.
- Does not have a lot of friends. He is quite to himself.
- Is not in any sport club but has a handful of online friends.

Characteristics & skills

- Introvert
- Intelligent
- Stubborn
- Is motivated to improve his game skills

Figure 43. Persona - Jamiro, large.

Appendix D Dialogue design for experimental condition

Geel: moet ingevuld worden tijdens experiment

Informele introductie

Robot zwaait

Robot: “Hoi! Goed dat je er bent. Ik ben NAO, hoe heet jij?”

Kind: *naam*

Robot: “Hoi *naam*, wat leuk om je te ontmoeten.

Zoals je ziet, ik ben een robot. Ik kan ook praten, luisteren en bewegen, maar misschien niet zo goed als jij en een beetje anders. Soms reageer ik een beetje langzaam en soms kraakt mijn lichaam als ik me beweeg. Let daar maar niet op. Ik kijk ernaar uit om samen wat te gaan doen. Heb jij wel eens eerder een robot gezien?”

Kind: *ja/nee*

Robot: “En wat vind je van robots?”

Kind: *antwoord*

Robot: “Cool! Ik ben denk ik wel de leukste robot die je hebt gezien. Laten we elkaar nog wat beter leren kennen. Wat is jouw favoriete dier?”

Kind: *dier*

Robot: “Oh wat leuk, waarom is dat jouw favoriete dier?”

Kind: *antwoord*

Robot: “Ga door. Vertel me meer”

Kind: *verder antwoord*

Robot: “Ik ken een *dier* genaamd Buddy. Buddy houdt van aandacht. Ik vind hem erg lief.”

Inhoudelijke introductie

Robot: “Goed, genoeg gekletst. Laten we beginnen met de leeropdrachten. We gaan samen twee video’s kijken en een spel spelen. Ze gaan over gezond eten. Gezond eten is heel belangrijk. Je wordt minder snel ziek, je krijgt minder snel gaatjes in je tanden, je hebt meer energie, kan beter nadenken en je voelt je blijer. Waarom vind jij het belangrijk om gezond te eten?” → **RP 2**

Kind geeft (enigszins) logische motivatie → Robot: “Heel goed!”

Kind zegt dat hij/zij het niet weet of geeft een onlogische reden → Robot: “De meeste andere kinderen vinden het belangrijk dat ze minder snel ziek worden. Dat is een goede reden om te leren over gezond eten.”

Kind of onderzoeker opent leerdoelenpagina op tablet/chromebook

Robot: “Op het scherm zie je wat ik hoop dat je gaat leren vandaag. Kun jij ze aan mij voorlezen?→ **RP 1**

Kind: *leest leerdoelen op*.

Robot: "Heel goed. Telkens als we een leerdoel gehaald hebben, zet je een vinkje in het vakje wat ervoor staat. Zullen we beginnen met de eerste video? Klik maar op *video 1*".

Robot hurkt

Video 1

Video 1 wordt gespeeld

Robot staat op

Robot: "Zo, dat vond ik een leuke video! Ik heb veel geleerd. Jij ook?"

Het kind komt terug op het leerdoelenscherm.

Robot: "Laten we samen kijken naar de leerdoelen van video. → **RP 1** We beginnen bij het eerste leerdoel. Denk je dat je nu weet wat de Schijf van Vijf is? → **RP 3**

Kind: ja → Robot: "Goed om te horen! De Schijf van Vijf is een hulpje van het Voedingscentrum om gezonder en duurzamer te eten en drinken. Vink dit leerdoel maar af."

Kind: nee → Robot: "Dat geeft niet, ik zal je helpen. De Schijf van Vijf is een hulpje van het Voedingscentrum om gezonder en duurzamer te eten en drinken. Vink dit leerdoel maar af."

Robot: "Dan nu het tweede leerdoel. Denk je dat je nu weet wat gezond is om te eten en drinken?"

Kind: ja → Robot: "Goed om te horen! Gezond eten doe je door elke dag genoeg uit elk vak van de Schijf van Vijf te eten. Het gezondste is om elke dag wat anders te eten. Vink dit leerdoel maar af."

Kind: nee → Robot: "Dat geeft niet, ik zal je helpen. Gezond eten doe je door elke dag genoeg uit elk vak van de Schijf van Vijf te eten. Het gezondste is om elke dag wat anders te eten. Vink dit leerdoel maar af."

Robot: "Dan het derde leerdoel. Denk je dat je nu weet wat ongezond is om te eten en drinken?"

Kind: ja → Robot: "Goed om te horen! Snoep en koek zijn ongezond. Frisdrank en sap ook. Daar zit heel veel suiker in. Vink dit leerdoel maar af."

Kind: nee → Robot: "Dat geeft niet, ik zal je helpen. Snoep en koek zijn ongezond. Frisdrank en sap ook. Daar zit heel veel suiker in. Vink dit leerdoel maar af."

Robot: "Super! Zie je dat we al drie leerdoelen gehaald hebben? → **RP 4**. Laten we doorgaan met een spel. Ben je klaar om het spel te starten? Klik maar op *spel.*"

Robot hurkt

Spel

Het spel wordt gespeeld.

Robot staat op

Robot: “Goed gedaan! Jij deed het beter dan dat ik het zou hebben gedaan. Ik heb veel geleerd!”

Het kind komt terug op het leerdoelenscherm.

Robot: “Laten we samen kijken naar de leerdoelen van het spel. → **RP 1** Denk je dat je nu kunt beoordelen of een product gezond of ongezond is?” → **RP 3**

Kind: ja → Robot: “Goed om te horen! Dit doe je door te bedenken wat er in het product zit, bijvoorbeeld veel suiker of vitamines. Vink het leerdoel maar af.”

Kind: nee → Robot: “Dat geeft niet, ik zal je helpen. Dit doe je door te bedenken wat er in het product zit, bijvoorbeeld veel suiker of vitamines. Vink het leerdoel maar af.”

Robot: “Dan nu het tweede leerdoel van het spel. Denk je dat je nu rekening kunt houden met vetten, suikers, zout, vezels en vitamines?”

Kind: ja → Robot: “Goed om te horen! Veel suiker, zout en verzadigd vet zijn meestal ongezond. Vezels en vitamines maken eten juist gezond. Vink dit leerdoel maar af.”

Kind: nee → Robot: “Dat geeft niet, ik zal je helpen. Veel suiker, zout en verzadigd vet zijn meestal ongezond. Vezels en vitamines maken eten juist gezond. Vink dit leerdoel maar af.”

Alle leerdoelen zijn nu afgevinkt.

Robot: “Super! Zie je dat we alle leerdoelen hebben gehaald?” → **RP 4**

Afsluiting

Robot: “We hebben veel geleerd. Hoe vind je dat?” → **RP 5**

Kind reageert positief → Robot: “Wat fijn! Dat vind ik ook.”

Kind reageert niet positief → Robot: “Weet je het zeker? Je hebt veel geleerd. Daar mag je blij mee zijn.”

Robot: “Wat vond je het leukste deel?”

Kind: *antwoord*

Robot: “En wat vond je het minst leuke deel?”

Kind: *antwoord*

Robot: “Aan het begin zei je dat je wilde leren over gezond eten omdat *verwijzing naar motivatie*. Is dat gelukt?” → **RP 2**

Kind: ja → Robot: “Goed om te horen!”

Kind: “nee” of “weet ik niet” → Robot: Dat geeft niet.

Robot: “Dit was het. Bedankt dat je samen met mij wilde leren.”

Robot applaudisseert

Appendix E Dialogue for control condition

Uitleg kleuren:

Grijs: weggelaten in controleconditie

Groen: toegevoegd in controleconditie

Geel: moet ingevuld worden tijdens experiment

Blauw: moet ik nog aanpassen op WoZ-pagina (niet op letten)

Robot zwaait

Robot: “Hoi! Goed dat je er bent. Ik ben NAO, hoe heet jij?”

Kind: *naam*

Robot: “Hoi *naam*, wat leuk om je te ontmoeten.

Zoals je ziet, ik ben een robot. Ik kan ook praten, luisteren en bewegen, maar misschien niet zo goed als jij en een beetje anders. Soms reageer ik een beetje langzaam en soms kraakt mijn lichaam als ik me beweeg. Let daar maar niet op. Ik kijk ernaar uit om samen wat te gaan doen. Heb jij wel eens eerder een robot gezien?”

Kind: *ja/nee*

Robot: “En wat vind je van robots?”

Kind: *antwoord*

Robot: “Cool! Ik ben denk ik wel de leukste robot die je hebt gezien. Laten we elkaar nog wat beter leren kennen. Wat is jouw lievelingsdier?”

Kind: *dier*

Robot: “Oh wat leuk, waarom is dat jouw lievelingsdier?”

Kind: *antwoord*

Robot: “Ga door. Vertel me meer”

Kind: *verder antwoord*

Robot: “Ik ken een *dier* genaamd Buddy. Buddy houdt van aandacht. Ik vind hem erg lief.”

Inhoudelijke introductie

Robot: “Goed, genoeg gekletst. Laten we beginnen met leren over gezond eten. We gaan samen een video kijken en een spel spelen. Ze gaan over gezond eten. Gezond eten is heel belangrijk. Je wordt minder snel ziek, je krijgt minder snel gaatjes in je tanden, je hebt meer energie, kan beter nadenken en je voelt je blijer~~. Waarom vind jij het belangrijk om gezond te eten?” → **RP 2**~~

~~**Happy flow:** Kind geeft (enigszins) logische motivatie → Robot: “Heel goed!”~~

~~**Alternative flow:** Kind zegt dat hij/zij het niet weet of geeft een onlogische reden → Robot: "De meeste andere kinderen vinden het belangrijk dat ze minder snel ziek worden. Dat is een goede reden om te leren over gezond eten."~~

Kind of onderzoeker opent leerdoelenpagina op tablet/chromebook

Robot: "Op het scherm zie je wat ik hoop dat je gaat leren vandaag. Dit noemen we de leerdoelen. ~~Kun jij ze aan mij voorlezen? → **RP 1**~~ Kijk er maar even naar.

~~Kind: *leest leerdoelen op*.~~

Wacht een paar seconden

Robot: "Heel goed. Telkens als we een leerdoel gehaald hebben, zet je een vinkje in het vakje wat ervoor staat. Laten we beginnen met de video. Klik maar op video".

Robot hurkt

<u>Video</u>

De video wordt gespeeld

Robot staat op

Robot: "Zo, dat vond ik een leuke video! Ik heb veel geleerd. Heb jij ook nieuwe dingen geleerd?"

Robot: "Klik maar op het kruisje rechtsboven om de video af te sluiten."

Het kind komt terug op het leerdoelenscherm.

~~Robot: "Laten we samen kijken naar de leerdoelen van video. → **RP 1** We beginnen bij het eerste leerdoel. Denk je dat je nu weet wat de Schijf van Vijf is? → **RP 3**~~

~~**Happy flow:** Kind: ja → Robot: "Goed om te horen! De Schijf van Vijf is een hulpje van het Voedingscentrum om gezonder en duurzamer te eten en drinken. Klik maar op het witte vierkantje om er een vinkje neer te zetten."~~

~~**Alternative flow:** Kind: nee → Robot: "Dat geeft niet, ik zal je helpen. De Schijf van Vijf is een hulpje van het Voedingscentrum om gezonder en duurzamer te eten en drinken. Klik maar op het witte vierkantje om er een vinkje neer te zetten."~~

Robot: "Door de video te kijken hebben we drie leerdoelen gehaald. Je mag daarom bij alle leerdoelen van de video een vinkje zetten. Klik eerst maar op het witte vakje voor *Weten wat de Schijf van Vijf is."*

~~Robot: "Dan nu het tweede leerdoel. Denk je dat je nu weet wat gezond is om te eten en drinken?"~~

~~**Happy flow:** Kind: ja → Robot: "Goed om te horen! Gezond eten doe je door elke dag genoeg uit elk vak van de Schijf van Vijf te eten. Het gezondste is om elke dag wat anders te eten. Zet maar weer een vinkje."~~

~~**Alternative flow:** Kind: nee → Robot: "Dat geeft niet, ik zal je helpen. Gezond eten doe je door elke dag genoeg uit elk vak van de Schijf van Vijf te eten. Het gezondste is om elke dag wat anders te eten. Zet maar weer een vinkje."~~

Robot: “Klik nu op het witte vakje voor *Weten wat gezond is om te eten en drinken”*

~~Robot: “Dan het derde leerdoel. Denk je dat je nu weet wat ongezond is om te eten en drinken?”~~

~~**Happy flow:** Kind: ja → Robot: “Goed om te horen! Snoep en koek zijn ongezond. Frisdrank en sap ook. Daar zit heel veel suiker in. Zet maar een vinkje.”~~

~~**Alternative flow:** Kind: nee → Robot: “Dat geeft niet, ik zal je helpen. Snoep en koek zijn ongezond. Frisdrank en sap ook. Daar zit heel veel suiker in. Zet maar een vinkje.”~~

Robot: “Klik nu op het witte vakje voor *Weten wat ongezond is om te eten en drinken”*

~~Robot: “Super! Zie je dat we al drie leerdoelen gehaald hebben? → **RP 4**~~.

Robot: Laten we doorgaan met een spel. Klik maar op Spel.”

Robot hurkt

Spel

Het spel wordt gespeeld.

Robot staat op

Robot: “Goed gedaan! Jij deed het beter dan dat ik het zou hebben gedaan. Ik heb veel geleerd! Heb jij ook nieuwe dingen geleerd?”

Robot: “Klik maar op het kruisje rechtsboven om het spel af te sluiten.”

Het kind komt terug op het leerdoelenscherm.

~~Robot: “Laten we samen kijken naar het leerdoel van het spel. → **RP 1** Denk je dat je nu kunt beoordelen of een product gezond of ongezond is?” → **RP 3**~~

~~**Happy flow:** Kind: ja → Robot: “Goed om te horen! Dit doe je door te bedenken wat er in het product zit, bijvoorbeeld veel suiker of vitamines. Veel suiker is ongezond en vitamines zijn gezond. Zet er maar een vinkje.”~~

~~**Alternative flow:** Kind: nee → Robot: “Dat geeft niet, ik zal je helpen. Dit doe je door te bedenken wat er in het product zit, bijvoorbeeld veel suiker of vitamines. Veel suikers zijn ongezond en vitamines zijn gezond. Zet er maar een vinkje.”~~

Robot: “Nu we het spel hebben gespeeld, hebben we het laatste leerdoel gehaald. Zet maar een vinkje bij *Kunnen beoordelen of een product gezond of ongezond is.*

Alle leerdoelen zijn nu afgevinkt.

~~Robot: “Super! Zie je dat we alle leerdoelen hebben gehaald?” → **RP 4**~~

Afsluiting

Robot: “We hebben veel geleerd. ~~Hoe vind je dat?” → **RP 5**~~

~~**Happy flow:** Kind reageert positief → Robot: “Wat fijn! Dat vind ik ook.”~~

~~**Alternative flow:** Kind reageert niet positief → Robot: "Weet je het zeker? Je hebt veel geleerd. Daar mag je blij mee zijn."~~

Robot: "Wat vond je het leukste deel?"

Kind: *antwoord*

Robot: "En wat vond je het minst leuke deel?"

Kind: *antwoord*

~~Robot: "Aan het begin zei je dat je wilde leren over gezond eten omdat *verwijzing naar motivatie*. Is dat gelukt?" → **RP 2**~~

~~**Happy flow:** Kind: ja → Robot: "Goed om te horen!"~~

~~**Alternative flow:** Kind: "nee" of "weet ik niet" → Robot: Dat is niet erg."~~

Robot: "Dit was het. Bedankt dat je samen met mij wilde leren."

Robot applaudisseert

Appendix F Information form for participants

Informatie over het onderzoek *Gezond leren eten met een robot*

Waar gaat het over?

We willen je vragen om mee te doen aan een onderzoek naar robots in de klas. We willen graag weten of robots kunnen helpen om beter te leren.

Je mag zelf beslissen of je meedoet.

Wat gaat er gebeuren?

Je gaat op school samen met de robot leren. Jullie zullen samen een video's kijken en een kort spel spelen. De robot zal je af en toe vragen stellen. Na afloop vul je een vragenlijst in.

Wanneer en hoelang?

Je gaat 1 keer met de robot leren. Dit zal op school zijn. Dit duurt ongeveer een kwartier.

Na 1 week vul je weer een vragenlijst in. Dan is er geen robot meer bij.

Wat zijn de voordelen en nadelen?

- \+ Je mag praten met een NAO-robot. Dat vinden de meeste kinderen leuk.
- \+ Je helpt mee aan het verbeteren van de robot. Zo kan hij nog beter worden in het helpen met leren!
- \+ Je leert over gezond eten. Dat is goed voor je gezondheid.
- – Je bent een kwartier weg uit de klas. Misschien mis je een deel van de les.

Belangrijk om te weten:

- Meedoen is **niet verplicht**
- Je mag altijd **stoppen zonder dat je hoeft te vertellen waarom**
- Je mag altijd **vragen stellen**.

Als je vragen hebt

Vragen kun je met je ouders bespreken. Of je kunt ze samen aan de onderzoeker Merel van den Berg stellen.

Je kunt de onderzoeker mailen op: merel.vandenberg@tno.nl

Schrijf jouw vragen hier op:

Toestemming

Zet hieronder je naam neer als je

- ○ Alles hebt gelezen
- ○ Alles begrijpt
- ○ Je mee wilt doen aan het onderzoek

…………………………………………………………………………………..

Appendix G Information form for parents

TNO innovation for life

Informatiebrief ouder - Onderzoek Gezond leren eten met een robot

Wie zijn wij?

Wij zijn onderzoekers van TNO. Dat is een onderzoeksinstelling waar veel onderzoek wordt gedaan.

Wat onderzoeken wij?

Er wordt steeds meer gekeken naar hoe robots kunnen helpen in het onderwijs. Ze kunnen bijvoorbeeld vragen stellen aan de leerling om te controleren of zij de stof goed snappen. Ook kunnen ze de leerling motiveren door ze aan te moedigen. In ons onderzoek kijken we of dit helpt wanneer kinderen leren over gezonde voeding. We kijken daarbij naar het volgende: hoe kan een robot kinderen bewuster maken van de leerdoelen en hun leervoortgang?

Hiervoor hebben wij de hulp van uw kind nodig. Alle leerlingen van groep ...van Harbour Bilingual worden gevraagd om mee te doen aan dit onderzoek. Ze zullen uitgenodigd worden om onder schooltijd een kwartier samen met de robot te leren. Aan het einde van deze sessie meten wij of de gesprekken met de robot goed hebben gewerkt. Met deze uitkomsten willen wij de gesprekstechnieken van de robot verbeteren, zodat zij ingezet kunnen worden om gezondheidsproblemen bij kinderen te voorkomen.

Wat moet uw kind doen? En wat doen wij?

- Wij willen te weten komen of bepaalde gesprekken met een robot goed helpen bij het leren.
- Aan de hand van een sessie van ± 15 minuten met uw kind en de robot verzamelen wij de gegevens. Van deze sessie wordt een geluidsopname gemaakt.
- Aan het einde van de sessie vult het kind een aantal vragen in over de stof en over de robot.

Hoe doet uw kind mee?

U en uw kind mogen zelf weten of jullie meedoen. Als uw kind mee wil doen vragen wij eerst toestemming aan hem/haar. Wij vragen ook om uw toestemming.

Wat gebeurt er met de gegevens als het onderzoek is afgelopen?

- De gegevens worden op het beschermde netwerk van TNO bewaard.
- Uw kind en u kunnen altijd de toestemming intrekken. In dat geval worden alle verzamelde gegevens van uw kind verwijderd.

Waarom zou uw kind meedoen?

- Uw kind krijgt de kans om met een NAO-robot te praten.
- Uw kind leert over gezonde voeding.
- Uw kind draagt bij aan onderzoek om gezondheidsproblemen bij kinderen te voorkomen.

Waarom zou uw kind niet meedoen?

- Uw kind mist ± 15 minuten van de normale lestijd.

Hoe beschermen wij de privacy van uw kind?

- Wij vragen tijdens het gesprek met de robot alleen de voornaam van uw kind. Verder vragen we niet om persoonlijke gegevens.
- De verkregen gegevens blijven geheim voor anderen. Alleen de onderzoeker van TNO heeft toegang.
- Wij delen de gegevens niet met anderen.
- Alle verkregen gegevens van uw kind krijgen een code die niets betekent (bijv. C1). Zo kan niemand raden van wie de code is.
- Alleen de onderzoeker kan zien welke naam bij welke code hoort. Zij maakt de lijst met codes en het wachtwoord.
- De gegevens worden vernietigd na 10 jaar.
-
- Onderzoekers moeten zich houden aan regels en afspraken om jouw privacy te beschermen. U kunt hierover meer lezen op: tno.nl/nl/over-tno/contact/corporate-legal/privacystatement/privacystatement-voor-tno-onderzoek

Contact: Merel van den Berg merel.vandenberg@tno.nl

Appendix H Consent form for parents

Toestemmingsverklaring

Informed consent form

Ondergetekende partijen geven toestemming voor deelname van zijn/haar kind aan het onderzoek getiteld *Gezond leren eten met een robot* bij TNO.

The undersigned parties consent to his/her child's participation in the study entitled Learning to eat healthily with a robot *at TNO.*

....

Naam kind ....

Name child

Geboortedatum ....

Date of birth

Ik bevestig dat ik de informatie over bovengenoemd onderzoek heb gelezen.

Ik begrijp de informatie.

I confirm that I have read the information about the study mentioned above. I understand the information.

De bedoelingen van het onderzoek en de daarbij gevolgde aanpak zijn tot mijn tevredenheid uitgelegd.

The intentions of the study and the approach followed have been explained to my satisfaction.

Ik heb de gelegenheid gehad om aanvullende vragen te stellen en deze vragen zijn naar tevredenheid beantwoord.

I have had the opportunity to ask additional questions and these questions have been answered satisfactorily.

Ik heb voldoende tijd gehad om over deelname van mijn kind na te denken.

I have had sufficient time to think about participation of my child.

Ik weet dat de deelname van mijn kind aan het onderzoek geheel vrijwillig is en dat ik de deelname op ieder moment kan beëindigen zonder dat ik daarvoor een reden hoef op te geven.

I know that the participation of my child in the study is completely voluntary and that I can withdraw my consent at any time without providing a reason.

Ik geef toestemming voor de verwerking van de persoonsgegevens van mijn kind voor de doelen zoals beschreven in de informatie.

I give permission to process the personal data of my child for the purposes described in the information.

Ik geef toestemming de onderzoeksgegevens van mijn kind te hergebruiken voor toekomstig onderzoek op het beschreven onderzoeksgebied op voorwaarde dat deze zo gecodeerd zijn, dat ze niet meer naar mijn kind terug te leiden zijn.

I give permission to reuse the research data of my child for future research in the research area described, provided that these are coded in such a way that they can no longer be traced back to my child.

Ik geef toestemming voor het bewaren van de gegevens van mijn kind en dat bevoegde leden van het onderzoeksteam en bevoegde inspecteurs hier inzage in hebben.

I give permission for the data of my child to be stored and for authorized members of the study team and authorized inspectors to have access to the data.

Voorts verklaar ik dat mijn kind geen mij bekende belemmeringen heeft om aan het onderzoek deel te nemen.

Furthermore, I declare that my child has no known impediments to participate in the study.

- ☐ **Optioneel (aankruisen):** Ik geef toestemming om een onherkenbare foto van mijn kind tijdens de sessie met de robot te gebruiken
  Optional (tick): *I give permission to use an unrecognizable picture of my child during the session with the robot*
- ☐ **Optioneel (aankruisen):** Ik geef toestemming om een herkenbare foto van mijn kind tijdens de sessie met de robot te gebruiken
  Optional (tick): *I give permission to use a recognizable picture of my child during the session with the robot*

Het toestemmingsformulier moet worden getekend door een gezaghebbende ouder of door de voogd, dus weghalen wat niet van toepassing is

The informed consent form must be signed by an authoritative parent or by the guardian, so remove what does not apply

Naam, handtekening en datum ondertekening gezaghebbende ouder/voogd

Name, signature and date of signing of authoritative parent/guardian

……………………………………………………………………………………………………

Ik heb me ervan vergewist dat ik de ouders / de voogd goed geïnformeerd heb over het onderzoek waaraan het kind gaat deelnemen.
I have made sure that I have properly informed the parents/guardian about the study in which the child is going to participate.

Naam, handtekening en datum ondertekening proefleider:

Name, signature and date of signing study leader

……………………………………………………………………………………………………

Appendix I Code of WoZ page

```
<html>
<head>
 <script src="https://ajax.googleapis.com/ajax/libs/jquery/3.3.1/jquery.min.js"></script>
 <script src="https://cdn.jsdelivr.net/npm/autobahn-browser@20.9.2/autobahn.min.js"></script>
</head>
<body>
 <div>
  <input id="realmInput">
  <button id="connectBtn">Verbind</button>
 </div>

 <div>
  <select id="behaviorSelector">
   <option value="BlocklyWaveRightArm">Zwaaien</option>
    <option value="BlocklyCrouch">Hurken</option>
    <option value="BlocklyStand">Opstaan</option>
    <option value="BlocklyLookingUp">Omhoog kijken</option>
    <option value="BlocklyPride">Handkusje</option>
    <option value="BlocklySneeze">Niezen</option>
    <option value="BlocklyApplause">Applaus</option>
    <option value="BlocklyTaiChiChuan">Tai chi</option>
  </select>
  <button id="playBtn">Behavior play</button>
 </div>

 <div><input id="robotSay39" placeholder="Tekst" size = "100"><span>Improvisatie</span></div>
 <div><input id="robotSay" placeholder="Tekst" size = "100" value = "Hoi! Goed dat je er bent. Ik ben Na-o, hoe heet jij?"></div>
 <div><input id="robotSay2" placeholder="Tekst" size = "100" value = "Hoi NAAM, wat leuk om je te ontmoeten. Zoals je ziet, ik ben een robot. Ik kan ook praten, luisteren en bewegen, maar misschien niet zo goed als jij en een beetje anders. Soms reageer ik een beetje langzaam en soms kraakt mijn lichaam als ik beweeg. Let daar maar niet op. Ik kijk ernaar uit om samen wat te gaan doen. Heb jij al wel eens eerder een robot gezien?"><span>AANPASSEN!</span></div>
 <div><input id="robotSay3" placeholder="Tekst" size = "100" value = "Cool. Ik ben denk ik wel de leukste robot die je hebt gezien. Laten we elkaar nog wat beter leren kennen. Wat is jouw lievelingsdier?"></div>
  <div><input id="robotSay4" placeholder="Tekst" size = "100" value = "Oh wat leuk. Waarom is dat jouw lievelingsdier?"></div>
  <div><input id="robotSay5" placeholder="Tekst" size = "100" value = "Ga door. Vertel me meer."></div>
  <div><input id="robotSay6" placeholder="Tekst" size = "100" value = "Ik ken een DIER genaamd Buddy. Buddy houdt erg van aandacht. Ik vind hem erg lief.  Goed, genoeg gekletst. Laten we beginnen met leren over gezond eten. Gezond eten is heel belangrijk. Je wordt minder snel ziek, je krijgt minder gaatjes in je tanden, je hebt meer energie, je kunt beter nadenken en je voelt je blijer. Op het scherm zie je wat ik hoop dat je gaat leren vandaag. Dit noemen we de leerdoelen. Kijk er maar even naar" ><span>AANPASSEN!</span></div>
  <div><input id="robotSay10" placeholder="Tekst" size = "100" value = "Heel goed. Telkens als we een leerdoel gehaald hebben, zet je een vinkje in het vakje wat ervoor staat. Laten we beginnen met de video. Klik maar op video."></div>
  <div><input id="robotSay11" placeholder="Tekst" size = "100" value = "Zo, dat vond ik een leuke video! Ik heb veel geleerd. Heb jij ook nieuwe dingen geleerd?"><span>Na video</span></div>
  <div><input id="robotSay110" placeholder="Tekst" size = "100" value = "Klik maar op het kruisje rechtsboven  om de video af te sluiten."></div>
  <div><input id="robotSay12" placeholder="Tekst" size = "100" value = "Door de video te kijken hebben we drie leerdoelen gehaald. Je mag daarom bij alle leerdoelen van de video een vinkje zetten. Klik eerst maar op het witte vakje voor Weten wat de Schijf van Vijf is."></div>
  <div><input id="robotSay15" placeholder="Tekst" size = "100" value = "Klik nu op het witte vakje voor Weten wat gezond is om te eten en drinken"></div>
  <div><input id="robotSay18" placeholder="Tekst" size = "100" value = "Klik nu op het witte vakje voor Weten wat ongezond is om te eten en drinken"></div>
  <div><input id="robotSay21" placeholder="Tekst" size = "100" value = "Laten we doorgaan met een spel. Klik maar op spel"></div>
  <div><input id="robotSay22" placeholder="Tekst" size = "100" value = "Goed gedaan! Jij deed het beter dan dat ik het zou hebben gedaan. Ik heb veel geleerd! Heb jij ook nieuwe dingen geleerd?"><span>Na spel </span></div>
  <div><input id="robotSay220" placeholder="Tekst" size = "100" value = "Klik maar op het kruisje rechtsboven om het spel af te sluiten."></div>
  <div><input id="robotSay23" placeholder="Tekst" size = "100" value = "Nu we het spel hebben gespeeld, hebben we het laatste leerdoel gehaald. Zet maar een vinkje bij Kunnen beoordelen of een product gezond of ongezond is."></div>
  <div><input id="robotSay30" placeholder="Tekst" size = "100" value = "We hebben veel geleerd. Wat vond je het leukste deel?"></div>
  <div><input id="robotSay34" placeholder="Tekst" size = "100" value = "En wat vond je het minst leuke deel?"></div>
  <div><input id="robotSay38" placeholder="Tekst" size = "100" value = "Dit was het. Bedankt dat je samen met mij wilde leren."></div>
```

```
</body>
<script>

var session = null // The session is basically our bridge between our code and the robot

////////////////////////////////////////////////////////////////////////////////
// This part wil connect our code with the robot                              //
////////////////////////////////////////////////////////////////////////////////

async function main(sess) {
  // This function will be called as soon as we have a successful connection
  // between our code and the realm of the robot.
  console.log("Robot is online!")
  session = sess // Set the global variable session

 // Dit is toegevoegd!
   // await session.call("rom.optional.behavior.play", [],
          //  {"name":"BlocklySitDown"})

}

function connectToRealm(realm) {
  console.log("Connecting to robot with realm: " + realm)

  var wamp = new autobahn.Connection({
                url: "wss://wamp.robotsindeklas.nl",
                realm: realm,
                protocols: ["wamp.2.msgpack"]
  })
  wamp.onopen = main
  wamp.open()
}

$("#connectBtn").click(function(event) {
  // This function will be called when somebody clicks on the connectBtn button.

  // The realm of a robot can be described as a small world in our cloud environment
  // that we should connect to, to control the robot. The code below, collects the
  // input value from the realmInput field and tries to connect to it.
  let realm = $("#realmInput").val()
  connectToRealm(realm)
})

////////////////////////////////////////////////////////////////////////////////
// Code required to let the robot say something with an input box             //
////////////////////////////////////////////////////////////////////////////////

async function say(text) {
  if (session == null) {
    console.log("Code has not yet been connected to a robot")
    return
  }
  console.log("The robot will say: " + text)
  // await session.call("rie.dialogue.say", [], {"text": text})
  // If you want animated speech, comment the line above and uncomment the line below
  await session.call("rie.dialogue.config.native_voice", [], {"use_native_voice": false} )
  await session.call("rie.dialogue.say_animated", [], {"text": text})
}

$("#robotSay").keypress(function(event) {
  if (event.keyCode == 13) {
    let text = $("#robotSay").val()
    say(text)
  }
})
```

Some repetitive code similar to the keypress function above is left out here

```
////////////////////////////////////////////////////////////////////////////////
// Code required to play a behavior on the robot via an select              //
////////////////////////////////////////////////////////////////////////////////

async function playBehavior(behavior) {
  if (session == null) {
    console.log("Code has not yet been connected to a robot")
    return
  }
  console.log("The robot will do: " + behavior)
  // First let the robot stand up:
  // await session.call("rom.optional.behavior.play", [], {"name": "BlocklyStand"})
  // Play the selected behavior:
  await session.call("rom.optional.behavior.play", [], {"name": behavior})
  // Let the robot go back to its resting position
  // await session.call("rom.optional.behavior.play", [], {"name": "BlocklyStand"})

}

$("#playBtn").click(function(event) {
  let behavior = $("#behaviorSelector").val()
  playBehavior(behavior)
})

</script>
</html>
```

Appendix J Questionnaire 1

Figure 44. Questionnaire 1 - page 1.

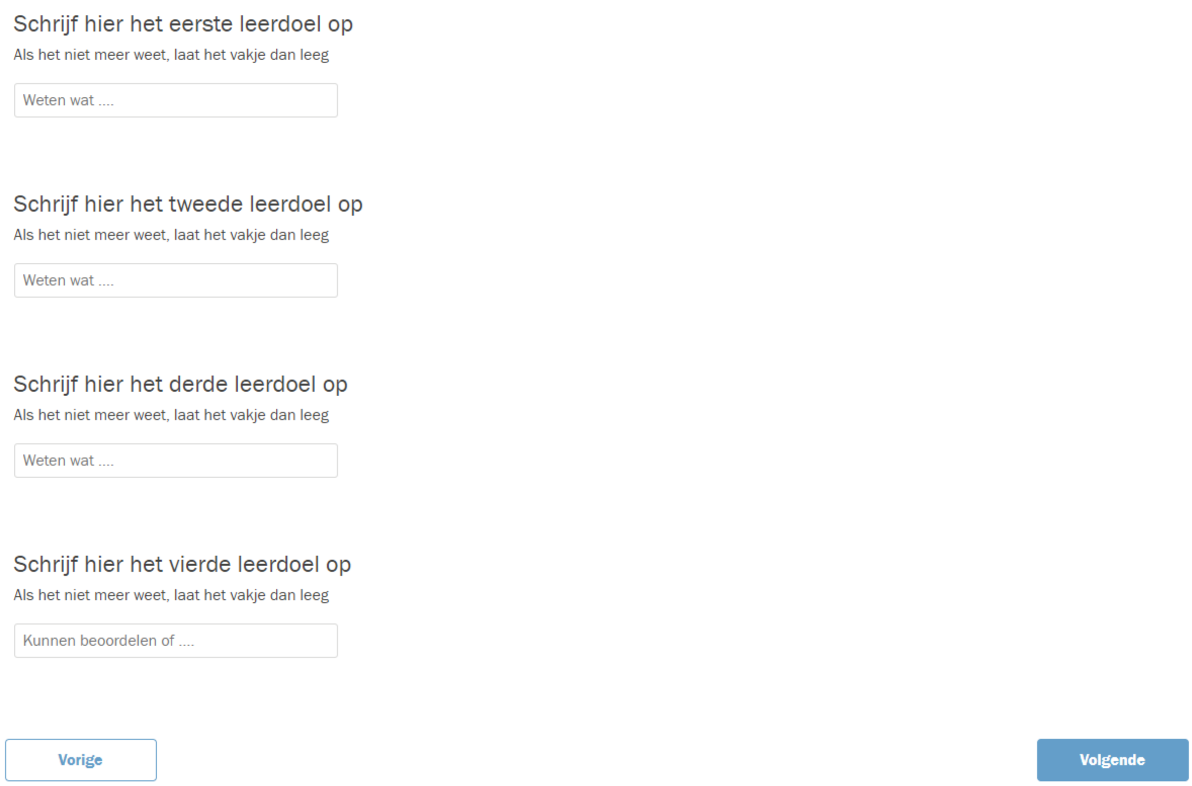


Figure 45. Questionnaire 1 - page 2.

1. Waar zijn voedingsstoffen goed voor?

Er zijn **2** goede antwoorden

- [ ] Dunner worden
- [ ] Fit blijven
- [ ] Groeien
- [ ] Minder gaatjes in je tanden
- [ ] Slimmer worden

2. Wat staat er **<u>wel</u>** in de Schijf van Vijf?

Er zijn **3** goede antwoorden

- [ ] Aardappels
- [ ] Cola
- [ ] Koekjes
- [ ] Thee
- [ ] Vlees

Figure 46. Questionnaire 1 - page 3.

3. Wat is het kleinste vak in de Schijf van Vijf?

- ○ Gezonde vetten zoals olie, margarine en halvarine
- ○ Vis, vlees, noten, peulvruchten, ei en zuivel
- ○ Volkorenbrood, volkorenpasta, zilvervliesrijst en aardappels
- ○ Water en thee zonder suiker

4. Waarom zijn frisdrank en sap ongezond?

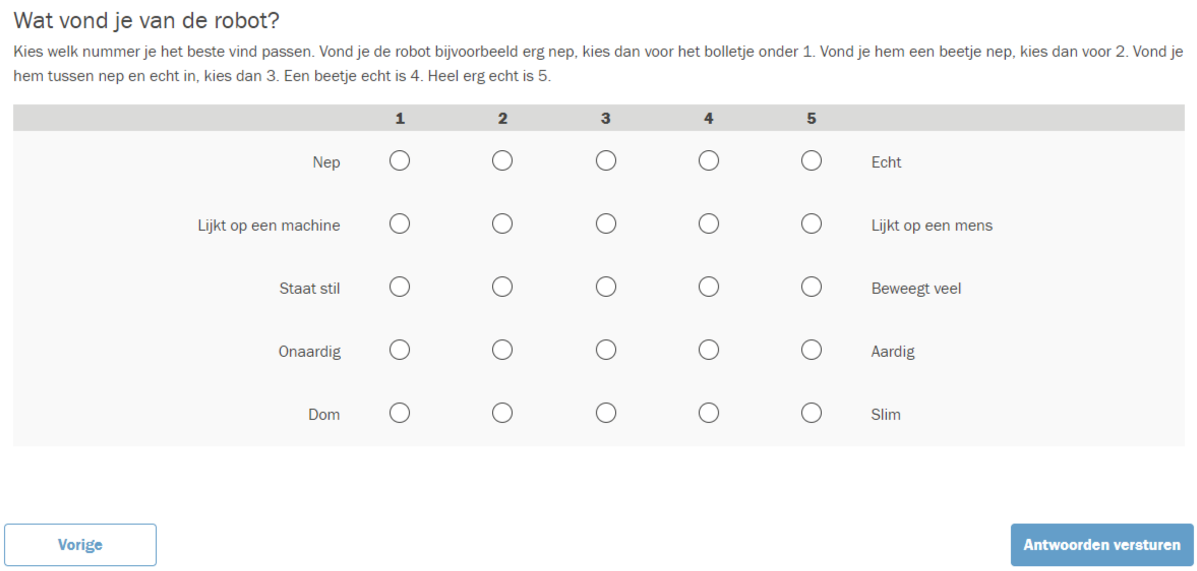
5. Waarom is chocoladepasta ongezond?

Noem **2** redenen

6. Welk zoet beleg is gezonder dan chocoladepasta?

Vorige | Volgende

Figure 47. Questionnaire 1 - page 4.



Wat vond je van de robot?

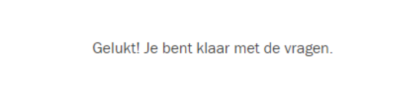
Kies welk nummer je het beste vind passen. Vond je de robot bijvoorbeeld erg nep, kies dan voor het bolletje onder 1. Vond je hem een beetje nep, kies dan voor 2. Vond je hem tussen nep en echt in, kies dan 3. Een beetje echt is 4. Heel erg echt is 5.

	1	2	3	4	5	
Nep	○	○	○	○	○	Echt
Lijkt op een machine	○	○	○	○	○	Lijkt op een mens
Staat stil	○	○	○	○	○	Beweegt veel
Onaardig	○	○	○	○	○	Aardig
Dom	○	○	○	○	○	Slim

Vorige

Antwoorden versturen

Figure 48. Questionnaire 1 - page 5.



Gelukt! Je bent klaar met de vragen.

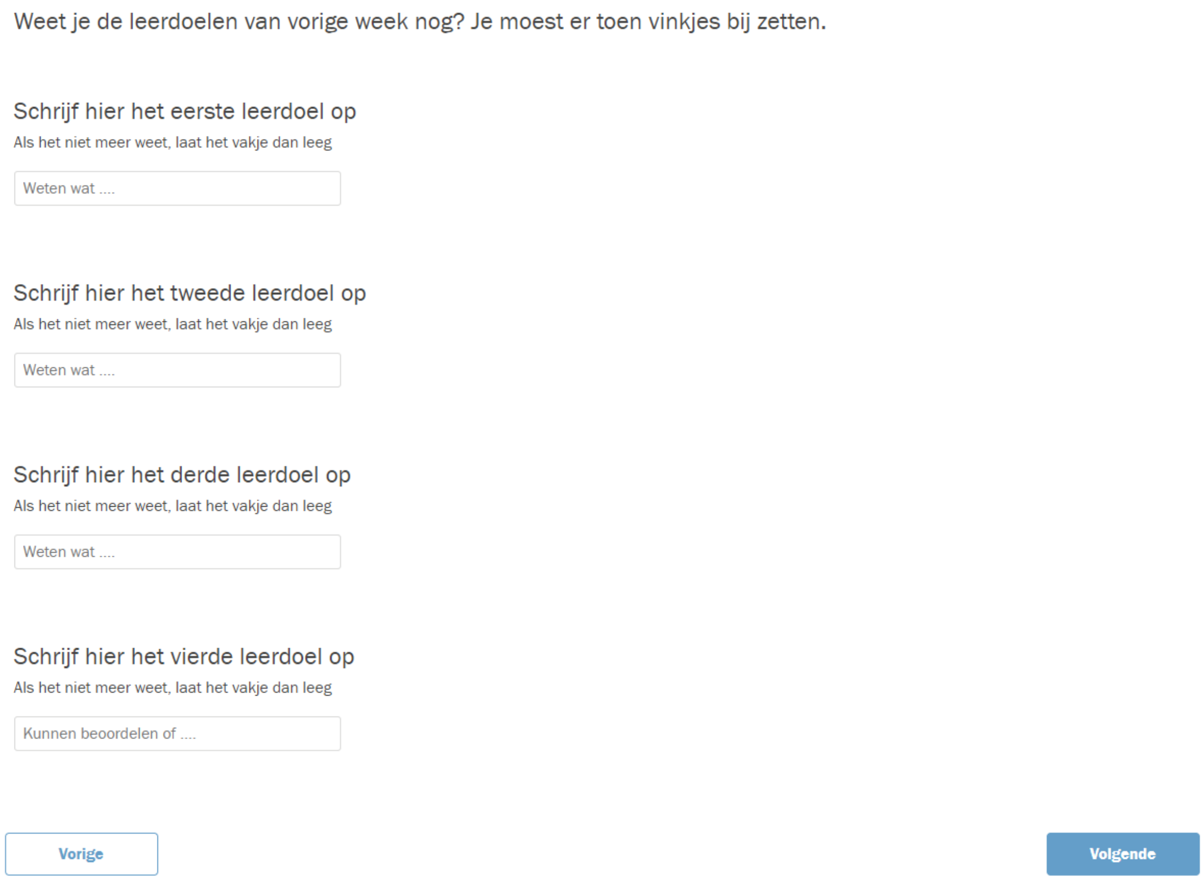
Figure 49. Questionnaire 1 - page 6.

Appendix K Questionnaire 2


De juf vult hier een code voor je in

Volgende

Figure 50. Questionnaire 2 - page 1.


Weet je de leerdoelen van vorige week nog? Je moest er toen vinkjes bij zetten.

Schrijf hier het eerste leerdoel op

Als het niet meer weet, laat het vakje dan leeg

Weten wat ....

Schrijf hier het tweede leerdoel op

Als het niet meer weet, laat het vakje dan leeg

Weten wat ....

Schrijf hier het derde leerdoel op

Als het niet meer weet, laat het vakje dan leeg

Weten wat ....

Schrijf hier het vierde leerdoel op

Als het niet meer weet, laat het vakje dan leeg

Kunnen beoordelen of ....

Vorige

Volgende

Figure 51. Questionnaire 2 - page 2.

1. Waar zijn voedingsstoffen goed voor?
Er zijn **2** goede antwoorden

- ☐ Dunner worden
- ☐ Fit blijven
- ☐ Groeien
- ☐ Minder gaatjes in je tanden
- ☐ Slimmer worden

2. Wat staat er **wel** in de Schijf van Vijf?
Er zijn **3** goede antwoorden

- ☐ Aardappels
- ☐ Cola
- ☐ Koekjes
- ☐ Thee
- ☐ Vlees

Figure 52. Questionnaire 2 - page 3.

3. Wat is het kleinste vak in de Schijf van Vijf?

- ○ Gezonde vetten zoals olie, margarine en halvarine
- ○ Vis, vlees, noten, peulvruchten, ei en zuivel
- ○ Volkorenbrood, volkorenpasta, zilvervliesrijst en aardappels
- ○ Water en thee zonder suiker

4. Waarom zijn frisdrank en sap ongezond?

5. Waarom is chocoladepasta ongezond?
Noem **2** redenen

6. Welk zoet beleg is gezonder dan chocoladepasta?

Vorige | Volgende

Figure 53. Questionnaire 2 - page 4.

Gelukt! Je bent klaar met de vragen.

Figure 54. Questionnaire 2 - page 5.

Appendix L Pilot studies

Pilot with TNO employees

Two pilot studies with TNO employees were conducted to test the procedure of the experiment. The first pilot study resulted in a number of observations that needed change:

1. The robot's voice changed throughout the conversation. The developer of Interactive Robotics was contacted to resolve this issue. In the end, an extra line was added to the WoZ code.
2. The experiment took approximately 25 minutes. This was 10 minutes longer than anticipated. The second video that was initially added to the learning material was therefore removed.
3. In this initial set-up, there was one blank text box on the WoZ page and the researcher copy-pasted all dialogue lines into this. This resulted in mistakes and delays in the conversation. Therefore, more text boxes were added that were prefilled with the dialogue sentences.
4. The participant indicated that one of the learning goals was difficult to understand (especially for children). The wording was changed.
5. The video was first incorporated as a link to the YouTube website. This resulted in a commercial being played. The video was therefore downloaded and linked as a local file.
6. The participant indicated that it should be clearer that the robot actively watched the video and game. This resulted in some try-outs for movements shortly before and after the video or game was played.
7. The participant indicated that it would be good if the robot indicated that they had watched the video or game too. Therefore, this statement was added to the dialogue twice: "I have learnt a lot. Have you also learnt new things?".
8. The participant proposed to add a question that measured the emotional state of the participant before and after the session. The AffectButton seemed appropriate for this. This was tested in the following pilot studies.

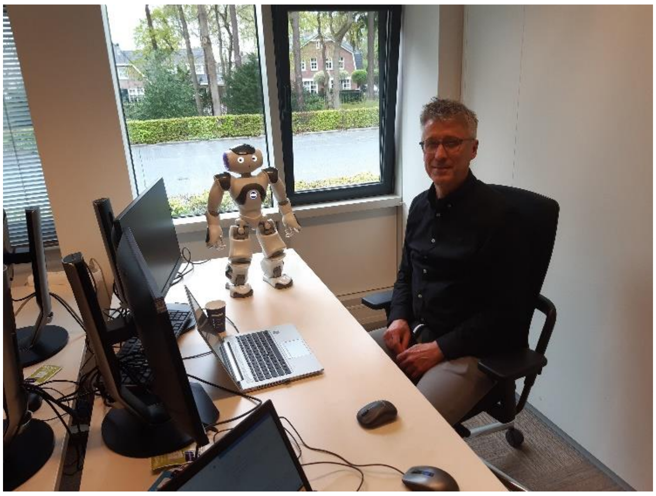


Figure 55. Pilot study 1.

During the second pilot, the following observations were made:

1. Now that all dialogue text was prefilled, the response time of the robot decreased. However, some mistakes in pasting the text in the code became apparent. These were corrected.

2. It was not clear for the participant where they had to click for the video and game. An extra sentence was added where the robot would say: “Click on video/game”.
3. One of the learning goals was removed, because the participant found it too difficult to understand (again, especially for children).
4. The audio of the explanation of the game was not played automatically. Apparently, this was a browser setting that had to be changed.

Figure 56. Pilot study 2.

Pilot with children

The next pilot study was conducted with four children in the right age category. They were all pupils in grade 4. Two of them tested the experimental condition and two of them tested the control condition. The pilot was meant to test both the procedure and the quality of the questionnaire questions.

Figure 57. Pilot study 3.

The following observations were made:

1. In general, the procedure went smoothly. The participants performed all actions correctly and indicated that the robot listened well to them. Sometimes they asked where to click.

2. An attempt was made to have the robot move subtlety during the video and game to raise awareness that the robot was watching the screen too, but this only seemed to distract participants.
3. The questionnaire could not be filled in twice on the same laptop. If cookies were removed and the browser was refreshed, it worked again. This was added to the procedure.
4. The incorporation of the AffectButton was tested. After a brief explanation, the participants seemed to understand how to use it. However, it took considerable time and the sessions often took 20 minutes already. Additionally, a simple way to record its outcome was not found. Combined with the fact that this would not measure a primary outcome variable, it was decided that the AffectButton was left out of the final experiment design.
5. One participant struggled to recall any of the learning goals. Text was added to the page that the fields could be left empty if the participant did not recall.
6. Participants struggled with the prompt "write down as many learning goals as you can", followed by an empty text box. Their answers indicated that they did not understand what was asked. Therefore, the text box was split into four boxes, one for each learning goal. In addition, the start of the phrases were prefilled in the field, i.e. "Weten wat…" and "Kunnen beoordelen of".
7. A number of the questions about the content were answered correctly by all participants. This indicated that they might have been too easy. Therefore, the following changes were implemented:
    a. An extra answer option was added to Q1 and Q3.
    b. Two questions were removed.
    c. Q6 was changed from a closed question to an open question.

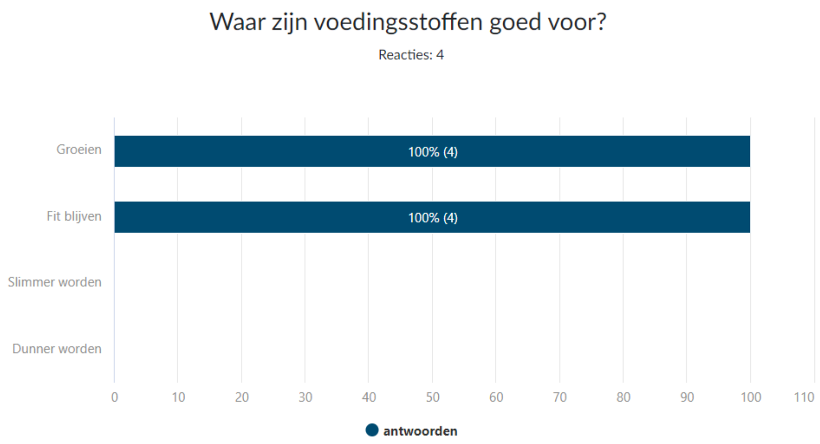


Figure 58. Pilot results - Q1.

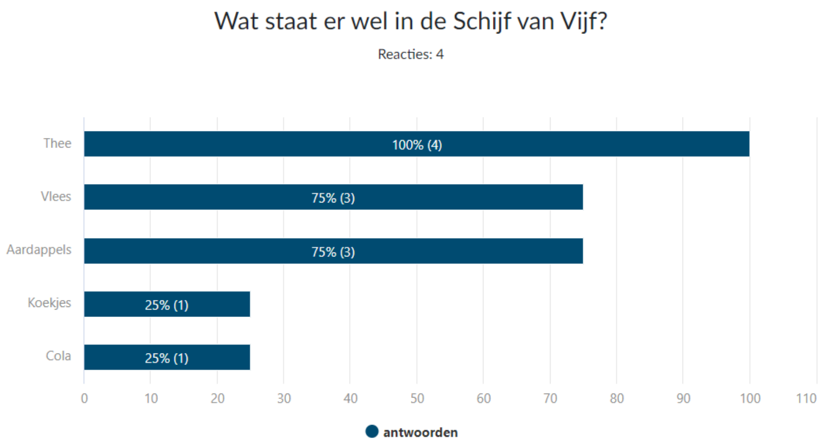


Figure 59. Pilot results - Q2.

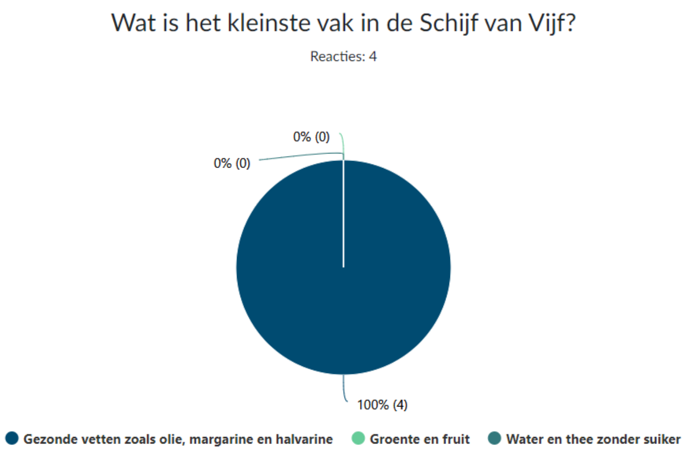


Figure 60. Pilot results - Q3.

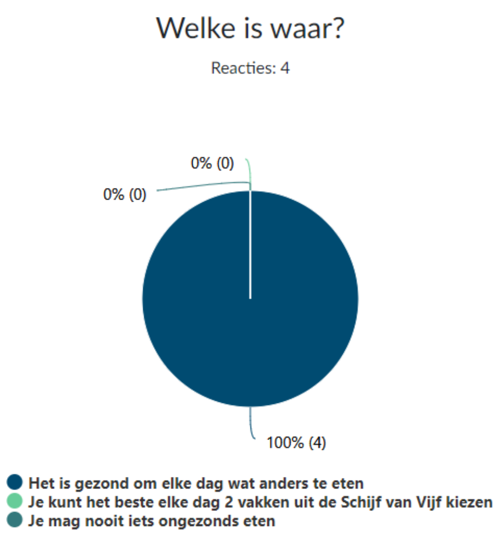


Figure 61. Pilot results - Q4.

Waarom zijn frisdrank en sap ongezond?

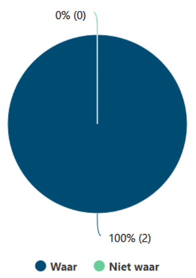
Reacties: 4

waarden
omdat er veel suiker in zit
omdat daar veel suiker in zit
sdvc
daar zit suiker in

Figure 62. Pilot results - Q5.

Een gezond ontbijt helpt je om na te denken

Reacties: 2 / Ontbrekende: 2



Antwoordoptie	absoluut	in%
Waar	2	100%
Niet waar	0	0%
Gemiddelde	**1**	
Aantal antwoorden	**2**	
Aantal NVT	**0**	
Aantal missings	**2**	

Figure 63. Pilot results – Q6.

Waarom is chocoladepasta ongezond?

Reacties: 3 / Ontbrekende: 1

waarden
chocola in zit en suiker
omdat er veel calorien in chocoladepasta zitten. en omdat er veel vet in zit
omdat daar vet in

Figure 64. Pilot results - Q7.

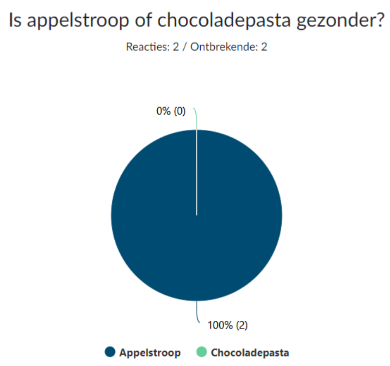


Figure 65. Pilot results - Q8.

Appendix M Draaiboek voor Gezond leren eten met een robot

Set-up

Praktische details
De school is Harbour Bilingual op de Coolhavenstraat 29A in Rotterdam. Vanaf Rotterdam Centraal zijn er veel verschillende opties.Tramlijn 8 richting Spangen is erg leuk, dan kom je langs de Euromast :).

De voordeur van de school zit standaard dicht. Je kunt aanbellen, maar soms wordt hij al voor je open gemaakt door iemand die voorbij loopt. Als ze vragen waar je voor komt, zeg dan dat je op het kantoor op zolder een onderzoek afneemt bij groep 5a. De beste manier om hier te komen is om direct naar links te gaan na de ingang, dan 1 set trappen omhoog te gaan, daar de gang over te lopen (daar zitten allerlei klassen), en dan bij het volgende traphuis omhoog te gaan. Als je dan door de deur gaat is het kantoor links van je. Je kunt me even bellen/appen als je het niet kunt vinden, maar misschien ben ik dan even bezig met een participant.

We gaan op ma t/m wo het experiment afnemen bij groep 6b. Hier zitten 28 kinderen in. De juf heet *naam juf*.
We gaan op do t/m vr het experiment afnemen bij groep 5a. Hier zitten 23 kinderen in. De juf heet *naam juf*.

De school heeft een eigen NAO die we als back-up kunnen gebruiken. Die staat nu ook op het kantoor.

Het wifi-wachtwoord is ******* (naam: Harbour Bilingual).

Benodigdheden
Meenemen Merel (onderzoeker)

- Eigen laptop TNO + oplader
- Andere laptop TNO + oplader
- Eigen laptop + oplader?
- NAO (incl ethernetkabel, accu, shirt)
- Stekkerblok
- Pen + papier
- Eigen drinken en evt eten (er zit een AH op 7 min loopafstand)
- Eten voor kinderen
- Bedankje voor klas (op woensdag + vrijdag)

Meenemen onderzoeksassistent

- Eigen laptop + oplader
- Pen + papier
- Eigen drinken en evt eten (er zit een AH op 7 min loopafstand)

Procedure

Rolverdeling
Let op: dit is een globale rolverdeling. Soms is het makkelijk om even in te springen en elkaars taak over te nemen.

Merel (onderzoeker)

- NAO besturen
- Participantnummer uitdelen en lijst bijhouden (op papier)
- Materiaal klaarzetten (leerdoelenlijst, survalyzer)
- Audio-opname starten en stoppen

Onderzoeksassistent

- Kinderen ophalen en terugbrengen
- Foto's maken (alleen bij toestemming!)
- Belangrijke observaties noteren (bijv. "halverwege de sessie met participant C1 is NAO gevallen")

Planning

Maandag	10.00-11.00	Aankomst Merel. Consent forms ophalen bij 6b. Ruimte klaarmaken. NAO verbinden.
	10.30-11.00	Aankomst onderzoeksassistent. Overzicht maken van consent forms: welke kinderen mogen meedoen en van welke kinderen mogen er foto's gemaakt worden. Helpen met klaarzetten.
	11.00-11.45	3 participanten. Start met controleconditie.
	11.45-13.15	Pauze (klas heeft gym + pauze)
	13.15-15.00	Ongeveer 7 participanten
	15.00-15.15	Opruimen en naar huis (vragen wat de juf als bedankje wil?)
Dinsdag	10.30-11.00	Aankomst Merel en onderzoeksassistent. Klaarzetten.
	11.00-11.45	3 participanten
	11.45-13.15	Pauze (klas heeft gym + pauze)
	13.15-15.00	Ongeveer 7 participanten

	15.00-15.15	Opruimen en naar huis
Woensdag	10.30-11.00	Aankomst Merel en onderzoeksassistent. Klaarzetten. Consent forms ophalen bij 5a (Merel maakt thuis alvast een overzicht).
	11.00-12.30	Ongeveer 8 participanten. Let op: als het goed is zijn dit vooral kinderen waarbij geen toestemming is gegeven. Die hoeven geen toets af te nemen en zijn dus sneller klaar. Mocht er tijd tekort zijn, dan meerdere kinderen tegelijk.
	12:30	Afscheid van klas. Bedankje overhandigen (?). Afstemmen met juf dat er over een week een toets moet worden afgenomen.
	12.30-12.45	Opruimen en naar huis.
Donderdag	10.30-11.00	Aankomst Merel. Klaarzetten.
	11.00-12:30	Ongeveer 4 participanten
	12:30-13.15	Pauze (denk ik, misschien heeft deze klas tussen 12.00 en 12.45 pauze)
	13.15-15.00	Ongeveer 5 participanten
	15.00-15.15	Opruimen en naar huis
Vrijdag	10.30-11.00	Aankomst Merel en onderzoeksassistent. Klaarzetten.
	11.00-12:30	Ongeveer 4 participanten
	12:30-13.15	Pauze (denk ik, misschien heeft deze klas tussen 12.00 en 12.45 pauze)
	13.15-14.50	Overige participanten (onderzoeksassistent eerder weg). We gaan het zeer waarschijnlijk niet redden om alle participanten 1-voor-1 te laten deelnemen. We kunnen op het laatst 1 of 2 groepjes van 2-4 kinderen tegelijkertijd laten deelnemen.
	14.50	Afscheid van klas. Traktatie en bedankje overhandigen. Afstemmen met juf dat er over een week een toets moet worden afgenomen.

	15.00-15.30	Opruimen en naar huis (onderzoeksassistent eerder weg)

Stappenplan per kind

1. Onderzoeksassistent haalt kind uit de klas nadat het vorige kind is teruggebracht. Merel zet ondertussen het materiaal klaar en start een nieuwe audio-opname. Onderzoeksassistent maakt alvast een gesprekje met het kind (bijv vragen of ze er zin in hebben).
2. Kind arriveert. Onderzoeksassistent begeleidt hem/haar naar de stoel. Onderzoeksassistent controleert of er consent is gegeven en of er foto's gemaakt mogen worden. Merel schrijft naam en participantnummer op.
3. De sessie met NAO start. In principe hoeft onderzoeksassistent hier niet veel te doen. Ze mogen vragen beantwoorden. Zij schrijven ook op als er iets bijzonders gebeurt (bijvoorbeeld NAO valt). Ze maken een foto als de ouders daarvoor toestemming hebben gegeven (het liefst waarbij je duidelijk ziet dat het kind en robot in gesprek zijn).
4. Zodra het gesprek met de NAO eindigt, zorgt Merel ervoor dat het kind naar de survey gaat en deze invult. We helpen bij vragen, maar geven geen inhoudelijke antwoorden. Als een kind moeite heeft met lezen, lees het dan voor. Als ze moeite hebben met typen, schrijf dan hun antwoorden op de open vragen op.
5. Zodra het kind klaar is, bedanken we hem/haar en vraagt Merel wat voor lekkers ze op vrijdag willen. Onderzoeksassistent loopt terug met ze naar de klas. Onderweg vragen ze aan het kind hoe ze het vonden en of ze niet door te vertellen aan andere kinderen wat ze precies hebben gedaan (want "dan is het niet zo leuk meer voor de andere kinderen"). We beginnen daarna weer bij stap 1.

Appendix N Additional figures

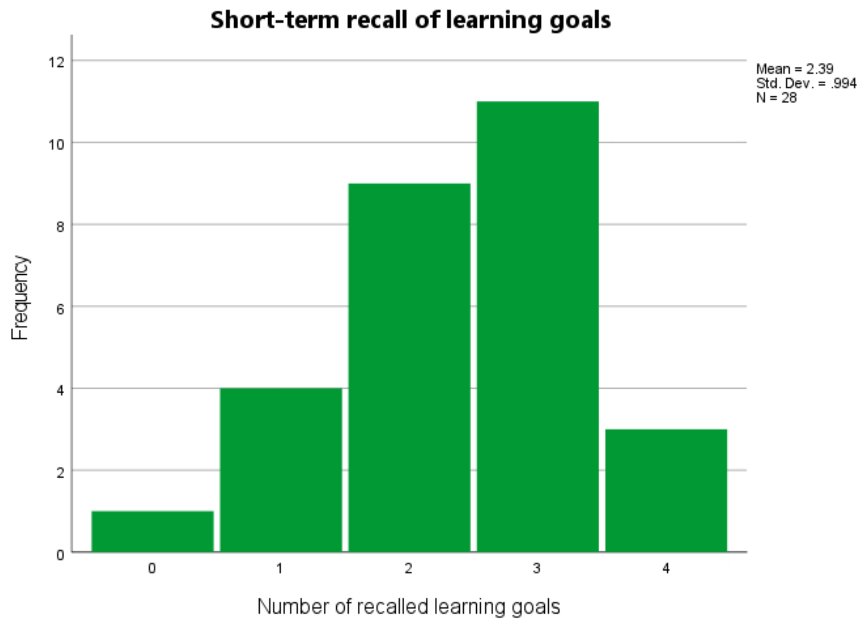


Figure 66. Short-term recall of learning goals – all participants.

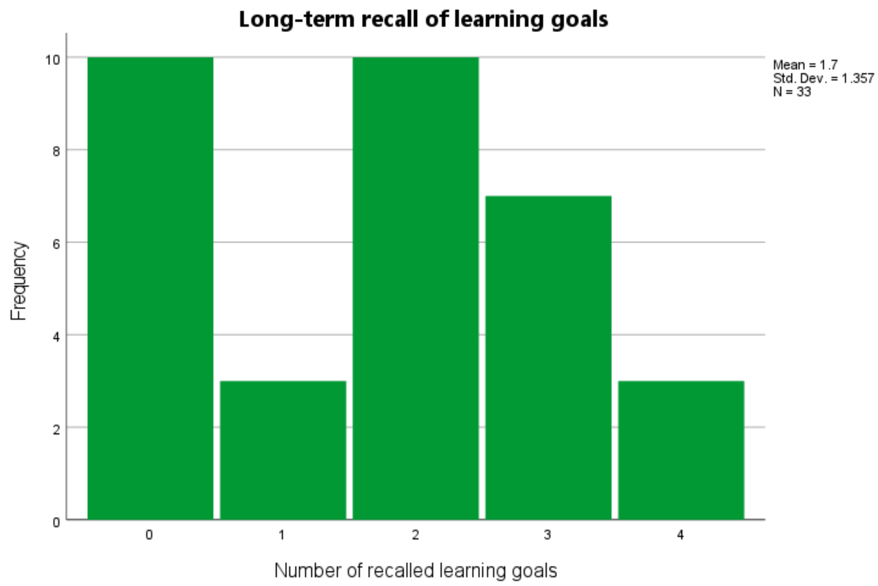


Figure 67. Long-term recall of learning goals - all participants.

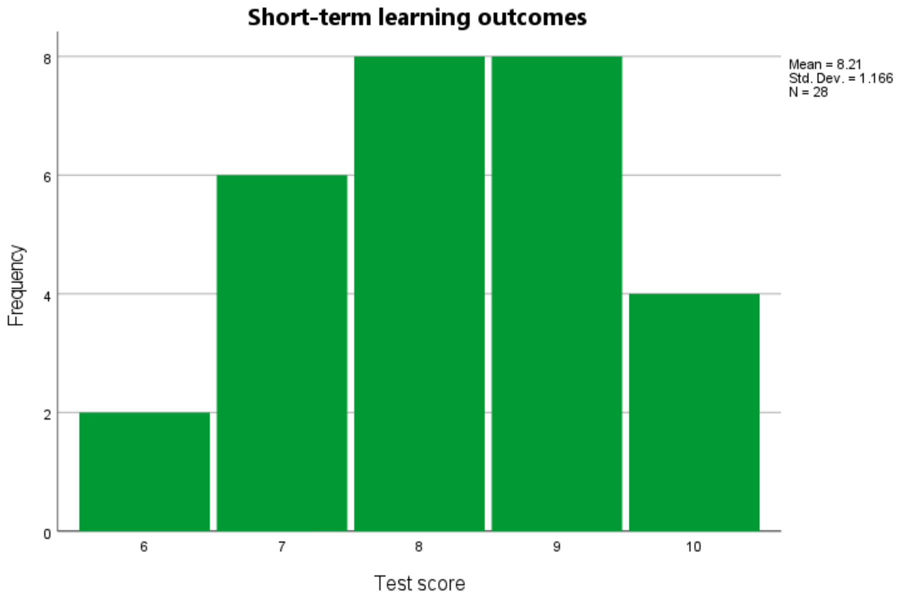


Figure 68. Short-term recall of learning outcomes – all participants.

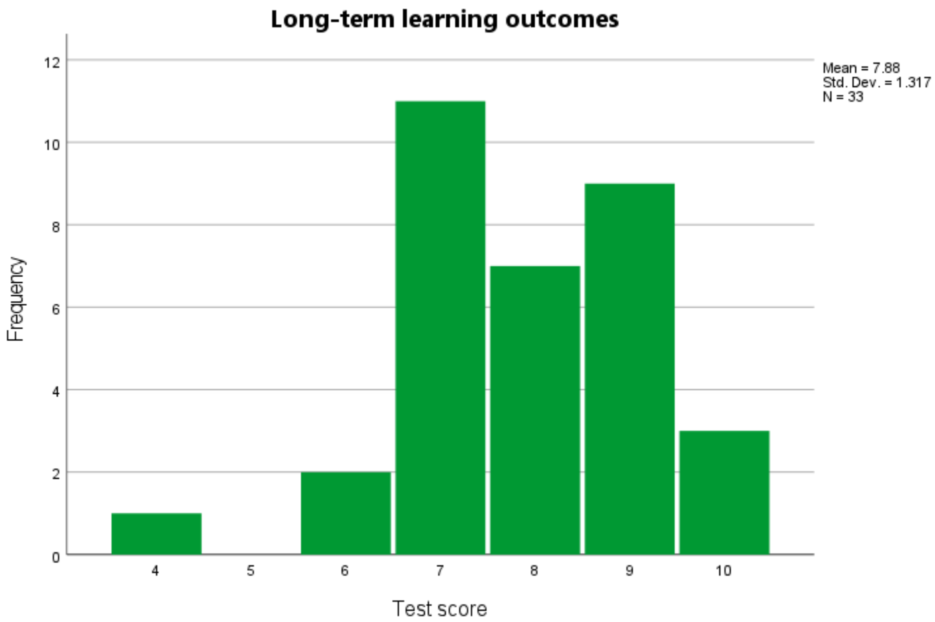


Figure 69. Long-term learning outcomes – all participants.

Appendix O SPSS output

Tests of Normality

Condition		Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
Control condition	LG_Total	.310	14	<.001	.865	14	.036
	Q_Total	.256	14	.013	.878	14	.054
Experimental condition	LG_Total	.278	14	.004	.878	14	.055
	Q_Total	.257	14	.013	.913	14	.174

a. Lilliefors Significance Correction

Figure 70. Shapiro-Wilk short-term.

Mann-Whitney Test

Ranks

	Condition	N	Mean Rank	Sum of Ranks
LG_Total	1	14	15.86	222.00
	2	14	13.14	184.00
	Total	28		

Test Statistics[a]

	LG_Total
Mann-Whitney U	79.000
Wilcoxon W	184.000
Z	-.919
Asymp. Sig. (2-tailed)	.358
Exact Sig. [2*(1-tailed Sig.)]	.401[b]

a. Grouping Variable: Condition

b. Not corrected for ties.

Figure 71. Mann-Whitney U-test short-term recall of learning goals.

T-Test

Group Statistics

	Condition	N	Mean	Std. Deviation	Std. Error Mean
Q_Total	1	14	8.36	1.336	.357
	2	14	8.07	.997	.267

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance		Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p			Lower	Upper
Q_Total	Equal variances assumed	4.025	.055	.641	26	.264	.527	.286	.446	-.630	1.202
	Equal variances not assumed			.641	24.052	.264	.527	.286	.446	-.634	1.205

Independent Samples Effect Sizes

		Standardizer[a]	Point Estimate	95% Confidence Interval	
				Lower	Upper
Q_Total	Cohen's d	1.179	.242	-.504	.984
	Hedges' correction	1.214	.235	-.489	.955
	Glass's delta	.997	.287	-.468	1.030

a. The denominator used in estimating the effect sizes.
Cohen's d uses the pooled standard deviation.
Hedges' correction uses the pooled standard deviation, plus a correction factor.
Glass's delta uses the sample standard deviation of the control group.

Figure 72. T-test short-term learning outcomes.

Tests of Normality[a,b]

Condition		Kolmogorov-Smirnov[c] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
1	LG_Total	.216	13	.100	.890	13	.098
	Q_Total	.250	13	.026	.857	13	.035
2	LG_Total	.264	11	.031	.878	11	.099
	Q_Total	.210	11	.191	.896	11	.165
3	LG_Total	.317	9	.009	.767	9	.009
	Q_Total	.257	9	.088	.903	9	.273

a. There are no valid cases for LG_Total in one or more split file. Statistics cannot be computed.

b. There are no valid cases for Q_Total in one or more split file. Statistics cannot be computed.

c. Lilliefors Significance Correction

Figure 73. Shapiro-Wilk test long-term.

Tests of Normality

Condition		Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
1	LG_Total	.237	24	.001	.878	24	.008
	Q_Total	.198	24	.016	.871	24	.005
3	LG_Total	.317	9	.009	.767	9	.009
	Q_Total	.257	9	.088	.903	9	.273

a. Lilliefors Significance Correction

Figure 74. Shapiro-Wilk test long-term – group vs indiviual.

T-Test

Group Statistics

	Condition	N	Mean	Std. Deviation	Std. Error Mean
LG_Total	1	13	1.92	1.320	.366
	2	11	2.18	1.328	.400

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
LG_Total	Equal variances assumed	.045	.834	-.477	22	.319	.638	-.259	.542	-1.384	.866
	Equal variances not assumed			-.477	21.305	.319	.638	-.259	.543	-1.386	.869

Figure 75. T-test long-term recall of learning goals.

Mann-Whitney Test

Ranks

	Condition	N	Mean Rank	Sum of Ranks
LG_Total	1	24	19.33	464.00
	3	9	10.78	97.00
	Total	33		

Test Statistics[a]

	LG_Total
Mann-Whitney U	52.000
Wilcoxon W	97.000
Z	-2.342
Asymp. Sig. (2-tailed)	.019
Exact Sig. [2*(1-tailed Sig.)]	.023[b]

a. Grouping Variable: Condition

b. Not corrected for ties.

Figure 76. Mann-Whitney U-test for long-term recall of learning goals – group and individual.

Mann-Whitney Test

Ranks

	Condition	N	Mean Rank	Sum of Ranks
Q_Total	1	13	12.73	165.50
	2	11	12.23	134.50
	Total	24		

Test Statistics[a]

	Q_Total
Mann-Whitney U	68.500
Wilcoxon W	134.500
Z	-.180
Asymp. Sig. (2-tailed)	.857
Exact Sig. [2*(1-tailed Sig.)]	.865[b]

a. Grouping Variable: Condition

b. Not corrected for ties.

Figure 77. Mann-Whitney U-test for long-term learning outcomes.

NPar Tests

Mann-Whitney Test

Ranks

	Condition	N	Mean Rank	Sum of Ranks
Q_Total	1	24	19.08	458.00
	3	9	11.44	103.00
	Total	33		

Test Statistics[a]

	Q_Total
Mann-Whitney U	58.000
Wilcoxon W	103.000
Z	-2.092
Asymp. Sig. (2-tailed)	.036
Exact Sig. [2*(1-tailed Sig.)]	.044[b]

a. Grouping Variable: Condition

b. Not corrected for ties.

Figure 78. Mann-Whitney U-test for long-term learning outcomes - group and individual.

Descriptive Statistics

Condition		N	Minimum	Maximum	Mean	Std. Deviation
Control condition	LG_Total	14	0	4	2.50	1.160
	Q_Total	14	6	10	8.36	1.336
	GNep_Echt	14	3	5	4.14	.770
	GLijktopeenmachine_Lijktopeenmens	14	1	4	2.29	.914
	GStaatstil_Beweegtveel	14	2	5	4.00	.877
	GOnaardig_Aardig	14	1	5	4.64	1.082
	GDom_Slim	14	1	5	4.50	1.160
	Valid N (listwise)	14				
Experimental condition	LG_Total	14	1	4	2.29	.825
	Q_Total	14	6	10	8.07	.997
	GNep_Echt	13	3	5	4.08	.641
	GLijktopeenmachine_Lijktopeenmens	13	1	5	3.00	1.080
	GStaatstil_Beweegtveel	13	2	5	4.00	1.000
	GOnaardig_Aardig	13	5	5	5.00	.000
	GDom_Slim	13	4	5	4.77	.439
	Valid N (listwise)	13				

Figure 79. Robot assessment descriptive statistics per condition.