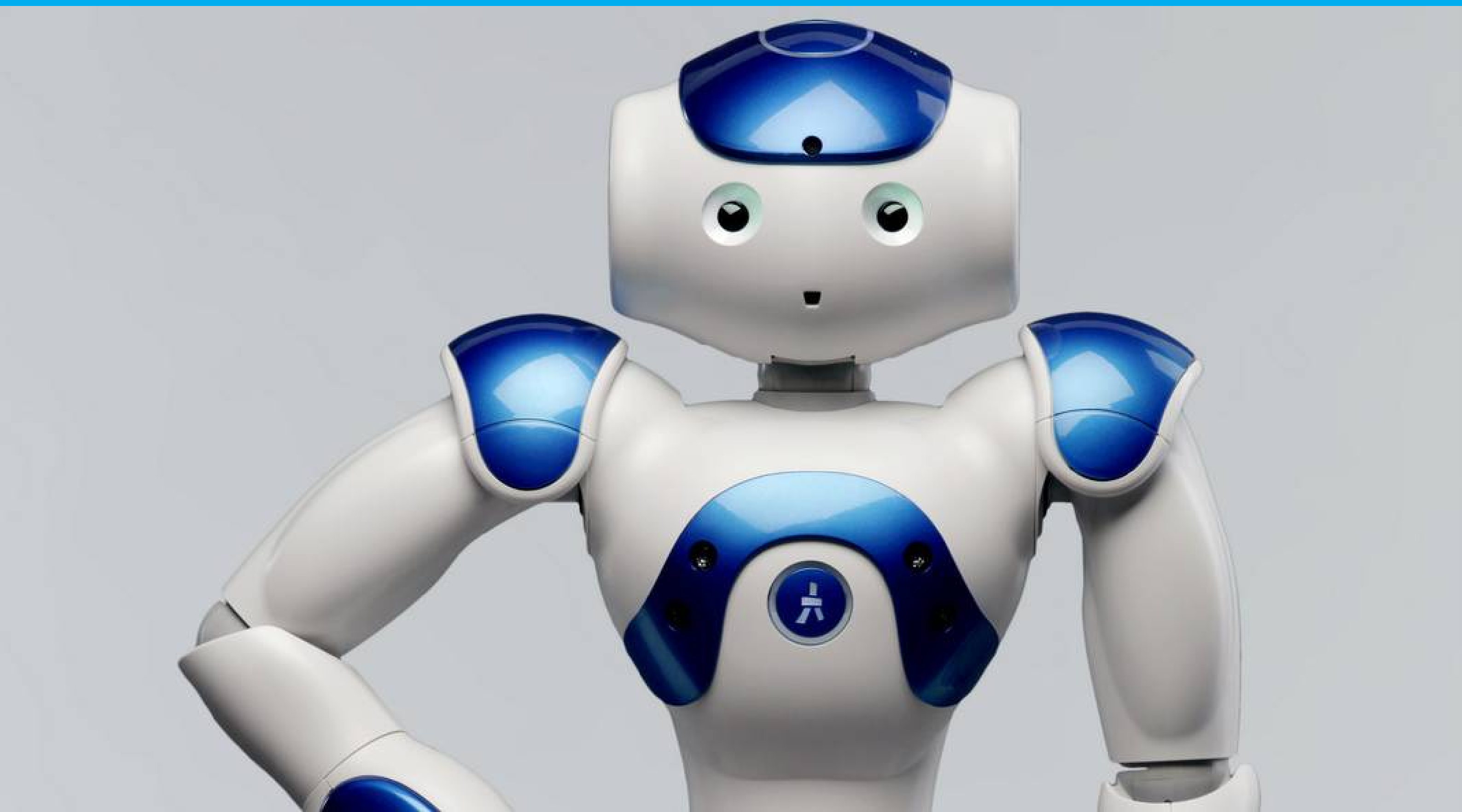
The Effects of a Social Robot's Gestures on Learning Outcomes

F.N. Moorlag



TU Delft

The Effects of a Social Robot's Gestures on Learning Outcomes

by

F.N. Moorlag

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Wednesday July 14, 2021 at 12:00 PM.

Student number:	4929209	
Project duration:	May 1, 2021 – July 14, 2021	
Thesis committee:	Dr. ir. J.C.F. De Winter,	TU Delft, supervisor
	Dr. ir. D. J. Broekens,	Interactive Robotics, supervisor
	Dr. ing. L. Marchal Crespo,	TU Delft

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

TUDelft INTERACTIVE ROBOTICS

PREFACE

Before you start reading this paper, there are a few things you should know. Firstly, this is a research involving a robot, a social robot named NAO. Like other robots, NAO is physically embodied and on top of that, NAO is able to produce human-like gestures by rotating its joints. Secondly, this thesis was an instructive, challenging, and amazing project to be involved in. The seeds for this project were sown in my first year, while following the courses Human-Machine Systems and The Human Controller. At that moment, the exact specifics of the thesis were unclear, but after my internship in Japan, the country that is forerunner in the field of social robotics, I was determined to continue exploring this fascinating area.

At first, I want to thank my supervisor Joost Broekens for showing his enthusiasm in this project from day one, for his guidance and support during the whole time span of project, and for allowing me to use and test with the robots as often as I needed. I also thank Bas and the other team members of Interactive Robotics, especially the developers Jurjen and Diony, who helped me to get acquainted with programming the robot.

A special thanks should be given to Joost de Winter, my TU Delft supervisor, for his knowledge, accurate comments, and valuable input during the virtual meetings. Without his feedback, this paper would not have been at this quality. He introduced me to other interesting studies and researchers, such as Dimitra Dodou, who I also would like to thank for her hard work on the meta-analysis.

I show my gratitude to Basisschool De Wijzer, De Paasbergschool, and the Pieter Bruegheschool for kindly welcoming NAO and me during the experiments. Additionally, to all the curious and energetic children who participated, I hope that your first interaction with a social robot has been a great experience.

Lastly, I want to thank my family and friends for always showing interest in my robot adventures and their support and encouragement. A special thanks to Sonia, Pieter, Floor, and Marnix for welcoming NAO as a guest when I was testing the robot at home during the lockdown. To the reader, I can only hope that reading this thesis sparks the same interest as I had when I started looking into social robots, now more than a year ago. It was a pleasant ride, I hope you enjoy reading this thesis as much as I enjoyed accomplishing it.

F.N.Moorlag
Amsterdam, 2021

CONTENTS

The Effects of a Social Robot's Gestures on Learning Outcomes

Moorlag, F. N.[1], Broekens, J.[2], De Winter, J. C. F. [3]
[1] Department of BioMechanical Engineering, Delft University of Technology, The Netherlands
[2] Affective Computing at Leiden Institute of Advanced Computer Science, Leiden University, The Netherlands
[1] Department of Cognitive Robotics, Delft University of Technology, The Netherlands

Abstract— **In recent years there has been an increasing interest in the development of social robots for educational purposes. A presumed advantage is their physical presence in the learner's referential world. However, it remains an open question whether this embodied presence of a robot can be utilized to support learning with robots' gestures. The aim of our study was to program a social robot that tutors mathematics through verbal explanations supported by deictic, beat, and iconic gestures that align with explanations, and to analyse its effectiveness on learning outcomes. The learning outcomes entail cognitive outcomes and affective outcomes. A between-subject experiment was performed to research the effect of the robot's use of supporting gestures on cognitive and affective learning outcomes compared to a robot that does not perform gestures or performs random gestures. In total, 78 children (M_{age}= 8.3 years) from Dutch primary schools participated. Results showed that the social robot effectively taught children mathematical equivalence by a significant increase in test scores between the pre- and post-test. No differences were found in test scores between participants in the supportive gesturing condition, the no gesturing condition, and the random gesturing condition. Participants showed high task and tutor engagement scores and head direction estimations revealed that participants were attentive towards the robot, but no differences were found between the three conditions. However, the gestures 'helpfulness' was rated significantly higher by participants in the supportive gesture condition than the random gesture condition, suggesting that children felt the gestures mattered.**

Index Terms—**Social robots, Human-Robot Interaction, Educational robots, Robot gestures**

I. Introduction

In recent years there has been an increasing interest in the development of social robots for educational purposes (Minkelen et al., 2020) and, as argued by Mubin et al. (2013), these robots show great potential to benefit the field of education where the robots may deliver content similarly to human tutors. A social robot is defined by Bartneck and Forlizzi (2004) as "a (semi) autonomous robot that interacts and communicates with humans by following expected behavioural norms". Like other robots, a social robot is physically embodied, whereas avatars or on-screen synthetic characters are not. When comparing these robots to traditional learning tools, the presumed advantage is their physical presence in the learner's referential world (Leyzberg et al., 2012). Wijnen et al. (2019) mentioned that this physical presence enables the robot to interact more naturally with the learner, for instance, by pointing, gazing, or gesturing.

Furthermore, it is argued that gesturing might have a 'grounding' effect in learning new concepts by linking existing perceptual and motor experiences to the meaning of the concept (De Wit et al., 2020). Gestures that support explanations are also mentioned to be a crucial part of communication with children. These supportive gestures consist of iconic, deictic, and beat gestures. Beat gestures rhythmically align with speech, deictic gestures are performed to direct a learner's attention towards a specific referent in the proximal or distal environment, whereas the shape of iconic gestures has some physical similarity to its referent (McNeil, 1992; Vogt et al., 2019).

Previous research has been conducted on the effect of gestures on learning outcomes in educational child-robot interactions. As mentioned by Schodde et al. (2019), cognitive learning typically refers to skills and knowledge to be learned by the student, whereas affective learning represents aspects such as motivation and attitudes toward the subject or content to be learned. Significantly higher learning gains and engagement levels were found for children that interacted with NAO with expressing gestures than for NAO without gestures during a second language tutoring session (De Wit et al., 2018). On the other hand, research by Vogt et al. (2019) did not result in significant differences in both cognitive and affective outcomes between children that were taught a second language by NAO with iconic gestures and NAO without iconic gestures. Consequently, it remains an open question whether the embodied presence of a social robot can be fully utilized to support learning with a robot's gesturing, and if so, what differences arise between supportive and random gestures.

The implementation of gestures and nonverbal feedback during mathematical tutoring sessions has so far not yet been researched, but this seems an interesting addition along with verbal interaction (Hindriks & Liebens, 2019). Therefore, the projected aim is to program a social robot that tutors mathematics through verbal explanations supported by deictic, beat, and iconic gestures that align with these explanations, and to analyse its effectiveness on learning outcomes. The following associated research question is proposed:

- What is the effect on learning outcomes of a social robot that tutors mathematics to primary school students with supporting gestures?

Hypothesis: it is hypothesised that the learning outcomes of students are positively affected by interacting with the social robot tutor that demonstrates supportive gestures that align with explanations when tutoring mathematics (H1). It is expected that the robot tutor that demonstrates supportive gestures during mathematical explanations more positively affects the student's cognitive learning outcomes than (1) a social without gestures ('alive') condition' or than (2) a social robot that demonstrates random gestures (H2). It is also expected that participants that interact with the supportive gesturing robot show different affective learning outcomes, with higher engagement and attention towards the robot, than participants that interact with a social without gestures or with a social robot that demonstrates random gestures (H3).

To test these expectations, an experiment was carried out in April 2021. The experimental method was based on related work, which is discussed in Chapter II, followed by an examination of the robot's explanation design for the supportive gesturing condition, the no gesturing condition, and the random gesturing condition in Chapter III. The experimental method will be introduced in chapter IV, and the results are presented thereafter. Finally, a discussion with limitations is considered in Chapter VI, and a conclusion with future recommendations in Chapter VII.

II. Related work

In this chapter, studies related to supporting gestures in education are reviewed. First, studies related to supporting gestures that align with explanations in current mathematical education for human-to-human interaction (HHI) are discussed in Section II-A. In Section II-B, an overview of research about supporting gestures in human-computer interaction (HCI) is presented. Lastly, human-robot interaction (HRI) research towards robots' gestures in education is introduced in Section II-C.

A. Supportive gestures in mathematical education in HHI

Human-robot interactions in education are often drawn upon ways in which human teachers deliver content to students in their learning activities (De Wit et al., 2018). Previous human-to-human interaction research has been conducted towards the effectiveness of observing teachers' gestures on learning mathematics. It was concluded by Goldin-Meadow et al. (1999) that gestures facilitated participants' comprehension of the teachers' math explanations when they matched speech, and hindered participants' comprehension of the math explanations when they mismatched speech. Alibaba and Nathan (2012) argued that mathematical cognition is embodied, based on action and perception, and grounded in the physical environment. Valenzeno et al. (2003) showed that pre-school children who were taught symmetry by an instructor who presented meaningful pointing gestures, learned more than children who were not exposed to gestures, measured through post-test scores. Related studies have found that observing teachers' gestures facilitates learning in other mathematical paradigms, including the learning of mathematical equivalence (e.g., $2+4+6=\_+6$; Perry et al., 1995; Singer & Goldin-Meadow, 2005; Cook et al., 2013; Congdon et al. 2017; Wakefield et al., 2018).

Wakefield et al. (2018) conducted experiments with children aged between 8-10 who were instructed mathematical equivalence by human instructors using Speech alone and Speech + (iconic and deictic) Gesture. The children had never solved mathematical equivalence problems before and conducted a pre- and post-test containing six missing addend equivalence problems. The problems were presented both in Form A, the last addend on the left side is the same as on the right side of the equals sign (e.g., $7+5+3=\_+3$), and in Form B, the first addend on the left side is repeated on the right side (e.g., $4+6+9=4+\_$). After the pre-test, the children watched six video instructions and solved six mathematical equivalence problems. Post-test scores revealed that children in the Speech + Gesture condition answered significantly more problems correctly than children in the Speech Alone condition.

Similar to Wakefield et al. (2018), Congdon et al. (2017) conducted the equalizer strategy to explain mathematical equivalence. This technique emphasizes that both sides of the equal sign must yield the same numerical value, and for each side, calculations can be performed separately (Goldin-Meadow et al., 1999). In the study by Condon et al. (2017), the instructor explained the problem $8+6+2=\_+2$ by saying; 'I want to make one side equal to the other side, eight plus six plus two is sixteen, and fourteen plus two is sixteen. So, one side is equal to the other side'. It was found that children performed significantly better on the post-test after receiving simultaneous Speech + Gesture instruction than after first the Speech and then Gesture instructions.

B. Supportive gestures in mathematical education in HCI

Furthermore, Cook et al. (2017) conducted research in which children observed lessons on mathematical equivalence explained by an avatar that either gestured or did not gesture, while head position, lip movements, and eye gaze remained identical across conditions. The avatar produced both content gestures (beat, iconic, and deictic), created to strengthen the conceptual content, and bimanual beat gestures, aimed to increase the avatar's charisma and appeal. Form A and Form B problems were both explained by the avatar, and children solved mathematical equivalence problems at the computer. From post-test scores, it appeared that children who observed the gesturing avatar learned significantly more, and they solved the mathematical equivalence problems more quickly.

C. Supportive gestures in education in HRI

The detailed and accurate motions humans make cannot yet be performed by social robots due to their limited degrees of freedom. De Wit et al. (2018) argued that this could lead to a loss of understanding when human gestures are being transferred directly to robots. However, research by Bremner & Leonards (2016) suggested that a social robot's iconic gestures are almost as comprehensible, compared to a human's gestures. Thus far, research towards gestures of social robots in education has mainly been focused on language learning. De Wit et al. (2018) let the social robot NAO perform iconic gestures and found significantly better learning gains and higher engagement for children with this robot than for the NAO without gestures. Conti et al. (2019) also found significantly higher test scores for children that engaged in memorizing games with an expressive NAO with gestures compared to children that interacted with a static NAO.

Contrastingly, research by Vogt et al. (2019) studied the effect between children that were taught a second language by NAO with iconic gestures and NAO without iconic gestures. No significant differences were found in both cognitive and affective learning outcomes. Additionally, De Wit et al. (2020) found that the robot's use of gestures - either repeated or varied - did not affect learning outcomes during a second language learning tutoring session. Thus far, it has not yet been researched what the effects are on learning outcomes of children that interact with a robot that performs iconic, deictic, and beat gestures compared a robot without gestures and a robot with random gestures.

III. Design of the robot's explanation

In this chapter, a design of interaction is introduced with the explanations and gestures that were automatically presented by a Softbank Robotics NAO v6. In Section III-A, the explanations and gestures are introduced and in Sections III-B, III-C, and III-D thereafter, the system design of the three robot conditions are considered, with the supportive gesturing condition, no gesturing ('alive') condition, and the random gesturing condition respectively. The overarching Python files with which the robot was automatically controlled are discussed in the last section.

A. Gestures and speech performed by the social robot

As mentioned in the previous chapter, most gesture studies in mathematical education are focused on mathematical equivalence. Furthermore, the equalizer strategy is often used and found to be effective (Congdon et al., 2017; Cook et al., 2017; Wakefield et al., 2018). Following these previous studies in HHI and HCI, it was decided that the social robot NAO would tutor mathematical equivalence through the equalizer strategy with gestures that align with speech, e.g. "One side equals the other side" with a right-handed sweep followed by a left-handed sweep (iconic gestures). For our study, it was chosen to execute a one-on-one tutoring interaction between a child and NAO, with the robot explaining this mathematical concept. It is believed that one-on-one tutoring is the best way to learn, and it offers the greatest potential for software systems (Belpaeme et al., 2018). The beat, deictic, and iconic gestures and explanations that were presented by NAO are shown in Appendix A in Tables 6-8. These explanations and gestures were based on research by Cook et al. (2017) as indicated by the last column of the tables.

The explanations and gestures that were performed by NAO during the tutoring sessions were developed through the platform 'https://platform.robotsindeklas.nl', created by 'Interactive Robotics', which is cloud connected to the robot. With the ROM (Robot Operation Module) and the RIE (Robot Interaction Engine), the NAO tutoring system was directly and indirectly controlled to manage its speech and movements. The scripts were created in Python 3.8 and through connection with the WAMP server, actions were generated by the robot[1,2]. Each explanation section with speech and an aligned iconic, beat, or deictic gesture was separately programmed, and these were sequentially called in one overarching file. As examined in the following sections, different techniques were implemented for deictic gestures, and iconic and beat gestures, as well as for the supportive gesture condition, for the no gesture ('alive') condition and the random gesture condition.

[1]The Gitlab with Python codes may be shared upon request

[2]https://www.youtube.com/channel/UCape8hI2xD4J1Imxus3MAWQ

B. System design for the supportive gesturing robot condition

In this section, the implementation of iconic and beat gestures for the supportive gesture robot is discussed first, followed by the implementation design of the deictic gestures.

1) Implementation of iconic and beat gestures: To align gestures with speech, various models and implementation techniques have been introduced. According to Holroyd and Rich (2012), languages that are based on Petri nets, have been used to control many virtual agents and is also making an introduction for the implementation in robot applications. These Petri nets were also used in the present study to coordinate the timing of speech and gesturing described in Section III-A and presented in Tables 6-8 in Appendix A. As shown in Figure 15 in Appendix B, these Petri nets consist of certain places (represented as circles) and transitions (represented as vertical bars). Each place represents a synchronisation point of a gesture, and the transitions represent changeovers between these points (Kopp et al., 2006; Holroyd & Rich, 2012). In each explanation section, the speech is partitioned into an intro and a main (and, if applicable, an end) part, in which the robot is supposed to gesture, as can be derived from the video material provided by Cook et al. (2017). An example is shown in Appendix B, where the robot presents a beat gesture while saying; ['about the equal sign'], and no movements are generated while saying; 'Today we are going to learn'.

The Python scripts for the explanation sections with iconic and beat gestures were thus constructed in similar frames and examined in Figure 19 in Appendix C. Each explanation section was separately programmed, and these were sequentially called in one overarching file. To control the movements of NAO, each of the gesture phases was planned in time frames with certain movement data. From the Aldebaran documentation (Robotics, 2015), joints and joint ranges (radians) that can be rotated by NAO were provided and these were used for the movement data to reproduce the gesture that is performed by Cook et al.'s (2017) virtual agent. The timing in milliseconds was retrieved from the videoclips provided Cook et al. (2017) and adjusted when this exceeded the angular velocity of the NAO robot's joints. Joints include head joints (yaw and pitch), left and right arm joints (shoulder pitch and roll, elbow yaw and roll, wrist yaw, and hand), and left and right leg joints (hip pitch and roll, knee pitch, and ankle pitch and roll). The elaboration of such a frame with timing, movement data, and synchronisation can also be found in Figure 19 in Appendix C.

2) Implementation of deictic gestures: To let NAO produce the deictic gestures, or pointing gestures, a different technique was developed in which the use of Aldebaran's landmark detection is exploited (Robotics, 2015). Through this vision module, the robot can recognise special landmarks, Naomarks, with specific patterns on them by capturing images with the two identical videocameras located in the forehead (up to 1280x960 resolution at 30 frames per second). The Naomarks can be placed at different locations in the robot's field of action and information can be obtained on the location of the robot with respect to those landmarks. In Figure 1, a mathematical equivalence problem that the NAO robot explained through pointing gestures, is shown with Naomarks, containing id's 68, 119, 107, 85 from left to right respectively, placed on top of the sum.

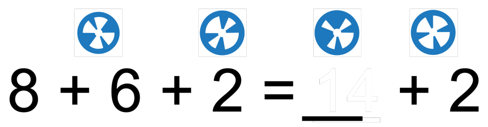


Figure 1: Mathematical equivalence problem with Naomarks 68, 119, 107, 85 from left to right, respectively, placed on top of the sum

Whenever a robot detects one of these Naomarks, the following information is extracted; [abs-x, abs-y, size, rel-x, rel-y, rot, id]. Through the absolute x-angle and the absolute y-angle, the location of that specific Naomark with respect to the robot's forehead, and so the pointing direction, can be obtained. In Figure 21 in Appendix C, the Python scripts for the explanation sections with deictic gestures are presented. At first, a head movement index was created to let the robot 'search' for the Naomarks. The head movement list consists of 16 datapoints that the robot followed while the vision was streaming to find Naomarks. Whenever Naomarks were detected, a dictionary was filled with 'time': capturing time, 'data': 'body': [abs-x, abs-y, size, rel-x, rel-y, rot, id]. In the case that more than one Naomark was captured, the 'body' dictionary was filled with multiple lists.

If the robot detected the specific Naomark that was needed for a certain explanation section, e.g., Naomark 107 where the robot said 'what goes inside the blank' after pointing at the blank spot, the absolute x- and y-angle (α and β respectively) were retrieved. With these angles, the desired pointing direction of the robot's head and arm were obtained. For Naomarks 107 and 85, on right side of the equal sign, NAO used the right shoulder joints to point when the angle was smaller than 0.1 rad (positive abs-x on the left side of the robot's viewing direction) and the left shoulder joint for angles larger than 0.1 rad. For Naomarks 68 and 119, located on the left side of the equal sign, NAO used the left shoulder joints to point if the angle was larger than -0.1 rad and the right shoulder joint for angles smaller than -0.1 rad. To prevent overshooting, the final pitch and yaw directions of the robot's head and shoulder joints arrived in steps, e.g. timestep-1 is $\frac{2}{3}\alpha$ and timestep-2 is α. This frame is shown in Figure 21 in Appendix C.

C. System design of the no gesturing robot condition

In this section, the system design for the second condition, the no gesture ('alive') condition is examined. The implementation of iconic and beat gestures is discussed first, followed by the implementation design of the deictic gestures. However, instead of performing gestures, the robot in the no gesture condition merely generated small movements to look alive.

1) Implementation of iconic and beat gestures: Similar techniques were implemented for the no gesture condition as for the supportive gesturing condition including synchronisation points, transitions, and partitioned speech. The same frame was implemented for the no gesture condition as for the gesture condition in Section III-B1(Figure 19 in Appendix C), including timing, head joints (yaw and pitch), left and right leg joints (hip pitch and roll, knee pitch, and ankle pitch and roll), and synchronisation. However, the arm joints (shoulder pitch and roll, elbow yaw and roll, wrist yaw, and hand) were left out. Through this approach, the frequency and intensity (duration) of the movements were equal to the supporting gesture condition, while the no gesture robot did not perform any gestures but looked 'alive'.

2) Implementation of deictic gestures: The same techniques were also implemented for the no gesture condition as for the supportive gesturing condition for the deictic gestures. In Figure 22 in Appendix C, the Python scripts for the explanation sections with deictic gestures are presented for the 'alive' condition. Again, a head movement index was created at first to let the robot 'search' for the Naomarks. If the robot detected the specific Naomark that was needed for a certain explanation section, e.g., Naomark 107 where the robot said; 'what goes inside the blank', the vision stream was ended and the robot returned to the $gesture_{end}$ and says the $speech_{main}$ (and is applicable $speech_{end}$). With this framework (as described in Figure 22 in Appendix C), the no supporting gesture robot looked 'alive', and the structure and durations of the explanation sections were similar to the explanation sections of the supportive gesturing robot.

D. System design of the random gesturing robot condition

The system design for the random gesturing condition is presented in this section. The implementation of iconic and beat gestures is discussed first, followed by the implementation design of the deictic gestures. However, instead of performing these gestures, the robot in the random gesturing condition performs randomly generated gestures.

1) Implementation of iconic and beat gestures: Again, similar techniques were implemented for the random gesture condition as for the supportive gesturing condition including synchronisation points, transitions, and partitioned speech. The random movements of the robot were generated from the 23 standard 'bodytalk' gestures that are available by default on the NAO. For each iconic, deictic, or beat gesture in the supportive gesturing condition, the robot performed one of these 23 gestures that were randomly selected per session. To ensure that the frequency and intensity of the movements were equal to the supporting gesture condition, the random movement was performed the same number of times as the supportive movement through the functions `yield sleep` (equal number of seconds as supporting gesture) and `yield stop` (Figure 20 in Appendix C).

2) Implementation of iconic and deictic gestures: For the random 'deictic' gestures, similar techniques were implemented for the random gesture condition as for the supportive and no-gesturing condition. The same mathematical equivalence problem that shown in Figure 1 with Naomarks, containing id's 68, 119, 107, 85 from left to right respectively, placed on top of the sum were used in this condition. However, instead of pointing towards these, the robot looked towards them and performed random gestures. The gestures in the random gesture condition were generated from the 23 standard 'bodytalk' gestures, in the same way as explained in the previous section. In Figure 23 in Appendix C, the Python scripts for the explanation sections with deictic gestures are presented for the random condition. Again, a head movement index is created at first to let the robot 'search' for the Naomarks. If the robot detected the specific Naomark that was needed for a certain explanation section, e.g., Naomark 107 where the robot said; 'what goes inside the blank', the vision stream was ended, and the robot performed the random gesture for the same amount of time as the supportive gesture in the condition with supporting gestures, and said the $speech_{main}$ (and if applicable $speech_{end}$). With this framework (Figure 23 in Appendix C), it is ensured that the frequency and intensity of the movements were equal to the supporting gesture condition.

E. Overarching robot control

1) Name storage and timing: Prior to the explanation sections, the participant's name was acquired through an `input()` function and stored as a json file. Throughout the tutoring session, the stored name was called and referred to by the robot in some explanation sections; e.g. 'Okay, *[name]* now it's your turn', to personalize the session (De Wit et al. 2020). This was equal for all conditions. Additionally, each of the gesturing sections were timed in the overarching file with the function `time()` for analysis.

2) Turning towards and away from the screen: Whenever the robot pointed to the screen, or only searched for a Naomark and performed a random or no gesture, the robot had to turn towards the screen first. This was achieved by calling 'turnleft' or 'turnright' behaviours that are available by default on the NAO. By capturing the duration of the turns, it was found that the robot needed to turn 3.7 seconds to the left to face the screen and after the deictic explanation sections, 3.7 seconds right to face the participant again. This is implemented by `yield sleep(3.7).`

3) Controlling the screen: During the tutoring session, the robot gave examples of mathematical equivalence that were presented on a screen, such as shown in Figure 1. At first, a blank was presented and once the robot had explained the mathematical equivalence problem, the right answer appeared on the screen e.g. for Figure 1, the sum $8 + 6 + 2 = \_ + 2$ changed into $8 + 6 + 2 = 14 + 2$. A PDF was created with the mathematical equivalence problems and their outcome on the next slide that were shown on the screen during the robot's explanation (Appendix D). To ensure that the presentation aligned with the explanation, an app with this file was created on the portal 'https://platform.robotsindeklas.nl' with. To show the next slide, `yield sess.publish ('rie.legacy.slide_control', ppt_ command = 'NEXT')` was called in the overarching file and the app automatically showed the next slide.

IV. Experimental method

A between-subject experiment was carried out in April 2021 to research the effect of the robot's use of supporting gestures on cognitive and affective learning outcomes compared to a robot that does not perform gestures or performs random gestures. The following experimental conditions were tested: 1) Supportive gestures, the robot performs iconic, beat, and deictic gestures that align with explanations. 2) No gestures, the robot uses speech to explain mathematical equivalence and does not perform iconic, beat, and deictic gestures that align with speech, but the robot moves to look 'alive'. 3) Random gestures, the robot uses speech to explain mathematical equivalence and performs random gestures. Other than the differences in the social robot's use of gesturing, the experimental conditions were identical.

Prior to the experiments, a pilot study was carried out at the University of Leiden in February 2021 to test the experimental method plans. Two children (aged 9, female and male) that had no prior knowledge of mathematical equivalence participated in the pilot study and interacted with the social robot. The procedure and robot's gestures in the pilot study were similar to the tutoring session of this study. The participants, measures for cognitive and affective outcomes, and the procedure are examined below.

A. Participants

In total, 78 participants (35 male and 43 female, M_{age}= 8.3 years, SD = 0.5 years) from Arnhem and surrounding areas participated in the experiment. Primary schools in this area were contacted and 17 children from Basisschool De Wijzer in Arnhem, 36 from De Pieter Brueghel School, and 25 from De Paasbergschool in Oosterbeek participated in the experiment. To take the novelty effect into account, children were introduced to NAO in a classroom introductory session prior to the one-to-one experiments. Similar to gestural robotics research by De Wit et al. (2020), the participants were pseudo-randomly assigned to one of the conditions with a balanced distribution of school, age, and gender. The study was approved by the research ethics committee of the Delft University of Technology and informed consent was given by the children's parents and teachers prior to their participation. The information letter and informed consent forms for the parents/legal representatives are shown in Appendix E and F. Furthermore, the experimenter followed the RIVM-guidelines concerning the COVID-19 situation as the prescribed 1.5-m distance was being maintained and the researcher wore a facemask.

B. Measures for cognitive learning outcomes

Children's mathematical equivalence knowledge was measured at different times by means of a test based on

Cook et al. (2017) and Wakefield et al. (2018). Pre-, mid-, post-, and delayed post-test questions consisted of ten novel equal addends equivalence problems. The tests started with three problems in Form A ($a + b + c = \_ + c$) and three problems in Form B ($p+q+r = p+\_$), which had identical formats to the problems explained by the robot. The tests ended with two problems with equal addends located in a matching position ($x + y + z = \_ + x$) and two problems with no equal addends ($l + m + n = o + \_$). The formats of the pre-, mid-, post-, and delayed post-test were thus identical, and the questions were randomly generated by a Python model; integers between 1 and 15 were summed up, and the final summations added up to integers between 10 and 20. The tests were identical for all participants and are represented in Appendix I, J, K, and L. No feedback was given to the participants when they were answering the questions on the tests to prevent learning that could arise because of this feedback.

C. Measurement methods for affective outcomes

The affective outcomes were measured using multiple methods, including a questionnaire and head directions estimated from recorded videos that were captured during the experiments. In the sections below, both methods are further discussed.

1) Questionnaire: The questionnaire was created with (1) a Task and Tutor Engagement Questionnaire (Serholt et al., 2014) and (2) a Self-Assessment Manakin (Lang and Bradley, 2005; 1994). The questionnaire was provided to the participants after interacting with the robot, and it was intended to measure their immediate reaction to the task. The statements were read out loud by the researcher, and the participants indicated whether they agreed (1 = completely disagree to 5 = completely agree), which was noted down by the researcher. In Tables 1 and 2, the questionnaire is shown; the corresponding Dutch version is presented in Appendix H.

The Task and Tutor Engagement Questionnaire was based on research by McGregor and Elliot (2002) and has been adjusted by Serholt et al. (2014) to assess how engaging the activity and interaction with the robot is perceived from the children's perspectives. The questionnaire, as shown in Table 1, highlights engagement, as well as the enjoyment and level of interaction with the robot instructor (Ryan & Patrick, 2001). It also reveals whether participants want to continue engaging with the robot. According to Serholt et al. (2014), this aims to examine the degree, at a minimal level, to which a socio-emotional bond had started to form between the robot and the student. In this study, 'instructor' was replaced by the word 'robot' and translated to Dutch (Appendix H).

Additionally, one extra question was added to the questionnaire about the robot's instructions. The robot did not adapt the tempo of the explanation per participant, so the instructions could be perceived too easy (and too slow) or too hard (and too fast). Therefore, participants were asked about the rapidity of robot's explanation; 'I found the robot's instructions too fast'.

To identify how the participants perceived the robot's gestures, three questions were asked specifically about the gestures (see Table 2). One open question was asked, after which two Likert-scale questions were presented. The statements were read out loud by the researcher, and the participants indicated whether they agreed (1=completely disagree to 5=completely agree), which was noted down by the researcher.

Table 1: Task and Tutor Engagement Questionnaire (Serholt et al., 2014)

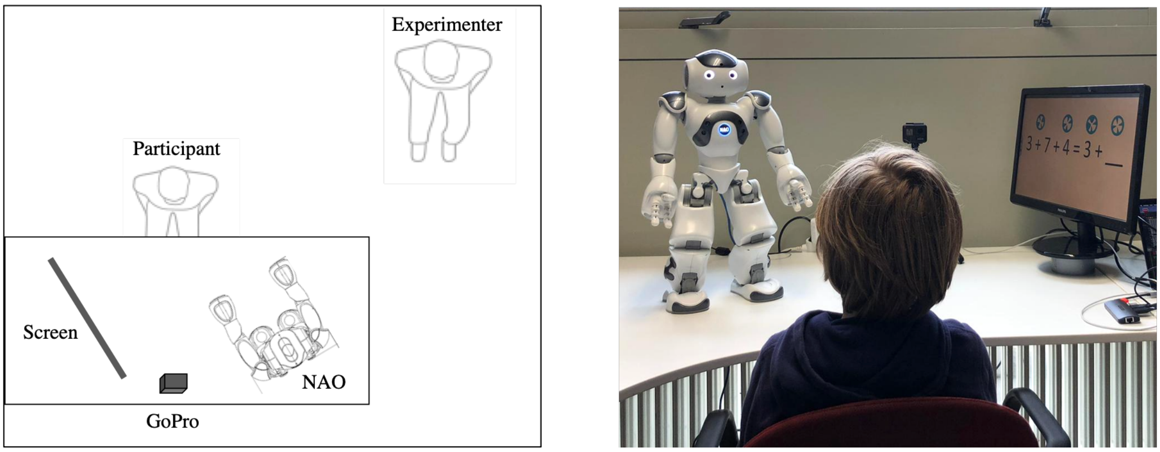
Task Engagement	Social Engagement
I enjoyed this activity	I would like to do another activity with this robot
I found this activity hard	I was wanted to show the robot I was doing a good job
I would like to continue with this activity	I wanted to keep practicing with the robot
It was important for me to do a good job	I found the robot helpful
I found this activity easy to understand	I found it easy to follow instructions presented to me
I found this activity boring	I found the robot boring

Table 2: Gesture-specific questions in the questionnaire

Question Type	Type
What did you think of this robot's gestures?	Open
I found the robot's gestures helpful	Likert-scale
I liked the robot's gestures	Likert-scale

Finally, the Self-Assessment Manikin (SAM) was implemented to elicit participants' affective responses. Lang and Bradley (2005; 1994) developed this instrument to measure affective responses on three dimensions: valence, arousal, and dominance. The pictographic format of this tool makes it accessible for participants with low literacy skills, including children. The first row of SAM presents the range of emotions in the pleasure dimension, with faces that go from a frown to a smile. In the second row, the arousal dimension is depicted, going from sleepy to excited. Finally, the last row shows feelings of dominance, where the biggest figure represents an in control and dominant being, as opposed to the smallest figure, that represents a submissive and controlled one (Hayashi et al., 2016). In this research, the pictographics were shown and explained to the participants, and they appointed their feelings themselves.

2) Head direction estimation: Besides the questionnaire, affective outcomes were also measured by estimating participant's head directions of video recordings. It is argued that gaze expresses how participants attend to varying stimuli, how engaging these stimuli are found, and what amount of attention is assigned to them (Argyle & Cook, 1976). According to Tanaka (2014), head direction and gaze are measures to interpret the level of engagement and interest towards the subject within an activity and interaction with others. Kennedy et al. (2015) suggest that gaze can be a reflection of a child's attention. In the present study, participants were video recorded using the GoPro HERO 7+ Black edition during the child-robot interaction. The GoPro captured videos with 1920x1080 resolution and 30 frames per second. Through OpenFace, a tool capable of facial landmark detection and head pose estimation with available source code (Baltrušaitis, 2018), head locations and rotations were extracted per frame by post-processing



(a) **(b)**

Figure 2: Experimental setup of the one-on-one interaction with participant, screen, the robot, and the video camera with a) the topview and b) the front view

the video recordings. By analysing the extracted data in Python, it was possible to acquire a global indication of whether the participant's head was directed towards NAO, the computer screen, or elsewhere.

D. Procedure

Prior to the experiment, an information letter was read to the children by the experimenter so the participants were prepared for what they could expect (Appendix G). The experimenter was always present during the sessions and seated on the left or right rear side of the children as shown in Figure 2. The children themselves sat behind a table facing the robot, the screen, and a GoPro. During the tutoring session, the robot was connected to the schools' internet with an ethernet cable.

The experiment started with a pre-test that was filled out by the participants. After that, the robot started the first explanation part which was followed by a mid-test that the participants filled out on paper. Subsequently, the robot continued with the second part of the explanation and afterwards the participants made the post-test. When the participants were filling out the tests on paper, NAO was sitting on the desk in the crouched position. At the end of the experiment, the children participated in a questionnaire that was read aloud by the experimenter. Each step is further examined below and an overview of the in total 30-minutes-lasting procedure is shown in Figure 3.

Pre-test. To measure participant's pre-existing knowledge of mathematical equivalence, the children were asked to complete the pre-test on paper, as shown in Appendix I. The experimenter timed the task duration and was available to assist but did not typically give assistance or feedback and provided minimal response when children do not know how to solve the problems (e.g., "You can skip the problems or guess if you do not know the answer.")

Tutoring sessions and mid-test. After the pre-test, children took part in a one-to-one interaction with NAO that was designed to explain the mathematical equivalence problems, either with supporting, random, or no gestures. NAO explained four problems, two of which were Form A and the other two were Form B. During the explanations, mathematical equivalence problems were shown at the screen at which the robot looked and pointed to. The experimenter sat down at the chair during the robot's explanations and did not say anything. After the explanation of the first two problems by NAO (Form A), the experimenter handed over a mid-test with mathematical equivalence problems on paper in the same format as the pre-test that the children solved (Appendix J). Children did not receive feedback from NAO or the experimenter on whether their answers were correct. When the children were finished with the mid-test, NAO explained the last two problems (Form B). During the robot's explanations of Form A and B, the children were captured by the GoPro to track their head directions for analysis.

Post-test. Subsequently, children were handed over a post-test by the experimenter and asked to complete this (Appendix K). Similar to the pre- and mid-test, the task was timed, and no feedback was given to the children when solving the problems on this test.

Questionnaire. At the end of the experiment, the experimenter sat down next to the children and read the questions from the questionnaire out loud from a paper and wrote down the children's answers (Section IV-C1 and Appendix H). The outcomes were inserted in Excel for further analysis of the affective learning outcomes.

Delayed post-test. One week after the experiment, the experimenter returned to the school to test the participants for retention with a delayed post-test. The children individually completed the delayed post-test and were timed during this task (Appendix L).

E. Analysis

To evaluate the learning outcomes, participant information and tests-scores of the pre-, mid-, post-, and delayed post-test scores, and durations of task-completing were derived and stored in one Excel file with each participant on

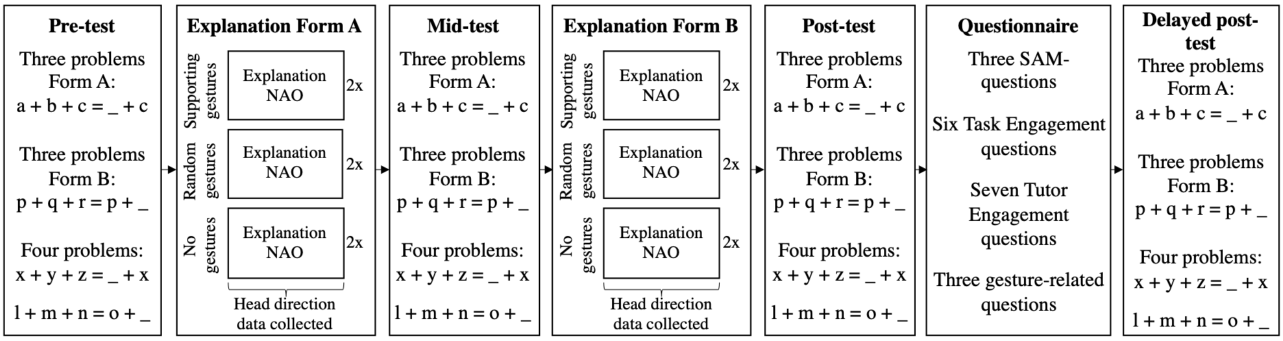


Figure 3: Diagram of the 30-minute enduring tutoring session procedure including the in order presented pre-test, explanation of Form A, mid-test, explanation of Form B, post-test, and questionnaire, and the delayed post-test one week after the session

a row. Likewise, ranked scores of questionnaire questions were stored in the data file, as well as head directions per explanation part (Form A and Form B) and head directions per specific explanatory section. To estimate the head directions, post-processing of the recorded videos was performed with OpenFace, and subsequently with python frameworks which are examined in next section.

1) OpenFace and framework to estimate head direction: By postprocessing the videos in OpenFace, the location of the head with respect to the camera in millimetres (T_x, T_y, and T_z) and the rotation of the head in radians with the camera being located at the origin (R_x, R_y, and R_z) were extracted per frame, with 30 frames per second. Positive T_x is on the right side of the camera, a positive T_y is downwards from the camera, and a positive T_z is away from the camera. The head rotations include pitch (R_x), yaw (R_y), and roll (R_z), and are in radians around the X, Y, and Z axes with the convention $R = R_x * R_y * R_z$, left-handed positive sign (Tadas Baltrušaitis, 2018). This indicates that the head pitch (R_x) is positive when the participant's head is nodding down and negative for the participant's head when nodding up. The yaw (R_y) is positive when the participant's head is rotating to his or her right and negative when rotating left (Figure 27 in Appendix M).

The setup of the experiment and the coordinates of the robot (X_R, Y_R, Z_R), the screen (X_S, Y_S, Z_S), and the participant (X_C, Y_C, Z_C), with the camera the origin (0, 0, 0), are shown in Figure 4. The robot- and screen coordinates were manually measured as minimum and maximum outer coordinates from the camera. Since the robot was moving during the explanation, X_{R1} and Z_{R1} were ascribed to the outer left landmark of the robot, whereas X_{R1} and Z_{R1} were ascribed to the outer right landmark of the robot. The coordinates Y_{S1} and Y_{R1} represent the top of the screen and the top of the robot (with y being negative upwards from the origin), respectively, and Y_{S1} and Y_{R1} represent the bottom of the screen and the bottom of the robot, respectively.

With the coordinates of the robot and the screen, and the location and rotation of the participant's head, it was estimated whether the participant's head is directed towards the robot, the computer screen or elsewhere. Through a framework in Python the head direction was estimated per frame and outcomes were saved in Excel for analysis. This framework and the formulas to derive head direction distances on the coordinate system with respect to the camera-origin in Figure 4 are examined in Appendix M.

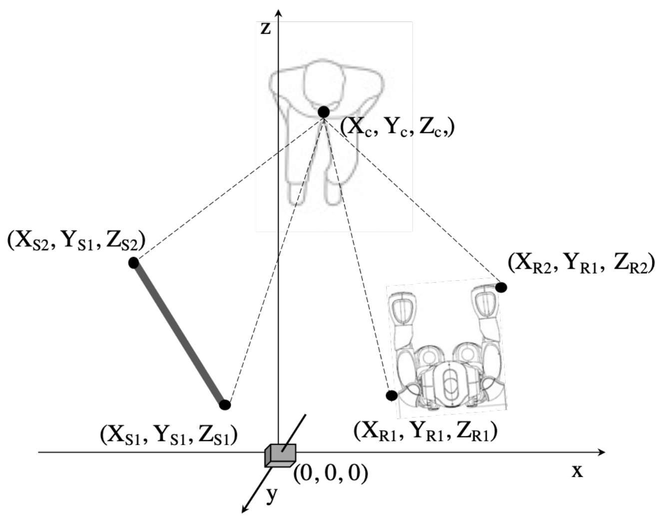


Figure 4: Mathematical equivalence problem with Naomarks 68, 119, 107, 85 from left to right, respectively, placed on top of the sum

2) Statistical analysis: As mentioned in Section IV-A, a total of 78 children participated in the experiments. One participant did not continue with the experiment, so data of a total of 77 participants were analysed and are presented in Chapter V. As one participant was not present during the delayed post-test, the test scores of 76 participants were considered for the analysis of the cognitive learning outcomes. Two participants were not recorded during the first explanation part (Form A) and four participants were not recorded during the second explanation part (Form B), so the head directions of 75 participants were considered for the analysis of the attention towards the robot, screen, or elsewhere during Form A. The head directions of 71 participants were considered for the analysis of the attention during Form B. Using IBM SPPS Statistics version 27, General Linear Models were used to evaluate participant's cognitive and affective learning outcomes and perform statistical analyses, that were validated through statcheck.io, which will be further examined in the next chapter.

Table 3: Demographic information of study participants, including condition, mean age, gender, school (W= De Wijzer, Pa = De Paasbergschool, Pi = De Pieter Brueghelschool), pre-test scores and number of participants per condition

Condition	N	Mean age (years)	Male/Female	School	Pre-test score
Supportive gestures	27	M_{age}= 8.3, SD = 0.45	10 M: 17 F	6 W: 8 Pa: 13 Pi	M_{score}= 2.77, SD = 3.47
No gestures	25	M_{age}= 8.4, SD = 0.56	10 M: 15 F	5 W: 9 Pa: 11Pi	M_{score}= 2.20, SD = 3.53
Random gestures	25	M_{age}= 8.3, SD = 0.54	15 M: 10 F	5 W: 8 Pa: 12 Pi	M_{score}= 2.20, SD = 3.80

V. RESULTS

The results of the experiments are represented in this chapter to answer the research question and test the hypotheses. First, the cognitive learning outcomes will be analysed, after which the affective learning outcomes will be examined, consisting of an analysis of the questionnaire and of the head direction estimation. The demographic information of the 77 children in the participant group is shown in Table 3. Univariate General Linear Model ANOVA with pre-test scores as dependent variable and condition as an independent fixed factor, showed no significant differences between the scores of the pre-tests, $F(2,74) = .228, p = .797, \eta_p^2 = .006$, indicating similar pre-test scores for all three conditions.

A. *Cognitive learning outcomes*

In this section, the results of the cognitive learning outcomes are examined. Table 4 and Figure 5 show the mean scores on the four tests (pre-, mid-, post-, and delayed post-test) per condition, indicating a similar increase in mathematical equivalence knowledge over time between conditions.

Table 4: Means and standard deviations of pre-, mid-, post-, and delayed post-test scores, and number of participants per condition

	Condition	Mean	Std. Dev.	N
Pre-test	Supportive gestures	2.77	3.536	26
	No gestures	2.20	3.488	25
	Random gestures	2.20	3.797	25
Mid-test	Supportive gestures	6.23	3.592	26
	No gestures	6.48	3.417	25
	Random gestures	6.24	3.756	25
Post-test	Supportive gestures	8.19	3.150	26
	No gestures	7.76	3.562	25
	Random gestures	7.16	3.760	25
Delayed post-test	Supportive gestures	7.65	3.566	26
	No gestures	6.96	3.646	25
	Random gestures	6.76	4.075	25

A General Linear Repeated Measures ANOVA was used to evaluate the participant's cognitive learning outcomes with test-scores (pre-test, mid-test, post-test, and delayed post-test) as within-subjects dependent variable, and condition as a between-subjects independent factor.

The analysis showed a significant effect of test, $F(3,219) = 77.992, p < .001, \eta_p^2 = .517$, which indicates that the participants learned about mathematical equivalence from their interactions with the robot regardless of the robot's gesturing condition. Multiple comparisons using Bonferroni adjustments, for which the p-value of the least significant differences is multiplied by the number of tests, revealed a significant difference between the pre-test and mid-test, $M_{dif} = 3.92, p < .001$, between the pre-test and post-test, $M_{dif} = 5.31, p < .001$, and between the pre-test and delayed post-test, $M_{dif} = 4.73, p < .001$. Additionally, a significant difference was also found between the mid- and post-test, $M_{dif} = 1.38, p = .001$. No significant differences were shown between the mid- and the delayed post-test, $M_{dif} = .80, p = .208$, and between the post- and delayed post-test, $M_{dif} = -.58, p = .235$.

1) Between-subject outcomes: No main effect was found of gesturing condition based on the linearly independent pairwise comparisons among the estimated marginal means of the tests, $F(2,73) = 0.28, p = .756, \eta_p^2 = .008$, which reveals that the robot's use of gestures – either supportive gestures, no gestures, or random gestures – did not significantly affect the cognitive learning outcomes. Additionally, Univariate General Linear Model ANOVA with scores as dependent variable and condition as independent fixed factor, also showed no significant differences between the scores of the mid-tests, $F(2,74) = .028, p = .972, \eta_p^2 = .001$, nor between the scores of the post-test, $F(2,74) = .649, p = .526, \eta_p^2 = .017$, and not for the scores of the delayed post-test, $F(2,73) = .398, p = .673, \eta_p^2 = .011$. Therefore, hypothesis H2 is rejected as no effect was found of gesturing condition on children's cognitive learning outcomes.

2) Within-subject outcomes: One-way repeated measures ANOVA with tests as within-subject variables showed that for the gesture condition, pairwise Bonferroni adjustments revealed significant differences between the pre- and the mid-test scores, $M_{dif} = 3.46, p < .001$, pre- and the post-test scores, $M_{dif} = 5.42, p < .001$, and between the pre- and the delayed-test scores, $M_{dif} = 4.89, p < .001$. For the no gesture condition, significant differences were found between the pre- and the mid-test scores, $M_{dif} = 4.28, p < .001$, pre- and the post-test scores, $M_{dif} = 5.56, p < .001$, and between the pre- and the delayed-test scores, $M_{dif} = 4.76, p < .001$. Finally, for the random gesture condition, differences were also significant between the pre- and the mid-test scores, $M_{dif} = 4.04, p < .001$, pre- and the post-test scores, $M_{dif} = 4.96, p < .001$, and between the pre- and the delayed-test scores, $M_{dif} = 4.56, p < .001$. An overview of the mean differences between test-scores per condition are presented in tables 9-11 in Appendix N. These effects are shown in Figure 5 and confirm hypothesis H1; that children learned mathematical equivalence from a social robot by showing significant higher scores on mid-, post-, and delayed post-tests than on pre-tests.

3) Duration: The duration to complete the tests was timed and a General Linear Repeated Measures ANOVA was used to evaluate the participants' durations to complete the tasks (pre-test, mid-test, post-test, and delayed post-

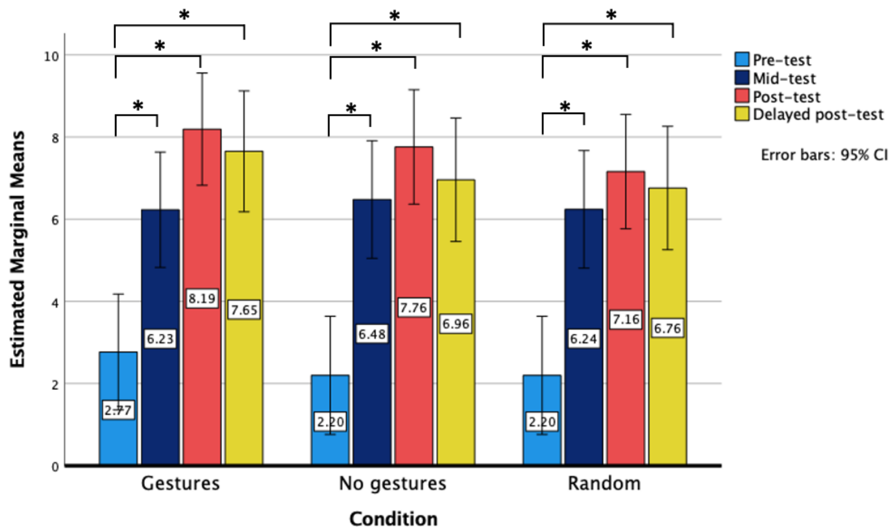


Figure 5: Mean pre-, mid-, post-, and delayed post-test scores and Confidence Intervals that are 95.0% as a function of condition (* $p < .001$)

test) as within-subjects dependent variable, and condition as between-subjects independent factor. The analysis showed a significant effect of test on duration of task, $F(3, 219) = 17.603, p < .001, \eta_p^2 = .194$. Pairwise Bonferroni adjustments revealed significant differences between duration to complete the pre- and the post-test, $M_{dif} = 26.48, p = .024$, demonstrating that the post-test was finished 26.48 seconds faster than the pre-test, and between the pre- and delayed post-test, $M_{dif} = 53.77, p < .001$, indicating that the delayed post-test was completed 53.77 seconds faster than the pre-test. Furthermore, a main difference was found between the duration to complete the mid-test and the delayed post-test, with the delayed post-test being completed $M_{dif} = 40.86$ seconds, $p < .001$, faster than the mid-test (Figure 29 in Appendix O).

No main effects were found on condition based on the linearly independent pairwise comparisons among the estimated marginal means of the durations, $F(2, 73) = .871, p = .423, \eta_p^2 = .023$, indicating that the robot's use of gestures – either supportive gestures, no gestures, or random gestures – did not affect the duration to complete the pre-, mid-, post-, or delayed post-test.

B. Affective learning outcomes from Questionnaire

1) Questionnaire Self-Assessment Manakin: To analyse the Self-Assessment Manakin for the valence, arousal, and dominance dimension, a Univariate General Linear Model procedure was carried out. In Figure 6, the means, and upper and lower bounds of the 95% Confidence Intervals are shown for each dimension per gesturing condition. From the figure, it can be seen that overall, the participants were positive, felt relaxed, and in charge during the interaction. The correlation matrix with the Pearson correlation coefficient between the dimensions is shown in Table 12 in Appendix P and indicates that the dimensions have no significant correlations. The One-way ANOVA with condition as independent fixed factor and dimension as dependent variable revealed that no significant effect of condition was found on valence, $F(2, 74) = .659, p = .521$, on arousal, $F(2, 74) = .091, p = .913$, and on dominance, $F(2, 74) = .050, p = .952$.

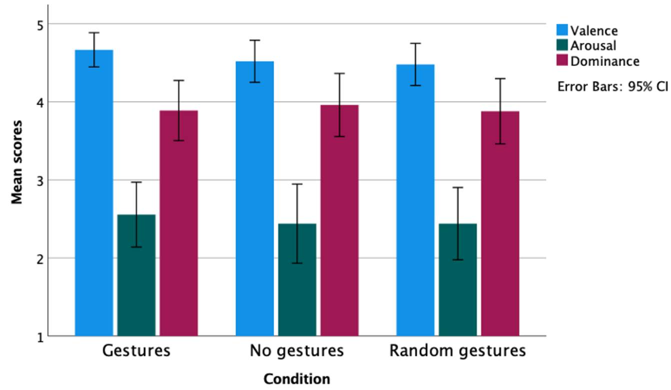


Figure 6: Mean Self-Assessment Manakin scores (5 = Strongly agree, 4 = Agree, 3 = Neither agree nor disagree, 2 = Disagree, 1 = Strongly disagree) as function of condition

2) Questionnaire Tutor and Task Engagement, Gesture specific questions: Figures 7 and 8 visualize the second part of the questionnaire, the task engagement questionnaire, the tutor engagement questionnaire, and the gesture-specific questions. A Multivariate General Linear Model ANOVA with questions as dependent variables and condition as independent fixed factor was executed. Outcomes showed that for the task and tutor engagement questionnaire, no significant differences were found between the conditions. For one gesture-specific question; 'I found the robot's gestures helpful', significant differences were found between the conditions, $F(2, 74) = 8.396, p < .001, \eta_p^2 = .185$. Participants in the gesture condition (M = 4.15, SD = 1.03) and in the no gesture condition (M = 3.68, SD = 1.07) rated this question significantly higher than participants in the random condition (M = 2.88, SD = 1.27). Adjustment for multiple comparisons using Bonferroni showed a significant difference between gesture and random gesture condition, $M_{dif} = 1.27, p < .001$, and between the no

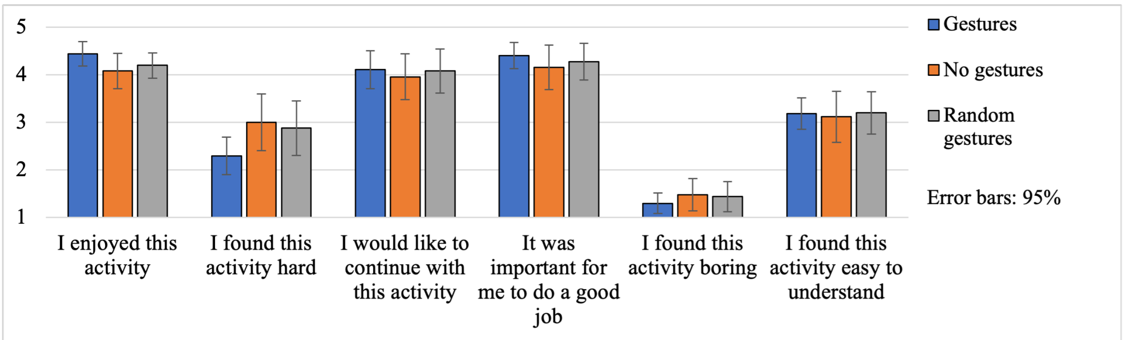


Figure 7: Mean Task Engagement scores (5 = Strongly agree, 4 = Agree, 3 = Neither agree nor disagree, 2 = Disagree, 1 = Strongly disagree) as a function of experimental condition ($*p < 0.05$ using Bonferroni adjustments)

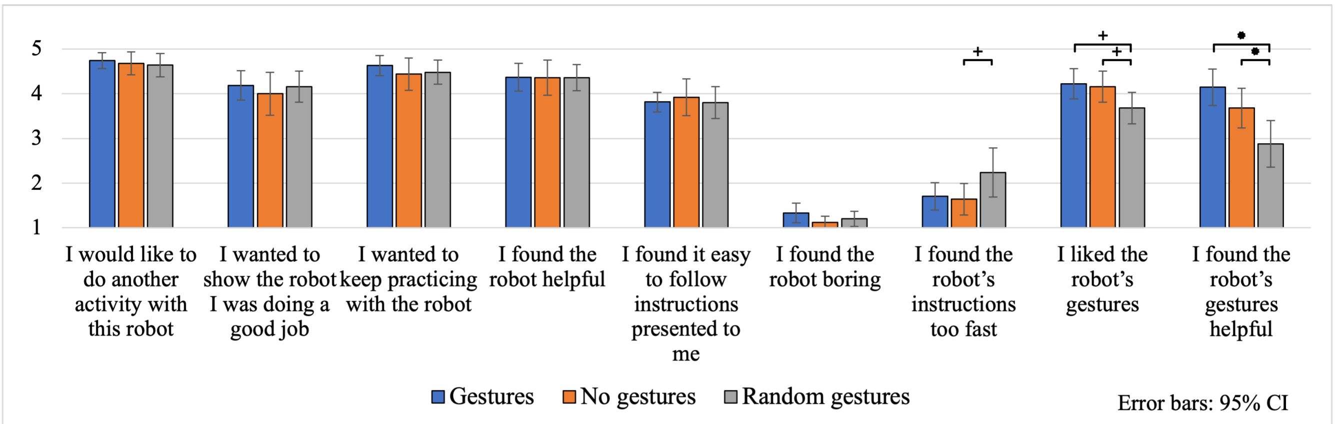


Figure 8: Mean Tutor Engagement and gestures specific scores (5 = Strongly agree, 4 = Agree, 3 = Neither agree nor disagree, 2 = Disagree, 1 = Strongly disagree) as a function of experimental condition (* significance with $p < 0.05$ using Bonferroni adjustments, + significance with $p < 0.05$ using LSD adjustments)

gesture and random condition, $M_{dif} = .800, p = .042$, with a significant mean difference at the 0.05 level. Additionally, the correlation matrices with the Pearson correlation coefficient between the questions are shown in Tables 13 and 14 in Appendix P and indicate which questions have significant correlations.

Furthermore, multiple comparisons adjustment using the Least Significant Difference (equivalent to no adjustment) revealed significant differences for two more questions; 'I found the robot's instructions too fast' and 'I liked the robot's gestures'. Participants in the no gesture condition (M = 1.64, SD = 0.86) rated 'I found the robot's instructions too fast' significantly lower than participants in the random condition (M = 2.24, SD = 1.33), with $M_{dif} = .600, p = .040$. Participants in the gesture condition rated 'I liked the robot's gestures' significantly higher (M = 4.22, SD = .85) than participants in the random condition (M = 3.68, SD = .85), $M_{dif} = .542, p = .024$, as well as the no gesture condition (M = 4.16, SD = .85) compared to the random condition, $M_{dif} = .480, p = .050$.

3) Open gesture question: Besides the closed questions, one open question was asked during the questionnaire; 'What did you think of this robot's gestures?' Children in the supportive gesture condition indicated; *"It is indeed useful that the robot says and portrays 'one side', and then the 'other side"*, and *"So the left side is the same as the right side, he presented that very clearly."* One participant reported that; *"Sometimes the gestures made it a bit difficult to hear."* Another participant commented on the robot's deictic gestures; *"Also handy because you can show where you are, by pointing to the screen."* One participant in the no gesture condition indicated that *"the robot's walking made a lot of noise"*, another participant added *"I was frightened when the robot started walking."* It was also mentioned that *"the teacher also does this, but the teacher also points towards the board"* Children that interacted with the robot with random gestures indicated that the gestures were *"hard to understand"* and *"I did not totally get them."* One participant mentioned that *"I found it nice because when you just stand still, you are less likely to be looked at."* In Table 15 in Appendix Q, the Dutch answers with English translations are presented.

Table 5 represents an overview of certain phrases (e.g., funny – grappig, frightening – spannend, and noisy – lawaaiierig) that were mentioned by participants in the three conditions. It can be noted that participants in the supportive gesturing condition often indicated the gestures as funny (8x), useful (9x) and good (8x), whereas participants in the no gesturing condition referred to the gestures as funny (7x) and frightening (5x) and participants in the random gesturing condition often mentioned nice (5x) and difficult to understand (6x).

Given the outcomes of the SAM, and the task and tutor engagement questionnaire, it can be noted that hypothesis H3 is partially rejected; that participants who

Table 5: Word count of given answers to the open question 'What did you think of the robot's gestures?' with participant ID numbers; condition (G = supportive gestures, N = no gestures, R = random gestures) – number – school (W= De Wijzer, Pa = De Paasbergschool, Pi = De Pieter Brueghelschool)

	Supportive gestures	**No gestures**	**Random gestures**
Funny	G6W, G3Pa, G4Pa, G2Pi, G3Pi, G10Pi, G9Pi, G12Pi	N3W, N4W, N1Pa, N2Pa, N2Pi, N3Pi, N10Pi	R1Pa
Nice	G2W, G5W, G3Pa, G5Pa, G3Pi, G4Pi, G6Pi	N6Pa, N7Pi, N8Pi	R1Pa, R2Pa, R2Pi, R9Pi, R5Pi
Useful	G6W, G5W, G2Pa, G5Pa, G1Pi, G4Pi, G10Pi, G11Pi, G7Pi	N4Pi	
Good	G1W, G1Pa, G2Pa, G7Pa, G8Pa, G4Pi, G11Pi, G8Pi	N6Pi	R7Pa, R11Pi, R3Pi, R8Pi
Proper/plain	G3W, G4W, G3Pi	N2Pa	R5W, R5Pa
Cute			R8Pa
Crazy	G2Pi		R6Pi
Noisy	G3W, G4W, G6Pa, G8Pi	N5W, N9Pa	R4W, R6Pa, R12Pi, R4Pi
Frightening		N3Pa, N4Pa, N8Pa, N7Pi, N11Pi	
Annoying		N5Pa	
Difficult to understand	G2Pi, G3Pi	N1W, N10Pi, N9Pi	R1W, R3W, R2Pa, R3Pa, R7Pa, R12Pi
I did not care		N2W, N7Pa, N1Pi	R1Pi
I do not know	G13Pi		R2W, R4Pa, R10Pi, R7Pi

interact with the supportive gesturing robot show different affective learning outcomes than participants that interact with a social without gestures or with a social robot that demonstrates random gestures. However, participants in the supportive gesturing condition rated the robot's gestures helpfulness significantly higher than the participants in the random condition. Furthermore, answers to open questions revealed that gestures were thought of as good, useful, and nice by participants in the supportive gesturing condition more often than by participants in the random or no gesturing condition, and therefore partially accepting H3.

C. Affective learning outcomes from head directions

In this section, the outcomes of the head directions will be evaluated to evaluate the participants' attention. First, the analysis of the head direction estimation per form will be analysed, followed by the analysis of the head direction per explanatory section. Finally, distributions of head direction distances during the intro (Table 6 in Appendix A) will be presented.

1) Head direction estimation per Form: In Figure 9, an overview of the estimated participants' head directions is shown for Form A and B per condition. The figures represent the average percentages of the time that participants looked towards the robot, towards the screen, or elsewhere. The arrows show the average number of transitions, indicating how often the participants shifted from one direction to another per explanation part given the 30 frames per second. From the figures can be noted that all participants – either supportive gestures, no gestures, or random gestures – looked at the robot around half of the time during both explanation parts. It can also be seen that for both explanatory parts, participants in the supportive gesturing condition switched more between the robot and elsewhere, and less between the screen and elsewhere than the other two conditions.

A Multivariate General Linear Model ANOVA was used to evaluate the participant's head direction with total percentages directed towards the robot, total percentages directed towards the screen, and total percentages directed elsewhere as dependent variables, and condition as between-subjects independent fixed factor. For the first explanation part, the introduction and Form A, no significant differences were found between the three conditions, neither for the total percentage directed towards the robot, $F(2,72) = .065, p = .937, \eta_p^2 = .002$, nor for the total percentage directed towards the screen, $F(2,72) = 1.027, p = .363, \eta_p^2 = .038$. A Multivariate General Linear Model ANOVA with transitions between the robot and elsewhere and transitions between the screen and elsewhere as dependent variables, and condition as between-subjects independent fixed factor showed no significant difference between the conditions, neither for the transition between the robot and elsewhere $F(2,73) = .599, p = .552, \eta_p^2 = .016$, nor for the transitions between the screen and elsewhere $F(2,73) = 1.680, p = .193, \eta_p^2 = .044$.

For the second explanation part, Form B, also no significant differences were found between the three conditions, neither for the total percentage directed towards the robot, $F(2,70) = .278, p = 0.758, \eta_p^2 = .008$, nor for the total percentage directed towards the screen, $F(2,70) = 1.111, p = .335, \eta_p^2 = .031$. Furthermore, no significant differences were found between the conditions for the number transitions between the robot and elsewhere, $F(2,73) = .845, p = .434, \eta_p^2 = .023$ and for the number of transitions between the screen and elsewhere $F(2,73) = .407, p = 667, \eta_p^2 = .011$.

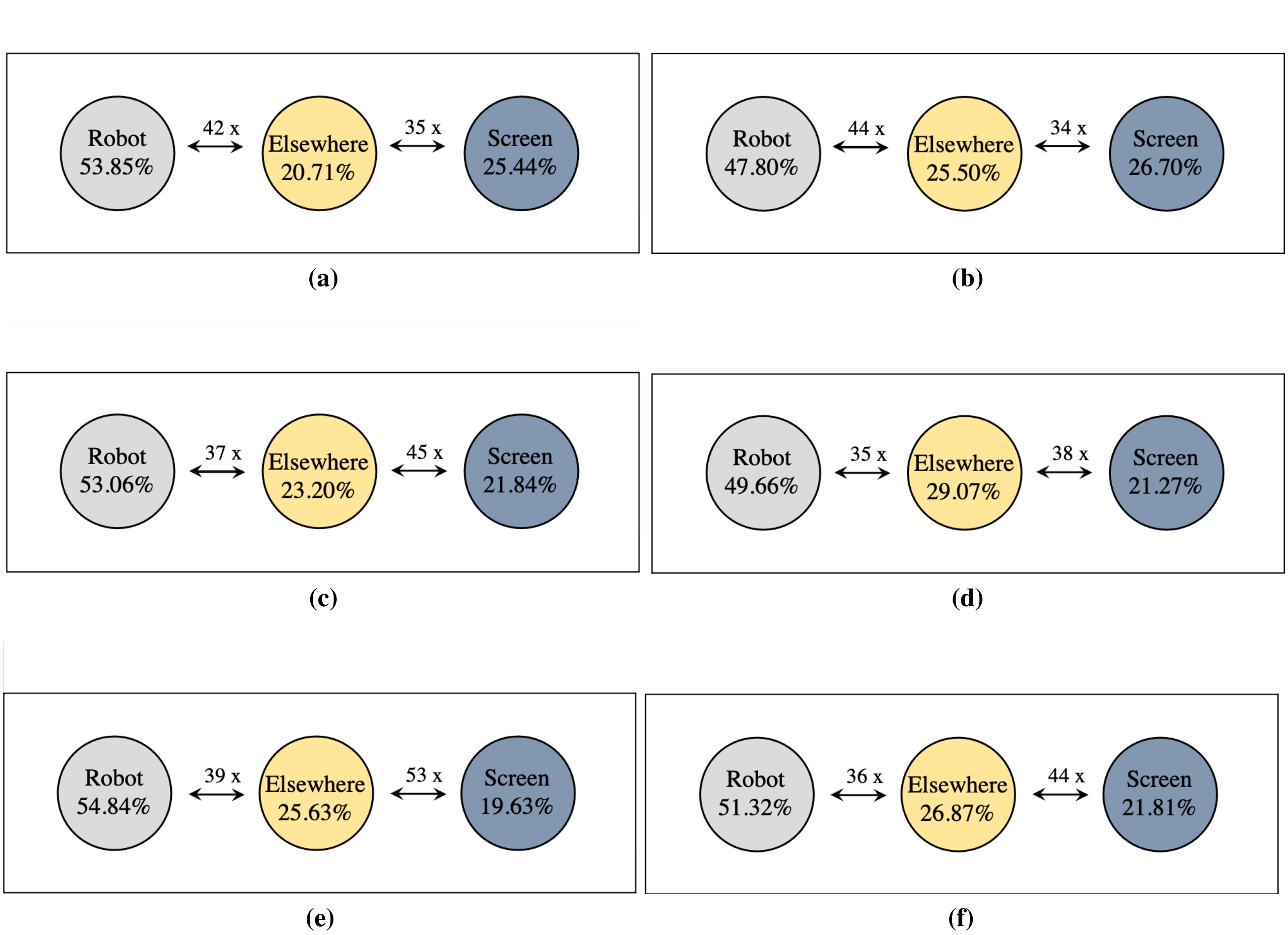


Figure 9: Mean percentages of head direction towards the robot, screen, or elsewhere and mean number of transitions made between the robot – elsewhere and between the screen – elsewhere for the a) supportive gesturing condition in Form A, b) supportive gesturing condition in Form B, c) no gesturing condition in Form A, d) no gesturing condition in Form B, e) random gesturing condition in Form A, f) random gesturing condition in Form B

2) Head direction estimation per explanatory section: The average percentages of the head directions towards the robot, screen, or elsewhere per explanatory section for Form A and Form B, with time in seconds on the horizontal axis are shown in Figure 10. From the figures, it can be noted that all participants – either supportive gestures, no gestures, or random gestures – looked at the robot most often during the iconic gesturing sections, regardless of the condition or Form. As the robot started to look and, for the gesture condition only, point towards the screen, the participants directed their heads more towards the screen or elsewhere. This could indicate that, regardless of gesturing condition, the attention towards the robot is related to the explanatory section.

A Multivariate General Linear Model ANOVA was used to evaluate the participant's head direction per explanatory section with total percentages directed towards the robot, total percentages directed towards the screen, and total percentages directed elsewhere as dependent variables, and experimental condition as between-subjects independent fixed factor. No significant differences were found between three conditions, for the total percentage directed towards the robot, for the total percentage directed towards the screen, and for the total percentage directed elsewhere for any of the explanatory sections. Given the non-significant effect of condition on the head directions, hypothesis H3 is partially rejected as no effect as was found of gesturing condition on children's attention.

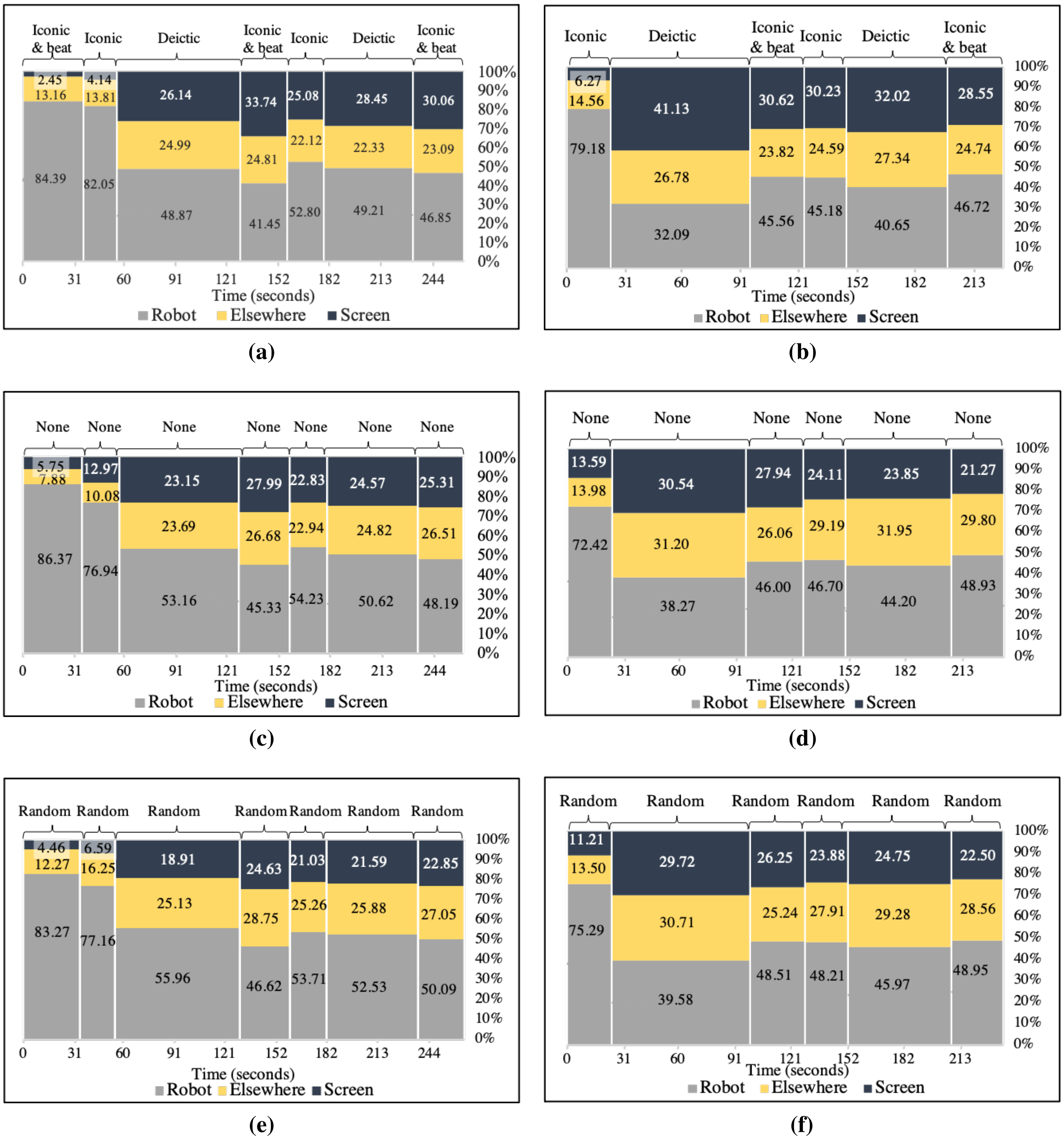


Figure 10: Mean percentages of head direction towards the robot, screen, or elsewhere during specific explanation sections for participants in the a) supportive gesturing condition in Form A, b) supportive gesturing condition in Form B, c) no gesturing condition in Form A, d) no gesturing condition in Form B, e) random gesturing condition in Form A, f) random gesturing condition in Form B

3) Distributions of head direction distances during the intro: In Figure 12, histograms with plotted standard normal distributions show the distribution of the distances in mm between camera-origin and direction of the head, when directed towards the robot, on the x-axis (with respect to the camera) during the intro per frame for all participants in the gesture condition, no gesture condition and random gesture condition respectively. In Figure 11, the distance in mm between camera-origin and direction of the head is illustrated. During the intro, which was part of Form A and endured 35.4 seconds on average, the robot did not turn left or right towards the screen or point towards the screen, and the supportive robot solely performed iconic and beat gestures (Table 6 in Appendix A). In Figure 28 in Appendix R, the distances are plotted over time for each participant per condition.

From Figure 12, it can be seen that the data of all the participants in the no gesture and in the random gesture condition roughly follow a normal distribution, whereas this does not apply for participants in the supportive gesture condition. The histograms may suggest that participants in the no gesture and random condition to a large extent directed their heads towards the centre of the robot, but that participants in the supportive gesture condition also directed their heads towards other components of the robot. This may indicate that different head direction patterns existed between the participants in the supportive gesture condition and the random and no gesture conditions. However, no

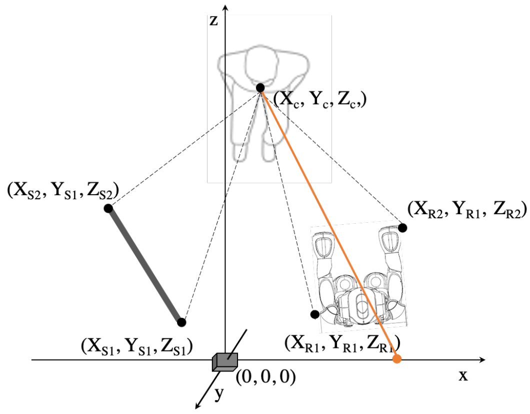
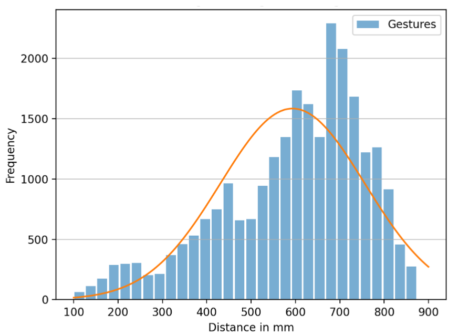
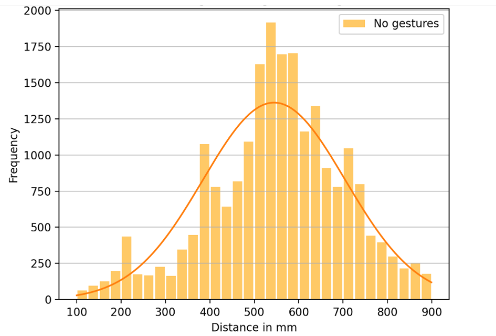


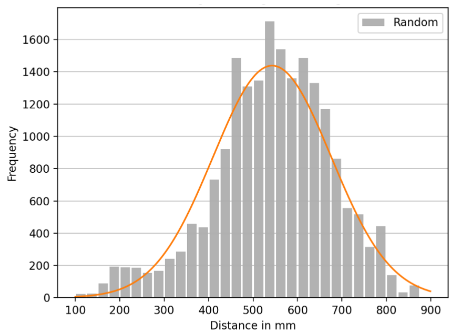
Figure 11: Illustration of distance in mm between camera-origin and direction of the head, when directed towards the robot, on the x-axis (with respect to the camera)



(a)



(b)



(c)

Figure 12: Histograms with plotted standard normal distributions of the distances in mm between camera-origin and direction of the participants' heads, when directed towards the robot, on the x-axis (with respect to the camera) during the intro per frame for a) Supportive gesture condition ($\mu = 593.53, \sigma = 163.20$), b) No gesture condition ($\mu = 544.58, \sigma = 160.56$), c) Random gesture condition ($\mu = 542.85, \sigma = 133.05$)

significant differences were found between the supportive gesturing condition, no gesturing condition, and random gesturing condition, with distance as dependent variable and condition as between-subjects independent fixed factor, $F(2, 72) = .585, p = .560, \eta_p^2 = .016$.

To further explore this, histograms were generated with plotted standard normal distributions that show the distribution of the distances in mm between the bottom of the robot and direction of the head on the y-axis, when directed towards the robot, during the intro per frame (Figure 14). These figures indicate to which height point on the y-axis the participant's head is directed, with the desk on which the robot is placed as 0-point and a positive distance upwards from the desk (Figure 13). In Figure 30 in Appendix R, the distances are plotted over time for each participant per condition. Figure 14 shows a higher mean distance for the no gesture condition than for the supportive and random gesture conditions. Therefore, the participants in the no gesture condition looked on average on a higher component of the robot than the gesturing conditions. This may indicate that different head direction patterns existed between the participants in the no gesture condition compared to the supportive and random gestures conditions. However, a Multivariate General Linear Model ANOVA showed no significant differences between the conditions, with distance as dependent variable and condition as between-subjects independent fixed factor, $F(2, 72) = 2.121, p = .127, \eta_p^2 = .056$.

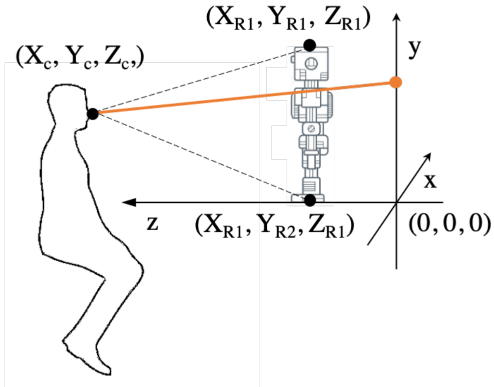
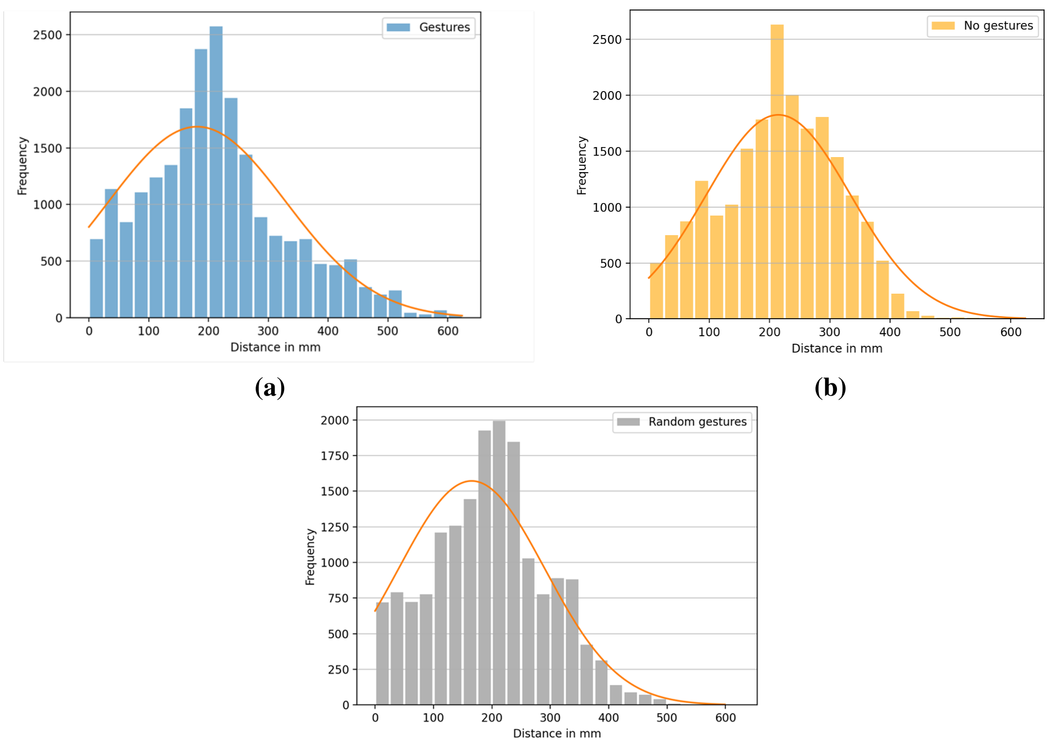


Figure 13: Illustration of distance in mm between bottom of the robot and direction of the head, when directed towards the robot, on the y-axis



(a) **(b)**

(c)

Figure 14: Histograms with plotted standard normal distributions of the distances in mm between the bottom of the robot and direction of the head on the y-axis, when directed towards the robot, during the intro per frame for a) Supportive gesture condition ($\mu = 181.04, \sigma = 148.49$), b) No gesture condition ($\mu = 214.69, \sigma = 119.85$), c) Random gesture condition ($\mu = 165.38, \sigma = 125.53$)

VI. Discussion

In this paper, a large-scale evaluation study is presented that was conducted to investigate to what extent social robots can contribute to mathematical equivalence tutoring with the use of supporting gestures. Three different robot conditions were compared; one with iconic, beat, and deictic gestures, one with no gestures, and one with random gestures. A total of 78 children from the fifth grade of three Dutch primary schools participated in the experiments in which they interacted one-on-one with the social robot. The participants conducted pre-, mid-, post-, and delayed post-tests to measure the cognitive learning outcomes. After the 30-min enduring tutoring session, they answered a questionnaire, including a Self-Assessment Manakin, task and tutor engagement questions, and gesture specific gestures, to measure the affective learning outcomes. To further evaluate the affective learning outcomes, the participants were video recorded during the robot's explanations to analyse the attention through head direction estimations.

To summarise the main findings, evidence was found to support hypothesis H1; that children learned mathematical equivalence from a social robot that demonstrated deictic, beat, and iconic gestures that align with explanations by showing significant higher scores on mid-, post-, and delayed post-tests than on pre-tests. However, no evidence was found to support hypothesis H2; that children who interacted with a social robot that demonstrated supportive gestures showed greater cognitive learning outcomes than children who interacted with the social robot with no gestures, or than children who interacted with the social robot with random gestures. It was found that all participants showed high tutor and task engagement by strongly agreeing to enjoying the task and social robot and wanting to continue with the robot. Finally, no evidence was found to support hypothesis H3; that a significant difference would be revealed between the affective learning outcomes of the participants in the supportive gesturing robot condition and the participants in the no gesturing robot condition and the random gesturing robot condition. However, participants in the supportive gesturing condition rated the robot's gestures helpfulness significantly higher than the participants in the random condition. Furthermore, answers to open questions revealed that gestures were thought of as good, useful, and nice by participants in the supportive gesturing condition more often than by participants in the random or no gesturing condition, and therefore partially accepting H3. Although previous studies with a social robot or avatar in educational context have demonstrated a positive effect of supportive gestures on learning compared to no gestures (De Wit et al., 2018; Conti et al., 2019; Cook et al., 2017), the present study does not confirm this. In the remainder of this section, these findings will be elaborated.

A. Cognitive learning outcomes

In line with the study by Cook et al. (2017), the children learned the concept of mathematical equivalence from a social robot that demonstrated deictic, beat, and iconic gestures that align with explanations. Contradictory, Cook et al. (2017) also found that children who observed the gesturing avatar learned significantly more, and they solved the mathematical equivalence problems more quickly than children who observed a non-gesturing avatar. This is compelling because identical gestures were used for the supportive gesturing robot in the present study as for Cook et al.'s (2017) gesturing avatar. One remark that multiple participants made; "Sometimes the gestures made it a bit difficult to hear" and the suggestion by Kennedy et al., (2015a) that robot's behaviour and movement could lead to more distraction, may suggest that gestures, and their noise, could draw attention away from the content. Furthermore, the effect of the screen was not considered in the present study. From head direction estimations resulted that during the deictic gesturing explanation sections, in which the robot pointed towards the screen, participants in the supportive gesturing condition, as well as participants in the no gesturing and random gesturing condition directed their head towards the screen more than during the iconic gesturing explanation sections. This indicates that all participants directed their head towards the screen in similar quantities, regardless of condition, and may have absorbed information from the sum displayed on the screen.

B. Affective learning outcomes

Analysis of the questionnaire and head direction estimations showed that in the present study, no significant differences were found in tutor engagement, task engagement, or attention between the three conditions which is similar to research by De Wit et al. (2018). This held for all questions in the questionnaire, except for one question; 'I found the robot's gestures helpful', which was rated significantly higher by participants in the supportive gesturing and no gesturing condition than participants in the random gesturing condition and shows that the manipulation was successful. Additionally, children in the supportive gesturing condition noted that the robot's gestures were 'useful' and 'good', whereas participants in the random gesturing condition mentioned that the gestures were 'difficult to understand', indicating that the supportive gestures may have been a useful addition to the explanation.

Contradictory to these findings, De Wit et al. (2020) did find significant differences in engagement between the gesturing conditions and the no gesturing condition. One major difference between the two studies is the number of tutoring sessions. Although the participants in the present study were introduced to the social robot during a classroom introductory lesson, the participants interacted one-on-one with NAO for the first time during the tutoring session and may have been influenced by the 'novelty effect'. According to Minkelen et al. (2020), this entails that the users are initially excited to interact with a new technology, such as a tutor robot, resulting in high engagement. When users become familiarised with the new technology and the novelty threshold is surpassed, the novelty effect wears off, and users' engagement tends to decrease. In the present study, participants in all three conditions showed high task and tutor engagement scores and directed their head towards NAO often, showing high attention towards the robot. However, when there would have been less effect of novelty, differences between the three gesturing conditions could have appeared.

C. On the experimental design

During the development of this project, several design choices were made which have been reported in the paper. Based on previous educational gesture research (Cook et al., 2017; De wit et al., 2020), the designed explanations and presented gestures were equal for the participants in each condition group. This entailed that the robot's explanation speed and difficulty level of the mathematical equivalence problem were identical for each child, regardless of the child's understanding of the concept. It could be an interesting feature to adapt the problem level and speed to the performance level of a student (Schadenberg et al., 2017). Such systems could feature a detailed learner model to profile children and their needs (e.g., knowledge state, engagement level and learning speed) to provide for personalised adaptation for each student.

Moreover, the experimental design was limited by an absence of gaze directions. Previous tests pointed out that the gaze direction was not accurately measured by the OpenFace Model and therefore, the head direction was chosen to measure attention. Post-testing showed that the head directions were estimated accurately by the model, but that participants' heads were sometimes directed elsewhere (e.g., towards the GoPro) while its eyes are directed towards either the robot or the screen. Each of the conditions encountered this matter and it did not differ between conditions, but gaze directions would have provided more accurate attention outcomes. Furthermore, timing the per specific gesture (e.g., point gesture towards the number 8) instead of timing per gesture group (e.g., deictic gestures) could have provided more accurate comparisons between conditions for head directions per specific gesture. Additionally, during the experiments, it was noted that, although the one-on-one interactions with the participant and the robot were held in a separate room, there still were background noises and distractions from children in the school hallways. Future robot designs should consider these real-life scenarios and test real use case to be able to develop social robots that can integrate in daily life and educational context.

VII. Conclusion and recommendations

A study in which a social robot was used to teach mathematics to primary school children is presented in this paper. The aim of the study was to program a social robot that tutors mathematics through verbal explanations supported by deictic, beat, and iconic gestures that align with these explanations, and to analyse its effectiveness on learning outcomes. Results showed that the social robot effectively taught children mathematical equivalence by an increase in test scores, but no differences were found in test scores between participants in the supportive gesturing condition, the no gesturing condition, and the random gesturing condition. Participants showed high task and tutor engagement scores and head direction estimations revealed that participants were attentive towards the robot, but no differences were found between the three conditions. However, the gestures 'helpfulness' was rated significantly higher by participants in the supportive gesture condition than the random gesture condition, suggesting that children felt the gestures mattered. Several design choices were made during development of the experiment (e.g., regarding the robot's gestures and interaction with the screen), which have been documented in this paper.

A. Future research

Arguably, the main contributions of the presented research are the introduction of the social robot as mathematical equivalence tutor that tutors mathematics with iconic, deictic, and beat gestures that align with explanations, and the comparison made between supportive gestures and random gestures in Human-Robot Interaction research as well as in Human-Computer Interaction research. It is believed that more research is needed to further investigate the effectiveness of supportive gestures in educational HRI. The present study consisted of one single tutoring session and did thus not investigate any potential long-term effects that supportive gestures of social robots might have compared to no gestures or random gestures. Besides that, the use of social robot's gestures could be further discovered in other teaching domains. Finally, it would be interesting to compare the learning outcomes of participants that interact with a social robot that uses gestures to support explanations and the learning outcomes of participants that interact with an avatar that uses the exact same gestures (e.g., the avatar created by Cook et al., 2017) to investigate whether there is an effect of the physical presence of a physically embodied robot. There is a need of more large-scale studies towards social robots in an educational context to increase the effectiveness of introducing robots to further benefit the field of education.

References

Alibali, M. W., & Nathan, M. J. (2012). Embodiment in Mathematics Teaching and Learning: Evidence From Learners' and Teachers' Gestures. *The Journal of the Learning Sciences*, 21(2), 247–286. WorldCat.org.

Argyle, M. & Cook., M. (1976). *Gaze and Mutual Gaze*. Cambridge University Press, New York, NY, USA, 1976.

Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L. P. (2018, May). Openface 2.0: Facial behavior analysis toolkit. *13th IEEE International Conference on Automatic Face & GestureRecognition (FG 2018)*. 59-66

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science Robotics*, 3(21). https://doi.org/10.1126/scirobotics.aat5954

Bradley, M.M., Lang, P.J. 1994. Measuring Emotion: The Self-Assessment Manikin and the Semantic Differential. *Journal of Behavioral Therapy and Experimental Psychiatry*, 25, 49-59

Bremner, P., & Leonards, U. (2016). Iconic gestures for robot avatars, recognition, and integration with speech. *Frontiers in psychology*, 7, 183.

Congdon, E. L., Novack, M. A., Brooks, N., Hemani-Lopez, N., O'Keefe, L., & Goldin- Meadow, S. (2017). Better together: Simultaneous presentation of speech and gesture in math instruction supports generalization and retention. *Learning and Instruction*, 50, 65–74. https://doi.org/doi:10.1016/j.learninstruc.2017.03.005

Cook, S. W., Duffy, R. G., & Fenn, K. M. (2013). Consolidation and transfer of learning after observing hand gesture. *Child Development*, 84(6), 1863–1871. https://doi.org/doi:10.1111/cdev.12097

Cook, Susan Wagner, Friedman, H. S., Duggan, K. A., Cui, J., & Popescu, V. (2017). Hand Gesture and Mathematics Learning: Lessons From an Avatar. *Cognitive Science*, 41(2), 518–535.

De Wit, J., Brandse, A., Krahmer, E., & Vogt, P. (2020). Varied human-like gestures for social robots: Investigating the effects on children's engagement and language learning. *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 359– 367. https://dl.acm.org/doi/abs/10.1145/3319502.3374815

De Wit, J., Schodde, T., Willemsen, B., Bergmann, K., de Haas, M., Kopp, S., Krahmer, E., & Vogt, P. (2018). The effect of a robot's gestures and adaptive tutoring on children's acquisition of second language vocabularies. *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction*, 50–58. https://dl.acm.org/doi/abs/10.1145/3171221.3171277 WorldCat.org. https://doi.org/10.1111/cogs.12344

Goldin-Meadow, S., Kim, S., & Singer, M. (1999). What the teacher's hands tell the student's mind about math. *Journal of Educational Psychology*, 91(4), 720–730.

Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychological Science*, 20(3), 267–272.

Hayashi, E. C., Posada, J. E. G., Maike, V. R., & Baranauskas, M. C. C. (2016). Exploring new formats of the Self-Assessment Manikin in the design with children. *Proceedings of the 15th Brazilian Symposium on Human Factors in Computing Systems*, 1-10

Hindriks, K. V., & Liebens, S. (2019). A robot math tutor that gives feedback. *International Conference on Social Robotics*, 601–610.

Holroyd, A., & Rich, C. (2012, March). Using the behavior markup language for human-robot interaction. 2012 7th ACM/IEEE *International Conference on Human-Robot Interaction (HRI)*, 147-148, IEEE.

Kennedy, J., Baxter, P., Senft, E., & Belpaeme, T. (2015). Higher nonverbal immediacy leads to greater learning gains in child-robot tutoring interactions. *International conference on social robotics.*

Kennedy, J., Baxter, P., & Belpaeme, T. (2015a). The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning. 2015 10th *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 67–74.

Kopp, S., Krenn, B., Marsella, S., Marshall, A. N., Pelachaud, C., Pirker, H., & Vilhjálmsson, H. (2006, August). Towards a common framework for multimodal generation: The behavior markup language. *International workshop on intelligent virtual agents*, 205-217, Springer, Berlin, Heidelberg.

Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (2005). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual*. Technical Report A-6. University of Florida, Gainesville, FL.

Le, Q., Huang, J., & Pelachaud, C. (2012). A common gesture and speech production framework for virtual and physical agents. *ACM international conference on multimodal interaction*.

Leyzberg, D., Spaulding, S., Toneva, M., & Scassellati, B. (2012). The physical presence of a robot tutor increases cognitive learning gains. *Proceedings of the annual meeting of the cognitive science society*. 34 (34).

McGregor, H. A., & Elliot., A. J., (2002) Achievement goals as predictors of achievement-relevant processes prior to task engagement. *Journal of Educational Psychology*, 94(2):381

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

Minkelen, P., Gruson, C., van Hees, P., Willems, M., de Wit, J., Aarts, R., Denissen, J., & Vogt, P. (2020). Using self-determination theory in social robots to increase motivation in l2 word learning. *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 369–377. https://dl.acm.org/doi/abs/10.1145/3319502.3374828

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

Nomura, T., Suzuki, T., Kanda, T., & Kato. K., Measurement of negative attitudes toward robots. *Interaction Studies*, 7(3):437–454, 2006.

Perry, M., Berch, D., & Singleton, J.L. (1995). Constructing shared understanding: The role of nonverbal input in learning contexts. *Journal of Contemporary Legal Issues*, 6, 213–236.

Ping, R. M., & Goldin-Meadow, S. (2008). Hands in the air: Using ungrounded iconic gestures to teach children conservation of quantity. *Developmental Psychology*, 44, 1277-1287. doi:10.1037/0012-1649.44.5.1277

Ryan, A. M., & Patrick, H. (2001). The classroom social environment and changes in adolescents' motivation and engagement during middle school. *American Educational Research Journal*, 38(2):437–460

Robotics, A. (2015). H25-Links—Aldebaran 2.1. 4.13 documentation.

Schadenberg, B. R., Neerincx, M. A., Cnossen, F., & Looije, R. (2017). Personalising game difficulty to keep children motivated to play with a social robot: A Bayesian approach. *Cognitive systems research*, 43, 222-231.

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

Singer, M. A., & Goldin-Meadow, S. (2005). Children Learn When Their Teacher's Gestures and Speech Differ. *Psychological Science*, 16(2), 85–89. WorldCat.org. https://doi.org/10.1111/j.0956-7976.2005.00786.x

Tanaka, F. (2014). Robotics for supporting childhood education. In Sankai, Y., Suzuki, K., and Hasegawa, Y. (2014). *Cybernics: Fusion of human, machine and information systems*, 185–195. Springer, Japan,

Valenzeno, L., Alibali, M., & Klatzky, R. (2003). Teachers' gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology*, 28, 187–204.

Vogt, P., van den Berghe, R., de Haas, M., Hoffman, L., Kanero, J., Mamus, E., Montanier, J.-M., Oranç, C., Oudgenoeg-Paz, O., Garcıa, D. H., Et al. (2019). Second language tutoring using social robots: A large-scale study. *14th ACM/IEEE International Conference on Human- Robot Interaction (HRI)*, 497–505.

Wakefield, E., Novack, M. A., Congdon, E. L., Franconeri, S., & Goldin-Meadow, S. (2018). Gesture helps learners learn, but not merely by guiding their visual attention. *Developmental Science*, 21(6). WorldCat.org. https://doi.org/10.1111/desc.12664

Appendix A - Beat, deictic, and iconic gestures and explanations per explanation section

Table 6: Script with gestures and explanations for the intro (only part of Problem script Form A)

Gesture	Category	Explanation section	Reference
Right-handed wave	Iconic gesture	'[Hi name].'	
Two handed outward-focused gesture	Beat gesture	'Today we are going to learn about [the equal sign].'	Cook et al. (2017)
Two handed outward-focused gesture	Beat gesture	'[The equal sign is a symbol] that tells us about the two things on both sides of it.'	
Left-handed outward-focused gesture	Beat gesture	'[Whatever is on one side of the equal sign] needs to be the same amount..'	Cook et al. (2017)
Right-handed outward-focused gesture	Beat gesture	'..[as whatever is on the other side of the equal sign].'	
Right-handed outward-focused gesture	Beat gesture	'Okay name, [now let's see how this works].'	

Table 7: Poblem script Form A (repeated for two problems - example problem: 8 + 6 + 2 = _ + 2)

Gesture	Category	Explanation	Reference
Two handed outward-focused gesture	Beat gesture	'Remember, [the equal sign means]..	
Right-handed balance gesture	Iconic gesture	'..[that the total amount on the left side must be the same]..'	
Left-handed balance gesture	Iconic gesture	'..[as the total amount on the right side].'	Cook et al. (2017)
One-handed point to blank	Deictic gesture	'This will help us figure out [ ] what goes inside the blank.'	
Right-Handed sweep right side	Iconic gesture	'[One side]..'	Goldin-Meadow,
Left-Handed sweep left side	Iconic gesture	'..[needs to equal the other side].'	Kim, & Singer,
Two-handed balance gesture	Iconic gesture	'You know you have the right answer, [when the two sides are the same amount].'	
		'Let's figure out how to do this'	
One-handed point to 8 and 6 on left side	Deictic gesture	'[ ] Eight plus six equals fourteen.'	
One-handed point to 2 on left side	Deictic gesture	'[ ] Fourteen plus two is sixteen'	
One-handed point to blank	Deictic gesture	'[And what number plus two] equals sixteen?'	
One-handed point to blank	Deictic gesture	'[Fourteen plus two equals sixteen].'	
Two-handed balance gesture	Iconic gesture	'If you look at both sides, [they equal the same amount].'	Cook et al. (2017)
Left-handed outward-focused gesture	Beat gesture	'[Which is sixteen],..'	
Right-handed outward-focused gesture	Beat gesture	'..[and sixteen].'	

		'So, one side equals the other side.'	Goldin-Meadow, Kim, & Singer, (1999), Albibaba & Nathan (2012), Cook et al. (2013), Cook et al. (2017), Wakefiel et al. (2018)
Two handed outward-focused gesture	Beat gesture	'Okay name, [let's look at the next sum].'	

Table 8: Problem script Form B (repeated for two problems – example problem: 5 + 3 + 9 = 5 + _)

Gesture	Category	Explanation	Reference
Two handed outward-focused gesture	Beat gesture	'Remember, [the equal sign means]..'	Cook et al. (2017)
Left-handed balance gesture	Iconic gesture	'..[that the total amount on the left side must be the same]..'	
Right-handed balance gesture	Iconic gesture	'..[as the total amount on the right side].'	
Right-Handed sweep right side	Iconic gesture	'[So, one side]..'	Goldin-Meadow, Kim, & Singer, (1999), Albibaba & Nathan (2012), Cook et al. (2013), Cook et al. (2017)
Left-Handed sweep left side	Iconic gesture	'..[needs to equal the other side].'	
		'Let's figure out how to do this'	
One-handed point to 5 and 3 on left side	Deictic gesture	'[ ] Five plus three is eight'	
One-handed point to 9 on left side	Deictic gesture	'[ ] Eight plus nine equals seventeen'	
One-handed point to blank	Deictic gesture	'[And five plus what number] equals seventeen?	
One-handed point to blank	Deictic gesture	'[Twelve plus five equals seventeen]'	
Two-handed balance gesture	Iconic gesture	'If you look at both sides, [they equal the same amount]..'	
Left-handed outward-focused gesture	Beat gesture	'..[which is fourteen]..'	
Right-handed outward-focused gesture	Beat gesture	'..[and fourteen].'	
		'So, one side equals the other side.'	Goldin-Meadow, Kim, & Singer, (1999), Albibaba & Nathan (2012), Cook et al. (2013), Cook et al. (2017), Wakefield et al. (2018)
Two handed outward-focused gesture	Beat gesture	'Okay name, [now it's your turn]!'	Cook et al. (2017)

Appendix B – Example of Petri nets used to coordinate the robot's speech and gestures

An example of using Petri nets to coordinate timing of speech and gestures is shown in this section, where the robot presents a beat gesture while saying; ['about the equal sign'], and no movements are generated while saying; 'Today we are going to learn'.

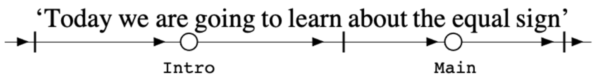


Figure 15: Synchronisation points and transitions based on Petri Nets for explanatory speech

Each gesture provided by Cook et al. (2017) can be partitioned into synchronisation points that follow phases which can consist of a start, ready, stroke-start, stroke, stroke-end, relax, and an end stage (Kopp et al., 2006). For the two-handed outward-focused beat gesture, an example is shown in Figure 16.

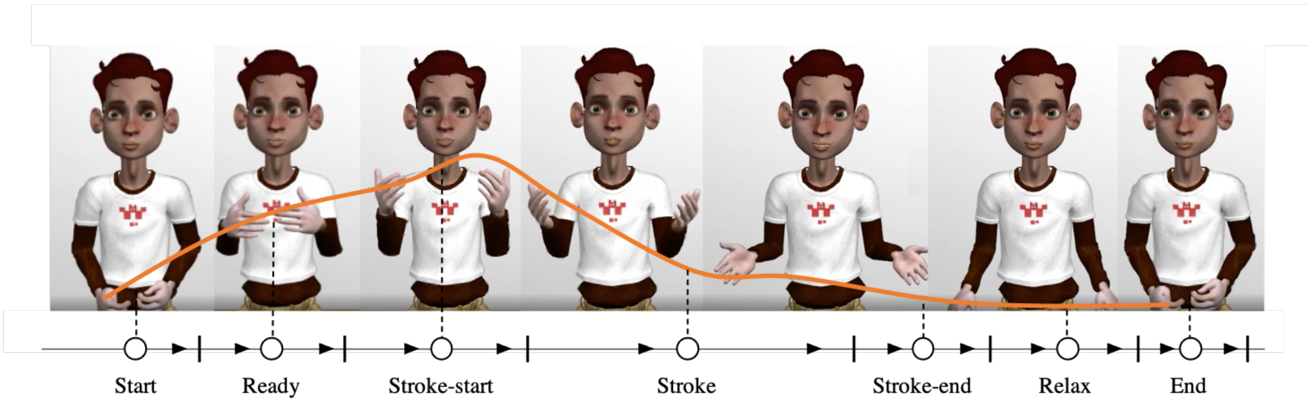


Figure 16: Synchronisation points and transitions based on Petri Nets for an explanatory beat gesture (footage by Cook et al., 2017)

The explanatory speech and beat gesture in Figures 15 and 16 can be aligned in a Petri net, as shown in Figure 17. The synchronisation point 'main' is now overlapping with the gesture phases 'ready' until 'relax'. To assure the synchronisation transitions, the function `yield` is called in the python script. For each explanation section in Tables 6-8 in Appendix A, the iconic and beat gestures were scripted in such Petri nets.

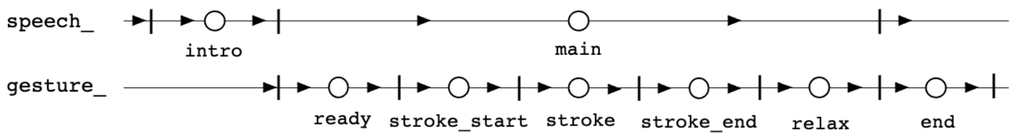


Figure 17: Petri net with synchronisation points and transitions to align speech and gesture

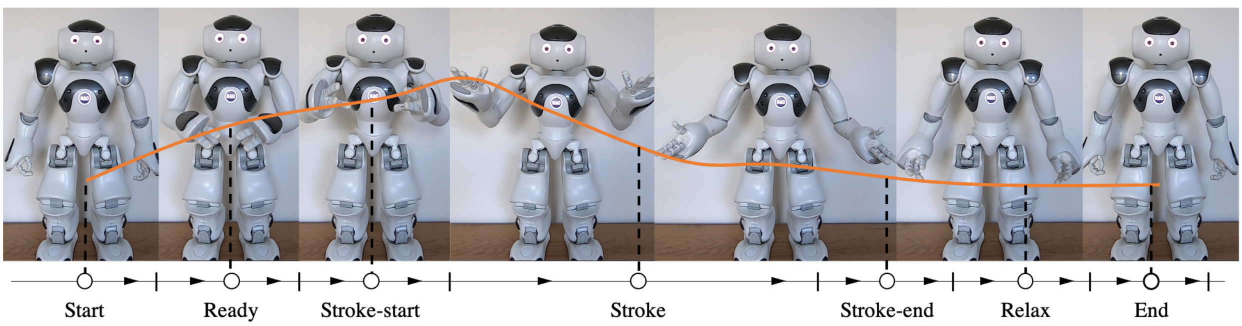


Figure 18: Synchronisation points and transitions for an explanatory beat gesture for NAO

The final synchronisation of speech and the beat gesture in Figure 17 is above in Figure 18.

APPENDIX C – PYTHON FRAMES DESCRIBING SYNCHRONISATION OF SPEECH AND GESTURES

Input:	$\boldsymbol{speech_{intro}}, \boldsymbol{speech_{main}}, \boldsymbol{speech_{end}}$ = 'Partitioned speech' $\boldsymbol{gesture_{start}}, \boldsymbol{gesture_{ready}}, \boldsymbol{gesture_{stroke\_start}}, \boldsymbol{gesture_{stroke}}, \boldsymbol{gesture_{stroke\_end}},$ $\boldsymbol{gesture_{relax}}$ = {'time': ms, 'data': {$joint_1$: radians, …, $joint_{1+n}$: radians}} $\boldsymbol{iconicgesture} = [\boldsymbol{gesture_{start}}, \dots, \boldsymbol{gesture_{relax}}]$ $\boldsymbol{gesture_{end}}$ = {'time': 1000, 'data': {$joint_1$: radians, …, $joint_{1+n}$: radians}}
Result:	Synchronised speech and gesture performed by NAO
Synchronization:	
	`yield` $speech_{intro}$ $gesture = iconicgesture$ `yield` $speech_{main}$ `yield` $gesture$ `yield` $speech_{end}$ `yield` $gesture_{end}$

Figure 19: Python frame of iconic gestures for the supportive gesturing condition with required input and results

Input:	$\boldsymbol{speech_{intro}}, \boldsymbol{speech_{main}}, \boldsymbol{speech_{end}}$ = 'Partitioned speech' $\boldsymbol{number_{random}}$ = random number between 1-23 $\boldsymbol{gesture}$ = 'Stand/BodyTalk/Speaking/BodyTalk_' $\boldsymbol{gesture_{end}}$ = {'time': 1000, 'data': {$joint_1$: radians, …, $joint_{1+n}$: radians}}
Result:	Synchronised speech and random gesture performed by NAO
Synchronization:	
	`yield` $speech_{intro}$ $gesture_{random} = gesture + number_{random}$ $say_{text} = speech_{main}$ `yield` sleep (equal number of seconds as supporting gesture) `yield` stop `yield` say_{text} `yield` $gesture_{random}$ `yield` $gesture_{end}$

Figure 20: Python frame of iconic gestures for the random gesturing condition with required input and results

Input:	$speech_{intro}, speech_{main}, speech_{end}$ = 'Partitioned speech' $head_1, \dots, head_{16}$ = {'time': 350, 'data': {'body.head.yaw': radians, 'body.head.pitch': radians}} $headmovements = [head_1, \dots, head_{16}]$ $gesture_{end}$ = {'time': 1000, 'data': {joint$_1$: radians, …, joint$_{1+n}$: radians}} $index = 0$ $naomark$ = {}
Result	List with estimation per frame whether the child is looking at the robot, screen, or elsewhere

Synchronization

```
yield speech_intro
yield call vision stream
yield subscribe on vision stream
searching_active = True
 while  searching_active = True repeat
    yield sleep 0.1
    if  index ≥ len (hadmovements) then
        index = 0
        yield headmovements[index]
        index + 1

     if  naomark ≠ {} then
        for  card in naomark['data']['body'] do
            if  card and # in card then
                yield close vision stream
                α_angle = card [0]
                β_angle= card [1]
                head_yaw_pitch = [{'time': 250, 'data': {'body.head.yaw': 2/3 α_angle,
                'body.head.pitch': 2/9 β_angle }}, {'time': 250, 'data': {'body.head.yaw': α_angle,
                'body.head.pitch': 1/3 β_angle }}]

                 if  α_angle ≤ 0.1 then
                     shoulder_roll_pitch = [{'time': 1000, 'data': {'body.arms.right.upper.roll': 2/3 α_angle,
                     'body.arms.right.upper.pitch': 2/3 β_angle }}, {'time': 1250, 'data':
                     {'body.arms.right.upper.roll': α_angle, 'body.arms.right.upper.pitch': β_angle }},
                     'body.arms.right.hand.fingers': 1}]

                elif  α_angle ≥ 0.1 then
                     shoulder_roll_pitch = [{'time': 1000, 'data': {'body.arms.left.upper.roll': 2/3 α_angle,
                     'body.arms.left.upper.pitch': 2/3 β_angle }}, {'time': 1250, 'data':
                     {'body.arms.left.upper.roll': α_angle, 'body.arms.left.upper.pitch': β_angle }},
                     'body.arms.left.hand.fingers': 1}]

                yield head_yaw_pitch
                yield shoulder_roll_pitch
                yield speech_main
                searching_active = False

Yield gesture_end
```

Figure 21: Python frame of deictic gestures for the supportive gesturing condition with required input and results for Naomark 107 or 85, for Naomark 68 and 85 `if` $\alpha_{angle} \leq 0.1$ `then` is replaced by `if` $\alpha_{angle} \leq -0.1$ `then` and `if` $_{angle} \leq 0.1$ `then` is replaced by `if` $\alpha_{angle} \leq -0.1$ `then`

Input: $speech_{intro}, speech_{main}, speech_{end}$ = 'Partitioned speech'
$head_1, \ldots, head_{16}$ = {'time': 400, 'data': {'body.head.yaw': radians, 'body.head.pitch': radians}}
$headmovements = [head_1, \ldots, head_{16}]$
$gesture_{end}$ = {'time': 1000, 'data': {$joint_1$: radians, …, $joint_{1+n}$: radians}}
$index = 0$
$naomark$ = {}

Result List with estimation per frame whether the child is looking at the robot, screen, or elsewhere

Synchronization

```
yield speech_intro
yield call vision stream
yield subscribe on vision stream
searching_active = True
while searching_active = True repeat
    yield sleep 0.1
    if index ≥ len (headmovements) then
        index = 0
        yield headmovements[index]
        index + 1

        if naomark ≠ {} then
            for card in naomark['data']['body'] do
                if card and # in card then
                    yield close vision stream
                    Yield gesture_end
                    yield speech_main
                    searching_active = False
```

Figure 22: Python frame of deictic gestures for the no gesturing condition with required input and results

Input: $speech_{intro}, speech_{main}, speech_{end}$ = 'Partitioned speech'
$head_1, \ldots, head_{16}$ = {'time': 400, 'data': {'body.head.yaw': radians, 'body.head.pitch': radians}}
$headmovements = [head_1, \ldots, head_{16}]$
$\boldsymbol{number_{random}}$ = random number between 1-23
$gesture$ = 'Stand/BodyTalk/Speaking/BodyTalk_'
$gesture_{end}$ = {'time': 1000, 'data': {$joint_1$: radians, …, $joint_{1+n}$: radians}}
$index = 0$
$naomark$ = {}

Result List with estimation per frame whether the child is looking at the robot, screen, or elsewhere

Synchronization

```
yield speech_intro
yield call vision stream
yield subscribe on vision stream
searching_active = True
while searching_active = True repeat
    yield sleep 0.1
    if index ≥ len (headmovements) then
        index = 0
        yield headmovements[index]
        index + 1

        if naomark ≠ {} then
            for card in naomark['data']['body'] do
                if card and # in card then
                    yield close vision stream

gesture_random = gesture + number_random
yield sleep (equal number of seconds as supporting gesture)
yield stop
yield gesture_random
yield speech_main
yield gesture_end
```

Figure 23: Python frame of deictic gestures for the random gesturing condition with required input and result

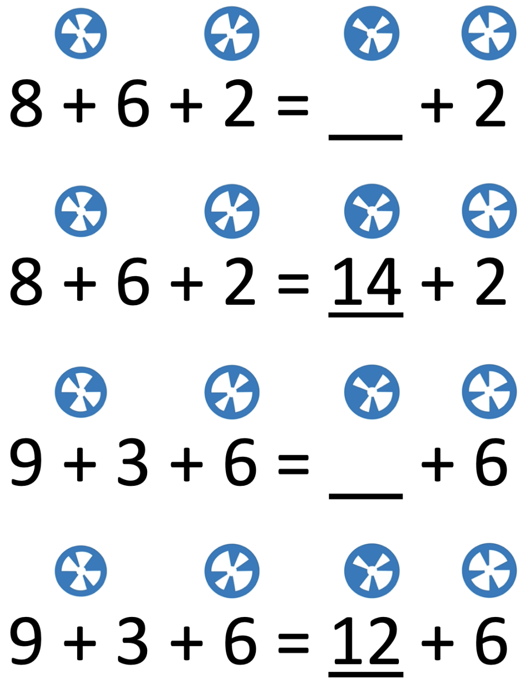


Figure 24: Mathematical equivalence problems of Form A that were explained by the robot and presented on the screen

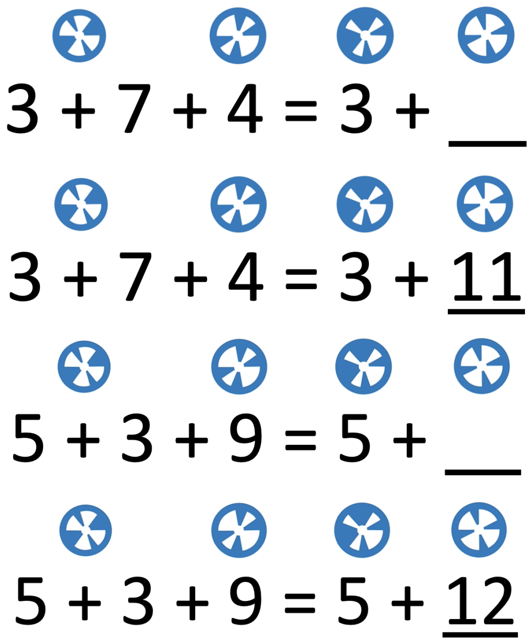
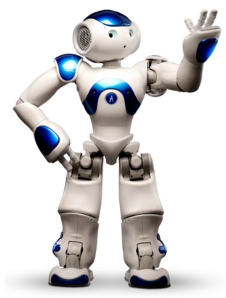


Figure 25: Mathematical equivalence problems of Form B that were explained by the robot and presented on the screen

Appendix E - Information letter for parents/legal representative of participants

DEELNEMER INFORMATIE BRIEF OUDERS/VERZORGERS



23-02-2021

Sociale robots in het onderwijs

Geachte heer/mevrouw,

Uw kind is gevraagd om deel te nemen aan het onderzoek: 'Sociale robots in het onderwijs'. Dit onderzoek wordt gedaan door Fleur Moorlag, MSc studente van de TU Delft. In deze brief vindt u informatie over het onderzoek. Voor vragen of opmerkingen kunt u altijd contact opnemen.

Details van het onderzoek

Achtergrond van het onderzoek

Sociale robots worden steeds vaker op scholen gebruikt waar kinderen vaak enthousiast zijn om met de robot te leren. Met dit onderzoek willen we bestuderen hoe een robot rekenproblemen het beste kan uitleggen aan basisschoolkinderen. Met de uitkomsten kunnen we in de toekomst de rekenuitleg van de robot verbeteren. Een voordeel van dit onderzoek is dat uw kind op een leuke manier iets over rekensommen kan leren van de sociale robot!

Doel van het onderzoek

Het doel van dit onderzoek is om uit te zoeken wat het effect is van de gebaren van een sociale robot tijdens de uitleg van rekensommen. De gegevens worden gebruikt voor een MSc thesis project van de TU Delft.

Wat houdt deelname aan het onderzoek in?

In de eerste sessie wordt de robot door de onderzoeker klassikaal geïntroduceerd door middel van een leuke en leerzame proefles (60 min). In de dagen die hierop volgen, vindt per leerling een tweede 1-op-1 sessie plaats met de rekenrobot en de onderzoeker (30 min). In deze sessie maakt de leerling een aantal rekenopgaven, en luistert en kijkt de leerling naar de uitleg van de robot. Voorafgaand geeft de onderzoeker informatie en naderhand vindt een korte debrief plaats.

Veiligheid en Privacy

Risico's van deelname

Er zijn geen risico's zijn bij deelname aan dit onderzoek. Tijdens het experiment volgen de onderzoeker en deelnemers de RIVM-richtlijnen met betrekking tot de huidige COVID-19 situatie aangezien de leraar, de onderzoeker en het kind de voorgeschreven afstand van 1,5 m in de klas kunnen aanhouden en zowel de leraar als de onderzoeker zullen mondmaskers dragen.

Procedures voor terugtrekking uit het onderzoek

Deelname aan dit onderzoek is geheel vrijwillig. Als uw kind meedoet aan het onderzoek, hebben u en uw kind de vrijheid om op elk moment terug te komen op deze beslissing. De verzamelde gegevens zullen direct daarna vernietigd worden. U kunt ook verzoeken om inzage in en rectificatie of verwijdering van persoonsgegevens. U hoeft geen verklaring te geven voor uw beslissing. Dit kunt u doen door contact op te nemen met Fleur Moorlag via het e-mailadres zoals vermeld onderaan deze brief.

Vertrouwelijkheid van gegevens

Met dit onderzoek worden de volgende persoonlijke gegevens verzameld en gebruikt: voornaam, leeftijd en geslacht, antwoorden van drie werkbladen met rekenopgaves, een enquêtevragenlijst en video-opnamen. Om de vertrouwelijkheid van persoonlijke gegevens van uw kind te waarborgen en te behouden, zullen de nodige beveiligingsmaatregelen worden genomen. De gegevens van uw kind worden opgeslagen in een beveiligde opslagomgeving bij de TU Delft. Alle gegevens worden vertrouwelijk verwerkt en opgeslagen met uitsluitend een deelnemersnummer. De naam van uw kind wordt alleen op het formulier voor geïnformeerde toestemming aan een deelnemersnummer gekoppeld. Het formulier voor geïnformeerde toestemming wordt digitaal opgeslagen op een aparte en veilige locatie. Zo blijven al uw gegevens vertrouwelijk. Alleen de onderzoeker (Fleur Moorlag) en haar twee begeleiders kunnen weten welk deelnemersnummer uw kind heeft. De tot personen herleidbare gegevens (toestemmingsformulier, werkbladen met rekensommen, video-opnames, en vragenlijsten) worden bewaard zo lang als nodig voor dataverwerking en analyse, en worden vernietigd na afronding van het afstudeeronderzoek.

De resultaten en geanonimiseerde data van deze studie zullen in mogelijke toekomstige wetenschappelijke publicaties (master thesis rapport, wetenschappelijke publicaties, rapporten) worden gepubliceerd. **Persoonlijke gegevens en video-opnames zijn alleen toegankelijk voor de onderzoeker en haar twee begeleiders voor analyse en zullen <u>nooit</u> gedeeld worden met derden.**

Beveiliging en privacy

De verwerking van de data zal in overeenstemming zijn met de AVG en gegevens worden alleen gebruikt voor het doel van dit onderzoek. Ook is een Data Protection Impact Assessment uitgevoerd.

Contactgegevens

Bij klachten over de vertrouwelijkheid van uw gegevens kunt u contact opnemen met de Functionaris Gegevensbescherming TU Delft (Erik van Leeuwen) via privacy-tud@tudelft.nl of rechtstreeks bij de Autoriteit Persoonsgegevens.

Bij voorbaat dank namens de onderzoekers voor uw eventuele medewerking,

Fleur Moorlag
(mail: f.n.moorlag@student.tudelft.nl)
Joost Broekens
(mail: d.j.broekens@liacs.leidenuniv.nl), en
Joost de Winter
(mail: j.c.f.dewinter@tudelft.nl)

Toestemmingsformulier voor sociale robots in het onderwijs

Vink de juiste vakjes aan	**Ja**	**Nee**
Deelnemen aan het onderzoek		
Ik heb de onderzoek informatie [DD / MM / JJJJ] gelezen en begrepen, of het is mij voorgelezen. Ik heb vragen kunnen stellen over de studie en mijn vragen zijn naar tevredenheid beantwoord.	□	□
Ik geef als ouder/verzorger vrijwillig toestemming voor mijn kind om deel te nemen aan dit onderzoek en begrijp dat hij/zij kan weigeren om vragen te beantwoorden en dat hij/zij zich op elk moment uit het onderzoek kan terugtrekken, zonder een reden op te geven.	□	□
Ik begrijp dat deelname aan het onderzoek een enquêtevragenlijst inhoudt die door de onderzoeker is ingevuld.	□	□
Ik begrijp dat deelname aan het onderzoek drie werkbladen met rekenopgaven omvat die door de deelnemer worden ingevuld.	□	□
Ik begrijp dat deelname aan het onderzoek inhoudt dat er een video van de deelnemer wordt opgenomen, die alleen beschikbaar zal zijn voor de geaccrediteerde onderzoekers.	□	□
Ik begrijp dat de onderzoeker en de deelnemers de RIVM-richtlijnen met betrekking tot de huidige COVID-19 situatie zullen volgen.	□	□
Gebruik van de informatie in het onderzoek		
Ik begrijp dat de informatie die mijn kind geeft, zal worden gebruikt voor onderzoeksdoeleinden in een MSc thesis project aan de Technische Universiteit Delft.	□	□
Ik begrijp dat persoonlijke informatie die over mijn kind is verzameld en die hem/haar kan identificeren, zoals zijn/haar naam, leeftijd, geslacht en videobeelden, **niet** buiten het studieteam zal worden gedeeld.	□	□
Toekomstig gebruik en hergebruik van de informatie door anderen		
Ik geef toestemming om de geanonimiseerde data van mijn kind, te archiveren in de datacollectie van de TU Delft zodat deze gebruikt kunnen worden voor toekomstig onderzoek en leren.	□	□

Handtekeningen

____ ____________________ ______

Naam deelnemer en wettelijke vertegenwoordiger Handtekening Datum

Ik heb het informatieblad nauwkeurig geschreven voor de ouder/verzorger van de potentiële deelnemer en, naar mijn beste vermogen, ervoor gezorgd dat de ouder/verzorger begrijpt waar hij/zij vrijelijk mee instemt.

Fleur Moorlag__________ ____________________ ______

Onderzoekers naam Handtekening Datum

Contactgegevens voor meer informatie:

Appendix G – Information letter participants

DEELNEMER INFORMATIE BRIEF LEERLINGEN

Hoi (naam leerling),

Je bent gevraagd om mee te doen aan een onderzoek over robots op scholen. Dit onderzoek wordt gedaan door mij, Fleur, van de universiteit in Delft. Door deze brief aan jou voor te lezen, geef ik je wat informatie, laat het me weten als je vragen hebt!

Robots zien we steeds vaker en ze worden nu ook op scholen gebruikt, misschien ook wel op jouw school. Heb jij al wel eens een robot gezien? Met het onderzoek willen we iets leren over hoe een robot het beste sommen kan uitleggen. Misschien kunnen we hiermee de reken uitleg van de robot nog meer verbeteren.

Het leuke is dat je iets over rekenen kunt leren van deze robot! De robot zal je nooit pijn doen.

Eerst word je gevraagd een werkblad in te vullen met rekensommen. Daarna legt de robot je iets uit en gaan jullie samen sommen oefenen. Hierna mag je nog twee keer een werkblad in te vullen. Aan het einde zal ik nog een paar vragen stellen over wat je van de robot vond.

Als je het niet leuk vindt om samen met deze robot sommen te oefenen, kan je me dat laten weten en kan je op elk moment stoppen.

Er zal een videocamera in het klaslokaal zijn en deze zal een video van je maken wanneer de robot je iets uitlegt. Met deze video leer ik iets over de robot. Behalve ik en twee andere onderzoekers ziet niemand anders de werkbladen en de video.

Heb je vragen over de robot, het onderzoek of mij?

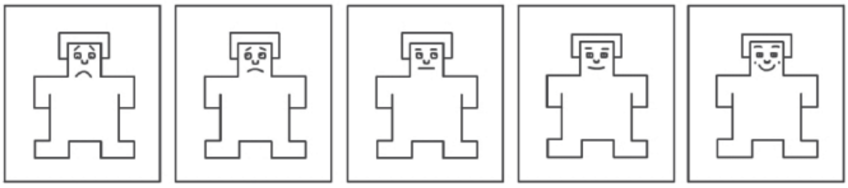
Fleur

APPENDIX H - QUESTIONNAIRE

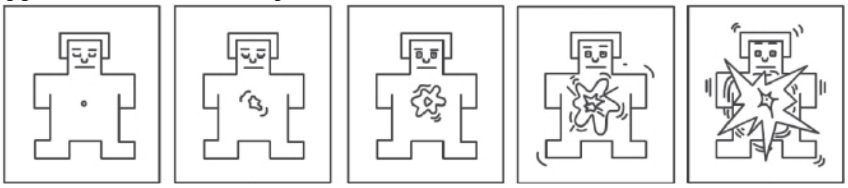
Name:____________________ Nr. : ______

To be read by researcher: Goed gedaan, dat waren de laatste sommen. NAO blijft hier nu zitten en ik ga jou nog een aantal vragen stellen over wat jij ervan vond om te leren met NAO. Hieronder staan een paar poppetjes, misschien teken je er zelf ook wel eens een. Kan jij mij per rij aanwijzen hoe jij je voelde tijdens het leren met deze robot?

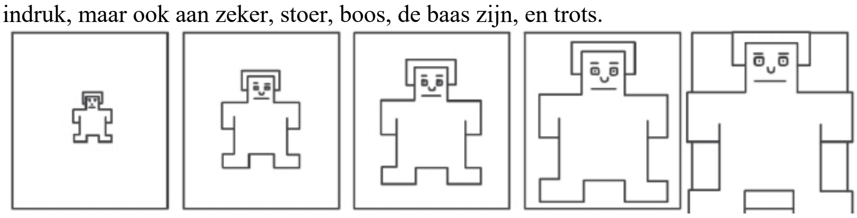
1. De eerste rij gaat over hoe **blij** je je voelde tijdens het leren. Voelde het leren goed of juist slecht. Denk aan blij, enthousiast, trots, tevreden en dankbaar, maar ook aan niet blij, verveeld, schaamte, ontevreden en boos.



2. De tweede rij gaat over hoe **rustig** je je voelde. Werd je rustig van het leren of juist actief. Denk aan rustig, saai, tevreden, moe, verdrietig, en dankbaar, en aan onrustig, actief, opgewonden, enthousiast, super wakker, en kwaad.



3. De derde rij gaat over hoe **onzeker** jij je voelde tijdens het leren. Voelde het leren met de robot als iets waarvan je dacht “Yes, dat doe ik wel even”, of dacht je juist “Hmm, ik weet niet of ik dat wel kan”. Denk aan onzeker, angstig, dankbaar, gehoorzaam, verdrietig, en onder de indruk, maar ook aan zeker, stoer, boos, de baas zijn, en trots.



Dan heb ik nu een paar vragen over hoe je het vond om over het ‘is-gelijk teken’ te leren. Ik lees de zinnen aan jou voor en dan kan jij zeggen of je het er (1) helemaal niet mee eens bent, (2) oneens mee bent, (3) niet eens en niet oneens mee bent, (4) eens mee bent, of (5) helemaal mee eens bent.

1. Ik vond het leuk om te leren over het is-gelijk teken

   O Helemaal oneens O Oneens O Niet eens of oneens O Eens O Helemaal eens

2. Ik vond het moeilijk om te leren over het is-gelijk teken

   O O O O O

3. Ik wil graag doorgaan met leren over het is-gelijk teken

   O O O O O

4. Ik wilde graag goed m'n best doen
O O O O O

5. Ik vond het saai om te leren over het is-gelijk teken
O O O O O

6. Ik vond het makkelijk
O O O O O

Oké, bedankt voor je antwoorden. Nu heb ik een paar vragen over hoe je het vond om samen met deze robot te leren. Ik lees weer de zinnen aan jou voor en kan jij dan aan mij aangeven of je het er (1) helemaal niet mee eens bent, (2) oneens mee bent, (3) niet eens en niet oneens mee bent, (4) eens mee bent, of (5) helemaal mee eens bent.

1. Ik wil graag nóg meer doen met deze robot
O Helemaal oneens O Oneens O Niet eens of oneens O Eens O Helemaal eens

2. Ik wilde aan deze robot laten zien dat ik goed m'n best deed
O O O O O

3. Ik wilde doorgaan met oefenen met deze robot
O O O O O

4. Ik vond dat deze robot goed hielp
O O O O O

5. Ik vond het makkelijk om deze robot te begrijpen
O O O O O

6. Ik vond deze robot saai
O O O O O

7. Ik vond de uitleg van de robot te snel gaan
O O O O O

En dan heb ik nu een paar vragen over de gebaren van deze robot.

8. Wat vond je van de gebaren van de robot?

Oké, bedankt voor je antwoord. Kun je me nu weer aangeven of je het (1) helemaal niet mee eens bent, (2) oneens mee bent, (3) niet eens en niet oneens mee bent, (4) eens mee bent, of (5) helemaal mee eens bent met deze zinnen?

9. Ik vond het leuk om naar de gebaren van deze robot te kijken
O O O O O

10. Ik vond dat de robot duidelijke gebaren maakte
O O O O O

APPENDIX I - PRE-TEST

Pre-test

Naam = ______________ Nr. = _______

2 + 11 + 7 = ___ + 7

4 + 10 + 6 = ___ + 6

12 + 2 + 5 = ___ + 5

2 + 4 + 7 = 2 + ___

7 + 5 + 1 = 7 + ___

9 + 8 + 2 = 9 + ___

7 + 4 + 9 = ___ + 7

3 + 2 + 11 = ___ + 3

7 + 8 + 4 = 5 + ___

4 + 7 + 2 = 8 + ___

Mid-test

Naam = _______________ Nr. = ________

11 + 3 + 4 = ___ + 4

7 + 5 + 2 = ___ + 2

4 + 4 + 8 = ___ + 8

10 + 2 + 7 = 10 + ___

2 + 15 + 3 = 2 + ___

5 + 10 + 4 = 5 + ___

3 + 4 + 9 = ___ + 3

2 + 10 + 8 = ___ + 2

9 + 4 + 1 = 8 + ___

6 + 7 + 2 = 5 + ___

Post-test

Naam = _______________ Nr. = ________

2 + 7 + 2 = ___ + 2

12 + 3 + 3 = ___ + 3

7 + 4 + 6 = ___ + 6

4 + 7 + 2 = 4 + ___

6 + 7 + 2 = 2 + ___

1 + 10 + 4 = 1 + ___

4 + 3 + 7 = ___ + 4

6 + 1 + 11 = ___ + 6

4 + 9 + 5 = 3 + ___

9 + 2 + 7 = 6 + ___

Delayed post-test
Naam = _______________ Nr. = ________

9 + 2 + 7 = ___ + 7

2 + 12 + 3 = ___ + 3

6 + 6 + 4 = ___ + 4

10 + 3 + 6 = ___ + 6

8 + 1 + 3 = ___ + 3

9 + 6 + 2 = ___ + 2

6 + 4 + 3 = ___ + 6

3 + 14 + 1 = ___ + 3

6 + 4 + 3 = 5 + ___

7 + 2 + 10 = 3 + ___

APPENDIX M – PYTHON FRAMEWORK FOR ESTIMATING PARTICIPANTS' HEAD DIRECTIONS

Data: $frame$ = number of frames with head location and rotation data
T_x = location of head with respect to camera in mm
T_y = location of head with respect to camera in mm
T_z = location of head with respect to camera in mm
R_x = rotation of head in radians around x axis (pitch)
left handed positive sign (positive for turning face right)
R_y = rotation of head in radians around x axis (yaw)
lefthanded positive sign (positive for turning face down)
$X_{R1}, Y_{R1}, Z_{R1}, X_{R2}, Y_{R2}, Z_{R2}$ = minimum and maximum coordinates for the robot, with camara as origin
$X_{S1}, Y_{S1}, Z_{S1}, X_{S2}, Y_{S2}, Z_{S2}$ = minimum and maximum coordinates for the robot, with camara as origin

Result: List with estimation per frame whether the child is looking at the robot, screen, or elsewhere

For *i* in framelist **do**

$X_cZ_{R1} = (T_z - Z_{R1}) * \tan(-R_y) + T_x$
$X_cZ_{R2} = (T_z - Z_{R2}) * \tan(-R_y) + T_x$
$Y_cZX_{R1} = ((T_z - Z_{R2})/\cos(R_y)) * \tan(R_x) + T_y$
$Y_cZX_{R2} = ((T_z - Z_{R2})/\cos(R_y)) * \tan(R_x) + T_y$

$X_cZ_{S1} = (T_z - Z_{S1}) * \tan(-R_y) + T_x$
$X_cZ_{S2} = (T_z - Z_{S2}) * \tan(-R_y) + T_x$
$Y_cZX_{S1} = ((T_z - Z_{S2})/\cos(R_y)) * \tan(R_x) + T_y$
$Y_cZX_{S2} = ((T_z - Z_{S2})/\cos(R_y)) * \tan(R_x) + T_y$

If $X_cZ_{R1} \geq X_{R1}$ and $X_cZ_{R2} \leq X_{R2}$ and $Y_cZX_{R1} \geq Y_{R1}$ and $Y_cZX_{R2} \leq Y_{R2}$ **then**
Headestimation = '*Robot*'

If $X_cZ_{S1} \leq X_{S1}$ and $X_cZ_{S2} \geq X_{S2}$ and $Y_cZX_{S1} \geq Y_{S1}$ and $Y_cZX_{S2} \leq Y_{S2}$ **then**
Headestimation = '*Screen*'

Figure 26: Python-model for estimating the participant's head direction per frame

Formulas to derive head direction distances during the intro:
To calculate the distances in mm between camera-origin and direction of the head, when directed towards the robot, on the x-axis (with respect to the camera) during the intro per frame, the formula: $T_x + tan(-R_y) * T_z$ was implemented. The formula $T_y + tan(R_x) * T_z$ was used to derive the distances in mm between direction of the head on the y-axis, when directed towards the robot, during the intro per frame (Figure 11). This was then multiplied by -1 and the camera height was added to find the height point on the y-axis, with the desk on which the robot was stated, as 0-point and a positive distance upwards from the desk (Figure 13).

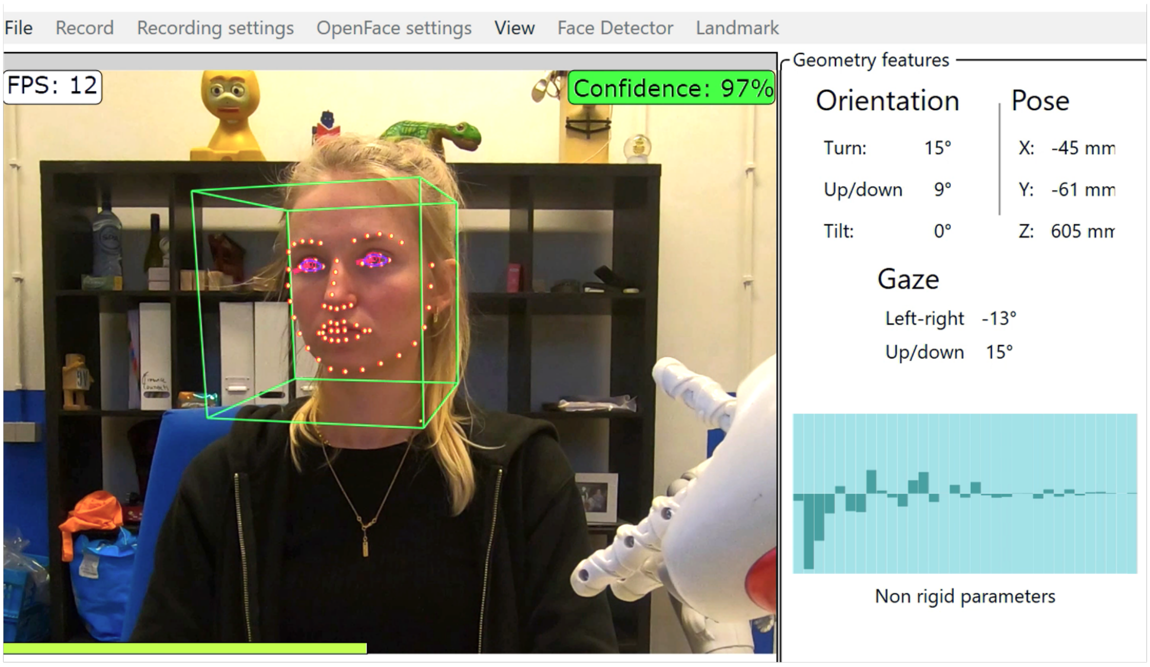


Figure 27: An example of the OpenFace model outcomes with location of the head with respect to the camera in millimetres (T_x, T_y,and T_z) and the rotation of the head (in this example in degrees) with the camera being located at the origin (R_x, R_y, and R_z)

APPENDIX N – MEAN DIFFERENCES BETWEEN TEST-SCORES PER CONDITION

Table 9: Mean differences between pre-, mid-, post-, and delayed post-test for the supportive gesturing condition (* significance at the .05 level)

	Pre-test	**Mid-test**	**Post-test**	**Delayed post-test**
Pre-test		M_{dif}= 3.46*, p <.001	M_{dif}= -5.42*, p <.001	M_{dif}= -4.89*, p <.001
Mid-test	M_{dif}= 3.46*, p <.001		M_{dif}= -1.96*, p = .036	M_{dif}= -1.43, p = .209
Post-test	M_{dif}= 5.42*, p <.001	M_{dif}=1.96*, p = .036		M_{dif}= 0.54, p = 1.000
Delayed post-test	M_{dif}= 4.89*, p <.001	M_{dif}=1.43, p = .209	M_{dif}= -0.54, p = 1.000	

Table 10: Mean differences between pre-, mid-, post-, and delayed post-test for the no gesturing condition (* significance at the .05 level)

	Pre-test	**Mid-test**	**Post-test**	**Delayed post-test**
Pre-test		M_{dif}= -4.28*, p <.001	M_{dif}= -5.56*, p <.001	M_{dif}= -4.76*, p <.001
Mid-test	M_{dif}= -4.28*, p <.001		M_{dif}= -1.28, p =.438	M_{dif}= -0.48, p =1.000
Post-test	Mdi= -5.56*, p <.001	M_{dif}= 1.28, p =.438		M_{dif}= 0.80, p = .834
Delayed post-test	M_{dif}= -4.76*, p <.001	M_{dif}= 0.48, p =1.000	M_{dif}= -0.80, p = .834	

Table 11: Mean differences between pre-, mid-, post-, and delayed post-test for the random gesturing condition (* significance at the .05 level)

	Pre-test	**Mid-test**	**Post-test**	**Delayed post-test**
Pre-test		M_{dif}= -4.04*, p <.001	M_{dif}= -4.96*, p <.001	M_{dif}= -4.56*, p <.001
Mid-test	M_{dif}= 4.04*, p <.001		M_{dif}= -0.92, p =.435	M_{dif}= -0.52, p =1.000
Post-test	M_{dif}= 4.96*, p <.001	M_{dif}= 0.92, p =.435		M_{dif}= 0.40, p =1.000
Delayed post-test	M_{dif}= 4.56*, p <.001	M_{dif}= 0.52, p =1.000	M_{dif}= -0.40, p =1.000	

Appendix O – Duration to complete the tests per condition

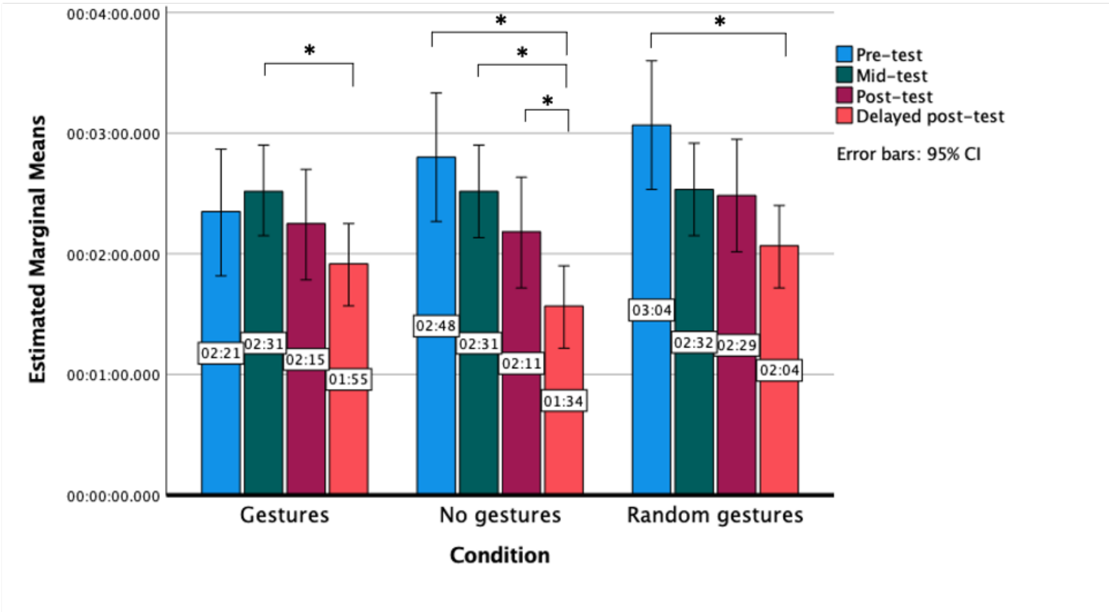


Figure 28: Mean durations to complete the pre-, mid-, post-, and delayed post-test with Confidence Intervals that are 95.0% as a function of condition ($*p < .001$)

Appendix P – Correlation matrices between questionnaire metrics

Table 12: Correlation matrix showing the Pearson correlation coefficient between the dimensions from the Self-Assessment Manakin (** Correlation is significant at the 0.01 level, 2-tailed, * Correlation is significant at the 0.05 level, 2-tailed)

	Valence	**Arousal**	**Dominance**
Valence	1	-.146 (p = .206)	.156 (p=.090)
Arousal	-.146 (p = .206)	1	-.056 (p=.629)
Dominance	.156 (p=.090)	-.056 (p=.629)	1

Table 13: Correlation matrix showing the Pearson correlation coefficient between the Task Engagement questions (** Correlation is significant at the 0.01 level, 2-tailed, * Correlation is significant at the 0.05 level, 2-tailed)

	I enjoyed this activity	**I found this activity hard**	**I would like to continue with this activity**	**It was important for me to do a good job**	**I found this activity boring**	**I found this activity easy to understand**
I enjoyed this activity	1	-.345** (p=0.002)	.552** (p<0.001)	.144 (p=.213)	-.536** (p<0.001)	.261* (p=.022)
I found this activity hard	-.345** (p=.002)	1	-.045 (p=.698)	.014 (p=.904)	.352** (p=.002)	-.691** (p<.001)
I would like to continue with this activity	.552** (p<.001)	-.045 (p=.698)	1	.311** (p=.006)	-.418** (p<.001)	.083 (p=.473)
It was important for me to do a good job	.144 (p=.213)	.014 (p=.904)	.311** (p=.006)	1	-.275* (p=.015)	.004 (p=.974)
I found this activity boring	-.536** (p<.001)	.352** (p=.002)	-.418** (p<.001)	-.275* (p=.015)	1	-.471** (p<.001)
I found this activity easy to understand	.261* (p=.022)	-.691** (p<.001)	.083 (p=.473)	.004 (p=.974)	-.471** (p<.001)	1

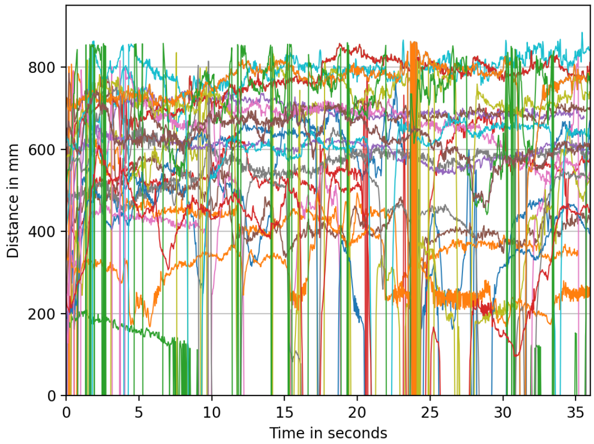
Table 14: Correlation matrix showing the Pearson correlation coefficient between the Tutor Engagement questions and gesture specific questions (** Correlation is significant at the 0.01 level, 2-tailed, * Correlation is significant at the 0.05 level, 2-tailed)

	I would like to do another activity with this robot	**I wanted to show this robot I was doing a good job**	**I wanted to keep practicing with the robot**	**I found the robot helpful**	**I found it easy to follow instructions presented to me**	**I found the robot boring**	**I found the robot's instructions too fast**	**I liked the robot's gestures**	**I found the robot's gestures helpful**
I would like to do another activity with this robot	1	.363** (p=.001)	.611** (p<.001)	.164 (p=.154)	.321** (p=.004)	-.502** (p<.001)	-.010 (p=.934)	.202 (p=.078)	.076 (p=.512)
I wanted to show this robot I was doing a good job	.363** (p=.001)	1	.504** (p<.001)	.288* (p=.011)	.281* (p=.013)	-.093 (p=.422)	-.036 (p=.753)	.219 (p=.055)	.144 (p=.210)
I wanted to keep practicing with the robot	611** (p<.001)	.504** (p<.001)	1	.336** (p=.003)	.283* (p=.013)	-.371** (p<.001)	-.151 (p=191)	.258*(p=.024)	.193 (p=.092)
I found the robot helpful	.164 (p=.154)	.288* (p=.011)	.336** (p=.003)	1	.287* (p=.011)	-.261* (p=.022)	-.215* (p=.028)	.154 (p=.181)	.273* (p=.016)
I found it easy to follow instructions presented to me	.321** (p=.004)	.281* (p=.013)	.283* (p=.013)	.287* (p=.011)	1	-.121 (p=.294)	-.168 (p=.145)	.024 (p=.834)	.264* (p=.020)
I found the robot boring	-.502** (p<.001)	-.093 (p=.422)	-.371** (p<.001)	-.261* (p=.022)	-.121 (p=.294)	1	-.073 (p=.528)	-.318** (p=.005)	-.022 (p=.847)
I found the robot's instruc-tions too fast	-.010 (p=.934)	-.036 (p=.753)	-.151 (p=191)	-.215* (p=.028)	-.168 (p=.145)	-.073 (p=.528)	1	-229* (p=.045)	-.306** (p=.007)
I liked the robot's gestures	.202 (p=.078)	.219 (p=.055)	.258* (p=.024)	.154 (p=.181)	.024 (p=.834)	-.318** (p=.005)	-229* (p=.045)	1	.243* (p=.033)
I found the robot's gestures helpful	.076 (p=.512)	.144 (p=.210)	.193 (p=.092)	.273* (p=.016)	.264* (p=.020)	-.022 (p=.847)	-.306** (p=.007)	.243* (p=.033)	1

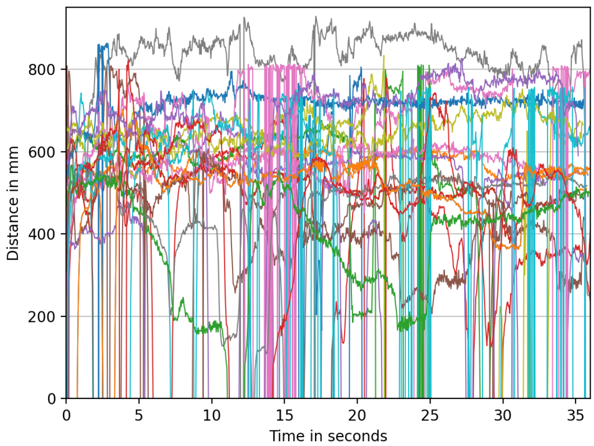
APPENDIX Q – ANSWERS TO OPEN QUESTION (DUTCH AND ENGLISH)

Table 15: Answers (in Dutch and English) given by the participants to the open question: 'What did you think of the robot's gestures?' per condition

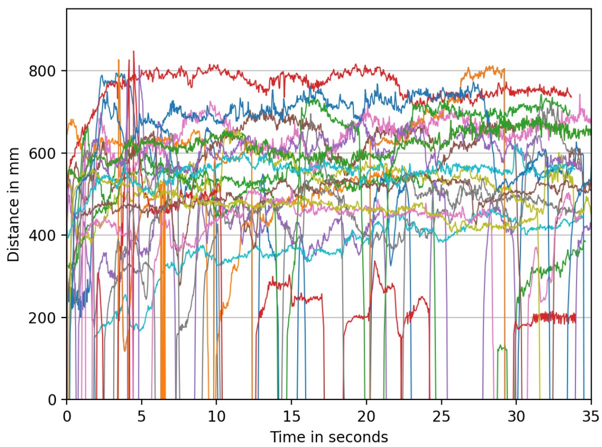
Condition	What did you think of the robot's gestures
Supportive gestures	"It is indeed useful that the robot says and portrays 'one side', and then the 'other side. (Wel handig dat die de ene kant zegt en uitbeeld en dan de andere kant.)" "So the left side is the same as the right side, he presented that very clearly. (Dus de linkerzijde is hetzelfde als de rechterzijde dat deed die heel duidelijk)." "Sometimes the gestures made it a bit difficult to hear. (Soms kon je het door de gebaren een beetje slecht verstaan)." "Also handy because you can show where you are, by pointing to the screen. (Ook wel handig want je kan dan laten zien waar, door op het bord te wijzen)." "I was able to better understand it because he pointed towards everything. (Ik kon het toen wel beter begrijpen omdat hij alles aanwees.)" "Handy if, for example, can't hear very well. (Handig als je bijvoorbeeld niet zo goed kan horen.)" "It was clear with his arms, but not so well with his fingers." (Je zag het met z'n armen wel goed, niet zo goed met z'n vingers.) "Useful because then you knew where he was explaining. (Wel handig want dan wist je waar die was met uitleggen.)"
No gestures	"The robot's walking made a lot of noise. (Het lopen maakte veel kabaal.)" "I was frightened when the robot started walking (Ik schrok toen hij ging lopen)." "The teacher also does this, but the teacher also points towards the board. (Dat doet de meester ook alleen de meester wijst ook naar het bord.)" "Sometimes the robot would turn and move its head but I didn't care. (Soms dan draaide de robot en bewoog hij z'n hoofd maar dat maakte mij niet uit.)" "Useful because it helped me understand it more. (Wel handig omdat ik het daardoor wel meer snapte.)" "Not very clear. (Niet heel duidelijk.)" "I could not really understand them. (Ik kon ze niet echt begrijpen.)"
Random gestures	"hard to understand (moeilijk om te begrijpen)" "I did not totally get them. (Ik snapte ze niet helemaal)" "I found it nice because when you just stand still, you are less likely to be looked at. (Ik vond het wel goed want als je alleen maar stil staat dan ga je minder snel naar diegene kijken)." "a little crazy (beetje gek)" "It would not have been so much fun without gestures. (Zonder was gebaren was het niet zo leuk geweest.)" "Everyone knows this moving is talking with your arms and hands. (Iedereen weet dat dit bewegen met je armen en handen praten is.)" "Very hard to explain how I thought of it. (Heel moeilijk om uit te leggen wat ik ervan vond.)" "Sometimes he did something and then I did not really understand the gestures. (Soms deed hij iets en dan snapte ik de gebaren niet echt.)" "What is this gesture? (Wat is dit gebaar?)" "He moved and that made a lot of noise. (Hij bewoog en dat maakte veel lawaai.)"



(a)

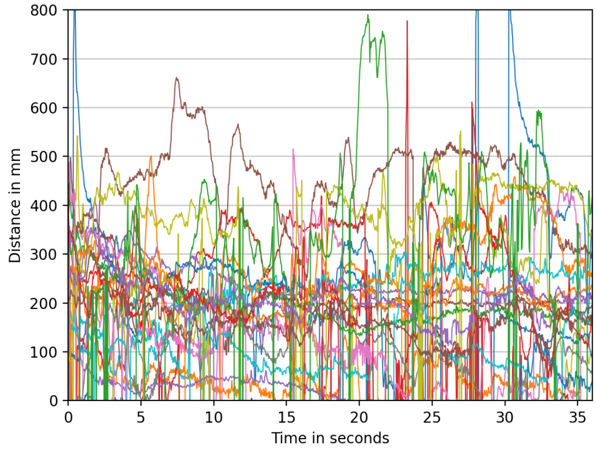


(b)

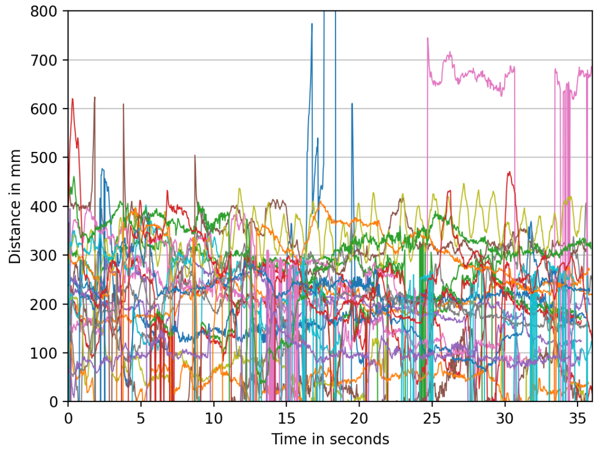


(c)

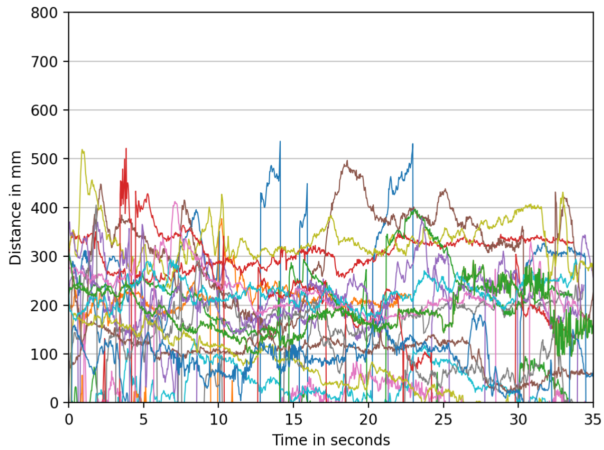
Figure 29: Distances between camera-origin and direction of the participants' heads, when directed towards the robot, on the x-axis (with respect to the camera) per participant over the intro time ($T_x + tan(-R_y) * T_z$), for the a) Supportive gesture condition ($\mu = 593.53, = 163.20$), b) No gesture condition ($\mu = 544.58, \sigma = 160.56$), and c) Random gesture condition ($\mu = 542.85, \sigma = 133.05$)
.



(a)

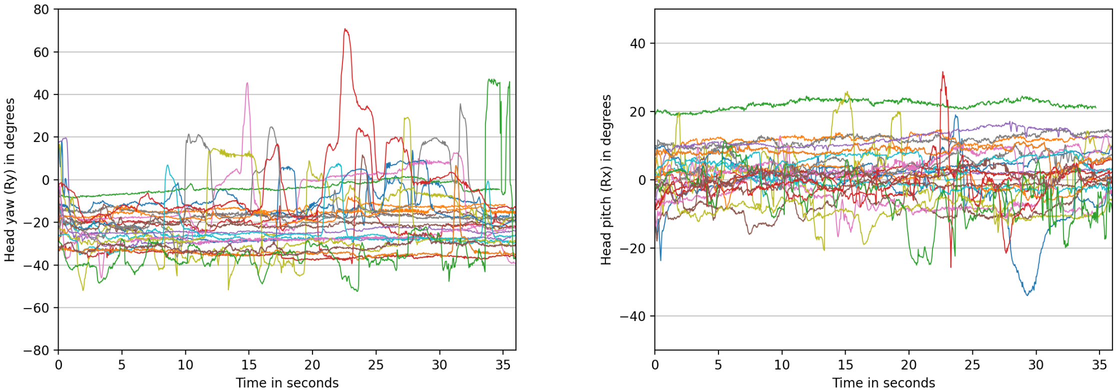


(b)



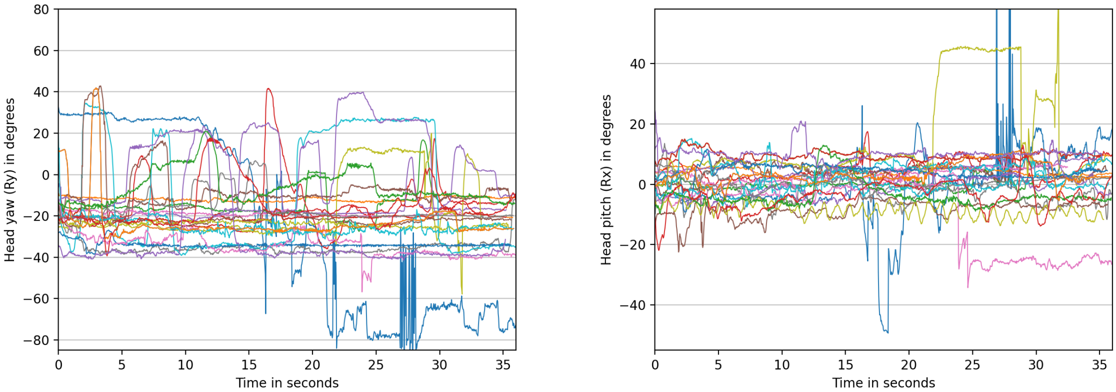
(c)

Figure 30: Distances in mm between the bottom of the robot and direction of the participants' heads on the y-axis, when directed towards the robot, per participant over the intro time for a) Supportive gesture condition ($\mu = 181.04, = 148.49$), b) No gesture condition ($\mu = 214.69, \sigma = 119.85$), and c) Random gesture condition ($\mu = 165.38, \sigma = 125.53$)



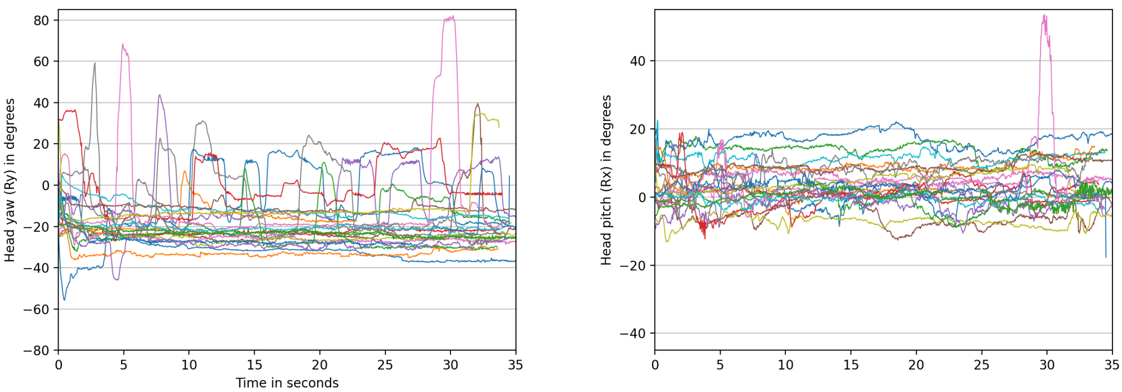
(a) **(b)**

Figure 31: a) Head yaw (R_y), $\mu = -21.87, \sigma = 12.53$ and b) head pitch (R_x), $\mu = 2.51, \sigma = 7.59$ in degrees per participant in the supportive gesturing condition during the intro time, head pitch (R_x) is positive when the participant's head is nodding down, head yaw (R_y) is positive when the participant's head is rotating to his or her right



(a) **(b)**

Figure 32: a) Head yaw (R_y), $\mu = -19.32, \sigma = 16.74$ and b) head pitch (R_x), $\mu = 2.12, \sigma = 8.29$ in degrees per participant in the no gesturing condition during the intro time



(a) **(b)**

Figure 33: a) Head yaw (R_y), $\mu = -18.45, \sigma = 13.01$ and b) head pitch (R_x), $\mu = 3.34, \sigma = 6.49$ in degrees per participant in the no gesturing condition during the intro time,