

# Back to School - Sustaining Recurring Child-Robot Educational Interactions After a Long Break

**Author(s)**

Ligthart, Mike E.U.; De Droog, Simone M.; Bossema, Marianne; Elloumi, Lamia; De Haas, Mirjam; Smakman, Matthijs H.J.; Hindriks, Koen V.; Ben Allouch, Somaya

**DOI**

[10.1145/3610977.3635001](https://doi.org/10.1145/3610977.3635001)

**Publication date**

2024

**Document Version**

Final published version

**Published in**

HRI 2024: proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction

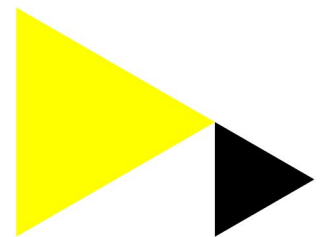
**License**

CC BY

[Link to publication](https://doi.org/10.1145/3610977.3635001)

**Citation for published version (APA):**

Ligthart, M. E. U., De Droog, S. M., Bossema, M., Elloumi, L., De Haas, M., Smakman, M. H. J., Hindriks, K. V., & Ben Allouch, S. (2024). Back to School - Sustaining Recurring Child-Robot Educational Interactions After a Long Break. In D. Grollman, & E. Broadbent (Eds.), *HRI 2024: proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 433-442). (ACM/IEEE International Conference on Human-Robot Interaction). Association for Computing Machinery. <https://doi.org/10.1145/3610977.3635001>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please contact the library: <https://www.amsterdamuas.com/library/contact>, or send a letter to: University Library (Library of the University of Amsterdam and Amsterdam University of Applied Sciences), Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Download date: 04 Dec 2025



# Back to School - Sustaining Recurring Child-Robot Educational Interactions After a Long Break

Mike E.U. Ligthart  
Vrije Universiteit Amsterdam  
Amsterdam, The Netherlands  
m.e.u.ligthart@vu.nl

Simone M. de Droog  
HU University of Applied Sciences  
Utrecht  
Utrecht, The Netherlands

Marianne Bossema  
Amsterdam University of Applied  
Sciences  
Amsterdam, The Netherlands

Lamia Elloumi  
Amsterdam University of Applied  
Sciences  
Amsterdam, The Netherlands

Mirjam de Haas  
HU University of Applied Sciences  
Utrecht  
Utrecht, The Netherlands

Matthijs H.J. Smakman  
HU University of Applied Sciences  
Utrecht  
Utrecht, The Netherlands

Koen V. Hindriks  
Vrije Universiteit Amsterdam  
Amsterdam, The Netherlands

Somaya Ben Allouch  
Amsterdam University of Applied  
Sciences  
University of Amsterdam  
Amsterdam, The Netherlands

## ABSTRACT

Maintaining the child-robot relationship after a significant break, such as a holiday, is an important step for developing sustainable social robots for education. We ran a four-session user study ( $n = 113$  children) that included a nine-month break between the third and fourth session. During the study, participants practiced math with the help of a social robot math tutor. We found that social personalization is an effective strategy to better sustain the child-robot relationship than the absence of social personalization. To become reacquainted after the long break, the robot summarizes a few pieces of information it had stored about the child. This gives children a feeling of being remembered, which is a key contributor to the effectiveness of social personalization. Enabling the robot to refer to information previously shared by the child is another key contributor to social personalization. Conditional for its effectiveness, however, is that children notice these memory references. Finally, although we found that children's interest in the tutoring content is related to relationship formation, personalizing the topics did not lead to more interest in the content. It seems likely that not all of the memory information that was used to personalize the content was up-to-date or socially relevant.

## CCS CONCEPTS

• **Applied computing** → **Education**; • **Human-centered computing** → **Empirical studies in interaction design**; • **Computing methodologies** → *Cognitive robotics*.

## KEYWORDS

Child-Robot Interaction, Long-term Engagement, Personalization, Longitudinal Study, Social Robotics, Mathematics Education

### ACM Reference Format:

Mike E.U. Ligthart, Simone M. de Droog, Marianne Bossema, Lamia Elloumi, Mirjam de Haas, Matthijs H.J. Smakman, Koen V. Hindriks, and Somaya Ben Allouch. 2024. Back to School - Sustaining Recurring Child-Robot Educational Interactions After a Long Break. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24)*, March 11–14, 2024, Boulder, CO, USA. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3610977.3635001>

## 1 INTRODUCTION

As part of an effort to reduce math deficiencies in primary education, we are developing a social robot math tutor. This work is part of the SOROCOVA project<sup>1</sup> and started after the COVID-19 pandemic led to a widening of the divide in math competences between children [47, 48]. Through a large scale user study ( $n = 130$  8-11 y.o.), that included three sessions within a week, we have shown that the math robot tutor is effective in improving children's math performance [37].

However, to truly reduce math deficiencies, children need to keep practicing throughout their school careers [24, 44]. Keeping children engaged during recurring interactions with a tutor robot spanning a long period of time is a key challenge in the field of human-robot interaction (HRI) [33, 38]. In previous work it has been shown that fostering the child-robot relationship provides a more sustainable motivation for engagement than the novelty effect [23, 38, 49]. Furthermore, a more relatable robot can potentially increase children's intrinsic motivation to work on math with the robot [10, 15].

However, an aspect of long-term HRI that has not yet been studied much is how to sustain a relationship after a significant break.



This work is licensed under a Creative Commons Attribution International 4.0 License.

HRI '24, March 11–14, 2024, Boulder, CO, USA  
© 2024 Copyright held by the owner/author(s).  
ACM ISBN 979-8-4007-0322-5/24/03.  
<https://doi.org/10.1145/3610977.3635001>

<sup>1</sup><http://www.sorocova.nl/>

In schools, holidays provide a natural gap between interaction sessions, but also breaks due to inconsistent use of robots by teaching staff or robot malfunctions are common [52]. In this paper, we investigate such a break in the interaction and how to best sustain the child-robot relationship in an educational setting.

After the three initial sessions (phase 1), we returned nine months later to the same six schools to host a fourth robot math tutoring session (phase 2) with the same participants ( $n = 113$ , 9-12 y.o.). The participants had moved up from group 6<sup>2</sup> to group 7.

To better foster the child-robot relationship, we implemented a memory-based personalization strategy [38] that makes the conversation more personal by referring to information that the child had shared previously [34], matching the conversational topics of the math dialogs to children's interests [37], and co-creating a secret handshake [40]. To address the break in the interaction, we developed a novel 'getting reacquainted' module for the memory-based personalization strategy to allow children to 'catch up' with the robot after a break.

In this paper, we focus on the social design components (discussed in Sections 2 and 3) of our robot math tutor and investigate to what extent they sustain the child-robot relationship after the nine month break (see Sections 4-7). In particular, we were interested in the contribution of memory references, personalizing the math dialogs, and the 'reacquainted moment' to keep children interested in the robot after the break.

## 2 RELATED WORK

### 2.1 Challenges of Long-Term Social Robot Tutors

Designing social robots that have to operate for a longer time in primary education is challenging for two reasons. Firstly, during their first twelve years, children experience crucial cognitive and social-emotional developments that rapidly change their educational needs, personal interests, and even their preferred interactions with robots [34, 51]. Secondly, children tend to learn more from tutors who are socially meaningful, often out of distrust of information coming from unfamiliar agents [8, 19, 32]. Over time, the robot thus needs to adapt its content and behaviors to reflect these personal changes and remain interesting, engaging, and relevant to the child [33, 55].

### 2.2 Social Personalization by Educational Robots

Personalization is an important concept in educational robotics [3, 55]. Often the focus lies on adapting the educational content [35], feedback [16], or learning strategies [36] to improve children's performance and competences [3, 22]. However, to address the social aspects of learning we also need social personalization; for example, by adapting the motivational prompts given by the tutor influenced by the affective state of the child [14].

In this paper we focus, in particular, on social personalization that contributes to relationship formation. Inspiration can be drawn from how student-teacher relationships develop. Children feeling heard and seen by the teacher is not only crucial for relationship

building [45], but also creates an inclusive learning atmosphere [41], benefiting both the student and the teacher [12].

One key improvement is for the child to feel remembered by the robot and to experience a development in the relationship [25, 38]. By asking the child to disclose information about themselves and to subsequently enable the robot to refer back to that information and use it to make the interaction more personal is a powerful strategy for giving children the feeling that they are heard and remembered by the robot, thereby fostering the relationship over time [21, 25, 30, 34, 39].

## 3 DESIGN RATIONALE FOR SOCIAL PERSONALIZATION

In this section, we present the specification of a memory-based personalization strategy that a robot math tutor can use to remain socially engaging and relevant, even after a break. In particular, we focus on three individual components to obtain this social relevancy: using memory references, personalizing the math dialogs, and the 'reacquainted moment'. We are interested in studying the effects of these components on sustaining the child-robot relationship after a break. To steer our investigation, we formulated hypotheses about the design as a whole and each individual component. We will test these hypotheses with a user study comparing the social personalization strategy with a non-personalized equivalent alternative.

### 3.1 Social Robot Math Tutor

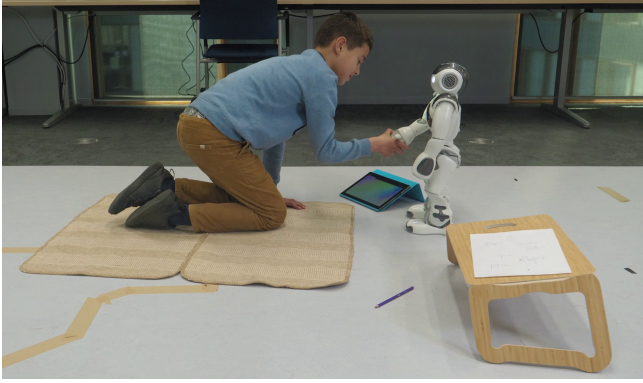
The key design principle of the social robot math tutor is to interweave social interaction with practicing math. This principle is theoretically grounded in social constructivism, which argues that learning is inherently a socially interactive process [1, 28, 46, 54]. One practical reason is that previous research has shown that social behaviors can actually distract from the educational activity if they are not directly linked [27, 29].

The main mode of interaction of the social robot tutor is a conversation. During every session the robot chitchats with the children and asks them questions about their hobbies and interests. This is followed by a series of math dialogs, in which the robot shares brief anecdotes about its fictional past jobs. A multiplication problem is embedded in each math dialog. For example, "I used to work as a dishwasher. During one shift, I had to clean 7 stacks of 14 plates. How many plates did I have to clean that shift? What is 7 times 14?".

The robot displays the problem on a tablet next to the robot. Children are given a pencil and paper to do the calculations (see Figure 1). While they are calculating the robot waits patiently. There are buttons located on the robot's feet. Children can press one button to signal to the robot that they are ready to verbally answer or press the other button to signal that they need help. If speech recognition fails after two attempts, children can provide the answer via the tablet. After a verbal answer, the robot repeats the answer and gives children the opportunity to correct the answers if necessary.

If children signal for help or answer incorrectly, the robot breaks down the sum following the principles of progressive schematization [13, 20, 50]. For example, "14 can be split into 10 and 4. You can solve  $7 \times 14$  by adding the results of  $7 \times 10$  and  $7 \times 4$ ". For each

<sup>2</sup>Group 6 (9-10 y.o.) and 7 (10-11 y.o.) are labels of the Dutch school system



**Figure 1: Experimental set-up with the robot in the middle, a small lap table with a paper and pencil on the right, and a tablet on the left. The child and the robot are doing their secret handshake.**

new type of problem the child encounters, the robot proactively demonstrates how to breakdown each type of problem.

We developed a rule-based artificial cognitive agent<sup>3</sup> to allow the robot to autonomously manage a multi-session child-robot conversation.

### 3.2 Memory-based Personalization

Memory-based personalization is the primary strategy for the robot to make the tutoring appear more personal. It persistently stores children’s answers to the questions the robot asks during the chitchat. In this way, the robot collects what children’s hobbies, interests, and preferences are. The robot uses this information in three ways. It includes references to this information throughout the conversation in the current or following sessions. We term these memory references. Secondly, the robot uses the collected interests and preferences to personalize the math dialogs. And thirdly, the robot can summarize the collected information to communicate to the child that it remembers them after a break. This is a core feature of the ‘getting reacquainted’ component. These individual components are discussed in more detail below.

After the three sessions in the first phase, we found no main effect of personalization on the child-robot relationship [37]. In previous work, differences in relationship scores due to personalization started emerging after four or five fifteen-minute sessions. After the third session relationship scores started decreasing overall. However, the decline was less strong when the robot personalized the interaction [38]. Since the length of the exposures and the content of personalization in our study are similar, we expect similar effects even though there is a nine month break between session three and four in our study. Thus, we hypothesise that personalization will result in a stronger child-robot relationship after the fourth session compared to no personalization (H1<sub>a</sub>) and the child-robot relationship will decline overall between sessions three and four (H1<sub>b</sub>).

<sup>3</sup>Code is available here: <https://bitbucket.org/socialroboticshub/sorocova-back-to-school/src/main/>

### 3.3 Memory References

The key principle underpinning memory references are communicating to the child that the robot listens to them, recognizes them, and remembers them. The robot does this in two ways. The first is by creating a secret handshake in the first session. Children can choose a song fragment and physically move the robot’s arms around to create the handshake (see Figure 1). The robot displays this secret handshake during every greeting and goodbye. However, a default wave is used when the robot does not personalize.

Secondly, the robot refers back to information shared by the child in a previous conversation. The robot fills the slots of the templated dialogs to make a reference. A memory reference is often used to motivate the inclusion of a new conversational topic, for example “You like [horses] right? I have a cool story about [horses]”. Or to make connections between the current and past conversations; for example, “My robot friend’s head was [orange], just like your favorite color”. A non-memory equivalent utterance would be, for example, “My robot friend’s head was blue, just like its favorite color”.

The memory references are designed to be explicit and reinforce the child-robot relationship. During the first phase we observed that children interpreted statements that were not informed by the memory as memory references. We expect that after more exposure (i.e. more sessions) the difference between a genuine memory reference and non-memory equivalent statements will be more clear. We hypothesise that children will perceive there are more memory references in the personalization condition (H2<sub>a</sub>). Children will perceive less memory references after more exposure when it does not personalize (H2<sub>b</sub>) and perceive an equivalent amount of memory references after more exposure when it does personalize (H2<sub>c</sub>). Finally, we expect a positive relationship between the perception of memory references and relationship formation between the child and the robot (H2<sub>d</sub>).

### 3.4 Personalized Math Dialogs

We created a collection of templated math dialogs. Each math dialog centers around one topic. These topics are related to one or more of the topics discussed during chitchat. Some math dialogs follow-up on others. The topic and dependencies are specified in the metadata of each math dialog.

The math dialogs are personalized in two ways. Firstly, the artificial agent uses the information collected about the child and the interaction history to reason about which math dialog is the best match to include next in the conversation. When the robot does not personalize, the agent randomly selects a math dialog. Secondly, the robot uses slot filling to tailor a math dialog to a child’s known preference for that topic. For example, “In the restaurant that I worked at, I used to prepare a lot of [pizza]”. When the robot does not personalize, a predefined generic slot fill is used.

The aim of personalizing the math dialogs is to make them more interesting for children and reinforce the child-robot relationship. We expect that after more exposure, as the novelty of the math dialogs decreases, personalization becomes more important. We hypothesise that children will find the personalized math dialogs more interesting (H3<sub>a</sub>). When the robot does not personalize, children’s interest in the math dialogs decreases after more exposure

(H3<sub>b</sub>) and when the robot does personalize, children's interest in the math dialogs remains equivalent after more exposure (H3<sub>c</sub>). Finally, we expect a positive relationship between dialog interest and relationship formation (H3<sub>d</sub>).

### 3.5 Getting Reacquainted

To address the nine-month break between the two phases we designed a getting reacquainted moment at the start of phase 2. As always, the robot greets the child by their name and displays their secret handshake. Then the robot provides a short recap of the conversations the child and the robot had during phase 1. The primary goal of the recap is to reinforce the notion that the robot remembers the child and feels fondly about their previous interactions. It recalls five interests or preferences about the child. Finally, the robot provides a brief refresher of how the tutoring and communication works.

When the robot has no access to the memory and does not personalize, it instead greets the child with a default wave and the blanket statement "Hi, nice to see you again". Instead of recapping the previous interaction, the robot reintroduces itself, talks about why it likes to do math, but the robot does provide the tutorial recap.

The aim of the 'reacquainted moment' is not only for children to refamiliarize with the robot, but to also give them the feeling that they are remembered by the robot. This in turn should further foster their relationship. We hypothesise that personalization will lead to children feeling more remembered by the robot (H4<sub>a</sub>) and that there is positive relationship between feeling remembered and relationship formation (H4<sub>b</sub>).

## 4 METHOD

We ran a multi-session user study in two phases. First, participants practiced math with the help of a tutor robot for three sessions within a week (phase 1). Then the robot left the school for nine months and came back for a fourth session<sup>4</sup> with the same participants (phase 2).

### 4.1 Participants

In phase 1, 130 children (group 6; 8-10 y.o.; 63 boys and 67 girls) completed the experiment. In phase 2, 113 (group 7; 9-12 y.o.; 55 boys and 58 girls) children from the first phase (the original sample) completed the experiment. Five participants that participated in phase 2, did not complete the experiment for technical reasons. 17 participants that participated in phase 1, did not in phase 2 because they were not at school during the experiment (e.g. they were sick or had changed schools).

We followed the guideline of Hogg & Tanis (2020) to recruit at least 30 participants per condition (and at least 100 respondents in total) [18]. With 4 (original) conditions, our minimum desired sample was therefore 120 participants. But we aimed for 130-140 participants to account for possible dropouts during the recurring study.

During phase 1, participants of the same age, gender, and math level were randomly split over the experimental conditions. Those

<sup>4</sup>The original plan was to have again three sessions, but we did not have the human resources for more than one.

conditions were the same in phase 2. Participants and their legal guardians signed informed consent forms before participating in phase 1 that also gave consent for participation in phase 2. This study was approved by the ethical committee of the institution of the last author (ref. number: 2022-054032).

### 4.2 Experimental Design

The full experiment had a mixed factorial design, with personalization and scaffolding as the between-subject factors and sessions as a within-subjects factor. To address the hypotheses in this paper, we only focus on personalization (without vs. with) as the between-subjects factor and phases 1 and 2 as within-subjects factor. The dependent variables were measured once after phase 1 and once again after phase 2.

### 4.3 Quantitative Measures and Instruments

The four dependent variables in phase 2 are relationship formation, perceived memory references, dialog interest, and feeling remembered. They were measured using a self-report questionnaire. The relationship formation scale was based on the scale developed by [53] and [8] and contained six items (Cronbachs  $\alpha$  phase 1: .73, phase 2: .85). The items cover aspects of comfort, similarity, friendship, and willingness to continue interacting. To accommodate children, a 4-point Likert scale was used: No, definitely not; No, not so; Yes, a little; Yes, definitely so [9]. The other variables were measured using a single item manipulation check on the same 4-point scale. The full questionnaire is available as supplementary material.

### 4.4 Qualitative Measures and Analysis

We conducted semi-structured interviews after session 3 and before and after session 4. We analysed the answers to four different questions. To investigate the impact of the robot's memory we asked participants before session 4 to speculate about what they thought the robot might remember about them (1). As part of the manipulation check children rated whether they felt the robot remembered them. During the post interview for session we asked participants how they noticed the robot remembered them (or not) (2) and what they thought about the robot (not) doing that (3). To investigate differences in the experience of the relationship before and after the break we asked participants, after session 3 and 4, to explain why the robot did (not) suit them (4).

Due to time constraints we analysed about half ( $n = 3$  interviews per participant  $\times$  55 participants = 165) of the interviews (random sample while balancing the conditions ( $n_p = 29$  and  $n_{np} = 26$ ), gender, and math level). The interviews were recorded on audio and automatically transcribed with Whisper (medium model) [43]. We performed a thematic analysis [5] where we followed a structured data-driven coding approach [11]. First, the sample was divided over five coders who collected the participants' responses to the five questions and coded each response. Secondly, the codes were reviewed by another coder. Finally, during a collective discussion the codes were refined and grouped under themes.

### 4.5 Set-up and Procedure

The set-up and procedure for phases 1 and 2 were almost identical. The study took place in an otherwise unoccupied room in the school



during normal school days. A 57 cm tall V6 Nao (humanoid) robot was used (see Figure 1). It was placed on the ground. On one side, a 9.9 inch Lenovo Tab4 10 tablet was placed in a tablet stand. On the other side a lap table was placed with paper and pencil on it. A rug was placed in front of the robot to allow the participants to sit and a handycam camera was positioned behind the robot to record the participants' behaviors during the interaction. The robot operated autonomously and was started from a laptop by a research assistant. The research assistant remained in the room but was positioned far behind the participant to avoid unnecessary contact. The research assistant would only intervene in the case of a system crash. After a reboot, the participant could continue the interaction where they left off.

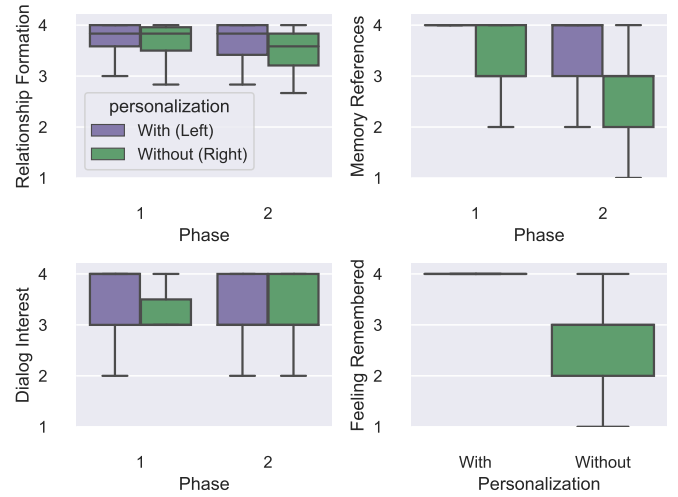
Participants came into the room one by one. There were three sessions on separate days within one week in phase 1. After nine months session 4 took place as part of phase 2. At the start of phases 1 and 2, participants received general instructions about the study and the robot, and were reminded that they could stop at any moment without giving any reason and without consequences. Session 1 in phase 1 started with a tutorial on how to talk to the robot and how the math exercises worked. Session 4 in phase 2 included a 'reacquaintance moment' (described in Section 3.5) and a recap tutorial.

All four sessions consisted of a math activity (specified in Section 3.1). After phases 1 and 2, participants could say goodbye to the robot and were taken to a separate room where they filled in the digital questionnaire and were interviewed by another research assistant who was unaware of the experimental condition to prevent a bias in the questioning.

## 5 RESULTS

The method of Brunner et al. (2002) [4], and the nparLD R-package [42], for non-parametric analysis of longitudinal data in factorial experiments (with the Wald-Type Statistic [WTS]) was used to perform a phase x personalization analysis ( $n_p = 55$  and  $n_{np} = 58$ ) on relationship formation, perceived memory reference, and dialog interest. Wilcoxon signed-rank and Mann-Whitney U post-hoc tests were to test for main effects of phase and personalization respectively. For hypotheses  $H2_c$  and  $H3_c$  Wilcoxon equivalence tests were run using the TOSTER R-package [31]. Reported effect sizes of similar constructs in comparable studies (e.g. [39]) are typically high ( $> .8$ ). To be on the safe side we selected the smallest effect size of interest (SESOI) to be .5.

Furthermore, a Mann-Whitney U test was used to check for a main personalization effect on feeling remembered. All data points are median [quartiles]. Spearman's rank-order correlation tests were run to assess the relationship between the individual items of the relationship formation questionnaire and perceived memory references, dialog interest, and feeling remembered (only in phase 2). The correlations for phase 1 are reported in Table 1 and for phase 2 in Table 2. Aside from the relationship formation scale, correlations are also reported for the individual items of this scale as well.



**Figure 2: Relationship formation (top left), perceived memory references (top right), dialog interest (bottom left) scores per condition per phase, and feeling remembered scores per condition (bottom right)**

### 5.1 Relationship Formation

The overall difference of relationship formation scores (see Figure 2; top left) between personalization conditions was found to be not statistically significant,  $WTS = 3.3$ ,  $p = .07$ . There was a main effect of phase on relationship formation,  $WTS = 15.3$ ,  $p = .00009$ ,  $d = .74$ . The interaction between phase and personalization was found to be not statistically significant,  $WTS = 1.7$ ,  $p = .19$ . The scores decreased between phases in both conditions. However, in the no-personalization condition the decline between phase 1 (3.8 [3.5, 3.95]) and 2 (3.6 [3.2, 3.8]) was stronger,  $V = 907$ ,  $p = .0003$ ,  $d = .80$ , than in the personalization condition; phase 1 (3.8 [3.6, 4.0]) and 2 (3.8 [3.4, 4.0]),  $V = 443$ ,  $p = .035$ . After the first phase, the relationship formation scores were not statistically significant,  $W = 1413$ ,  $p = .28$ , but after the second phase the participants reported statistically significantly more relationship formation in the personalization condition than in the no-personalization condition,  $W = 1256$ ,  $p = .048$ ,  $d = .37$ .

### 5.2 Perceived Memory References

There was a main effect of personalization on the perception of memory references (see Figure 2; top right),  $WTS = 68$ ,  $p < .00001$ ,  $d = 1.9$ . There was a main effect of phase,  $WTS = 6$ ,  $p = .01$ ,  $d = .57$ . The phase x personalization interaction was not statistically significant,  $WTS = 1.8$ ,  $p = .28$ . In phase 1, the scores in the personalization condition (4 [4, 4]) were statistically significantly higher than in the no-personalization condition (3 [3, 4]),  $W = 357$ ,  $p < .00001$ ,  $d = 1.03$ . In the no-personalization condition the scores statistically significantly decreased between phases,  $V = 294$ ,  $p = .03$ ,  $d = 1.24$ . The difference between the phases in the personalization condition was not statistically significant,  $V = 72$ ,  $p = .5$ . The TOST equivalence test, on the other hand, was statistically significant,  $V = 189$ ,  $p = .0012$ . This resulted in an

**Table 1: Spearman’s rank-order correlation between the individual relationship formation questionnaire items and perceived memory references and dialog interest scores for phase 1.**

Measure	R	1	2	3	4	5	6	M1	M2
R. Relationship formation	-								
1. Comfortable	.53***	-							
2. Suits me	.83***	.42***	-						
3. Feels like a friend	.70***	.11	.38***	-					
4. Continue chat	.57***	.19*	.32***	.48***	-				
5. More activities	.57***	.36***	.28***	.33***	.38***	-			
6. Meet more often	.42***	.25**	.26**	.12	.24*	.43***	-		
M1. Memory reference	.10	-.17	.07	.27*	.08	.12	-.04	-	
M2. dialog Interest	.28*	.26*	.17*	.34**	.07	.23*	.04	.21	-

\*\*\* sig. at 0.0001, \*\* sig. at .01, \* sig. at .05

**Table 2: Spearman’s rank-order correlation between the individual relationship formation questionnaire items and perceived memory references, dialog interest, and feeling remembered scores for phase 2.**

Measure	R	1	2	3	4	5	6	M1	M2	M3
R. Relationship formation	-									
1. Comfortable	.67***	-								
2. Suits me	.82***	.48***	-							
3. Feels like a friend	.81***	.50***	.54***	-						
4. Continue chat	.73***	.42***	.54***	.49***	-					
5. More activities	.59***	.33***	.36***	.35***	.52***	-				
6. Meet more often	.61***	.43***	.43***	.42***	.66***	.52***	-			
M1. Memory reference	.25**	.15	.34***	.23*	.10	.10	-.03	-		
M2. dialog Interest	.40***	.38***	.33***	.37***	.40***	.14	.34***	.26**	-	
M3. Remembrance	.19*	.06	.26**	.20*	.05	.02	.07	.56***	.16	-

\*\*\* sig. at 0.0001, \*\* sig. at .01, \* sig. at .05

even bigger gap between the personalization (4 [3, 4]) and no-personalization condition (3 [2, 3]) in study 2,  $W = 242$ ,  $p < .00001$ ,  $d = 1.5$ .

### 5.3 Math Dialog Interest

The overall differences between dialog interests scores neither between conditions and phase (see Figure 2; bottom left), nor their interaction, were statistically significant, all  $WTS < 3.5$ , all  $p's > .06$ . As reported in [37], the dialogs were statistically significantly more interesting in the personalization condition (4 [3, 4]) than in the no-personalization condition (3 [3, 3.5]) in phase 1,  $W = 551$ ,  $p = .016$ ,  $d = .488$ , but the difference between the personalization condition (3 [3, 4]) and the no-personalization condition (3 [3, 4]) was not statistically significant in phase 2,  $W = 697$ ,  $p = .48$ . In both conditions the interest scores were statistically equivalent in both phases,  $V_{np} = 87$ ,  $p_{np} = .002$  and  $V_p = 165$ ,  $p_p = .0004$ .

### 5.4 Feeling Remembered

Participants statistically significantly felt more remembered by the robot (see Figure 2; bottom right) in the personalization condition (4 [4, 4]) than in the no-personalization condition (3 [2, 3]),  $W = 394$ ,  $p < .00001$ ,  $d = .91$ .

Before session 4 we asked participants what they expected the robot to remember about them. Slightly more than half of the sample expected the robot to have a memory (Personalization: 69% vs. No-Personalization: 42%). We identified that participants expected the robot to have a social memory and remember their name, interests, or the secret handshake (P: 48% vs. NP: 31%), their math level (P:

14% vs. NP: 4%), or remember things “because it is a computer” or “it probably has access to the cloud” (P: 17% vs. NP: 12%). Finally, in the no-personalization condition participants were less sure what to expect, 46% could not answer the question versus 20% in the personalization condition.

After session 4 we asked how they could tell the robot remembered them or not. In the personalization condition 93% of the sample explicitly referred to the robot recalling their interests (81%), name (48%), or the secret handshake (44%) as the reason they felt remembered. The remaining 7% felt remembered, but could not specify why. In the no-personalization condition 19% (incorrectly) thought the robot recalled their interests, name, or secret handshake (all 40%). Another 19% mentioned the lack of memory references as a reason they did not feel remembered. 38% felt remembered, but could not specify why. Finally, 12% could not answer the question.

Participants (13%) who did not feel remembered (i.e. score 1 or 2; all in no-personalization condition) were mostly ambivalent (71%) or sometimes disappointed (29%) about it. A positive point remarked by one participant is that it offered the opportunity to share more current information with the robot after the break. For those that felt remembered by the robot (87%) it was hard to articulate what they thought about it. The majority (79%) gave nondescript positive remarks (e.g. “it was cool the robot did that”). A specific remark that was given more than once (7%) was “[I appreciated] that the robot did not have to ask the same questions again, because he can remember my answers from last time.”

We also asked participants why the robot suited them. After 3 sessions 15% of the sample argued that the robot was a good fit because it had similar interests and preferences. The feeling of similarity increased to 40% after session 4, often citing the memory references used during the ‘getting reacquainted’ moment explicitly. The feeling of similarity replaced enjoyment of the interaction as a key motivation for a good fit. Enjoyment was mentioned by 31% of the sample after session 3 and by 20% after session 4.

## 6 DISCUSSION

### 6.1 Effects of Social Personalization on the Child-Robot Tutor Relationship After a Break

Participants felt a significantly stronger relationship after the break (i.e. after session 4) when it personalized the interaction than when it did not personalize. We can therefore accept hypothesis  $H1_a$ . The difference in relationship scores between personalization and no personalization was not statistically significant before the break (i.e. after session 3). Overall, the relationship scores were high across the board, but they did decline slightly, and statistically significantly, between session 3 and 4. We can therefore accept hypothesis  $H1_b$ .

It prompts the question of what causes the decline in the relationship scores. Perhaps, the break cooled the relationship somewhat. Interestingly, however, is that Lighthart et al. (2022) reported a similar decline in relationship scores between a third and fifth interaction, but without a break. Those interactions were similar in length (i.e. between 15 and 20 minutes per session) and format (personalized conversation). Those scholars argue that it does not mean that children experience a weaker relationship over time. Instead, they suggest that children have not experienced a relationship with the

robot and, as a consequence, rate the relationship scale items based on what they hope the relationship will be. After more exposure to the robot and experiencing more of the relationship, they rate the items based on that experience [39]. In most cases it does not live up to the high expectations [6], reducing the gratification, thus resulting in lower scores [26]. Our results add to the evidence supporting this explanation. The internal consistency of the relationship formation scale improved between phases 1 and 2. This suggests that participants were more sure about how to rate the items after phase 2 than after phase 1.

Ligthart et al. (2022) furthermore reported that, similar to our findings, only after the fifth session a main effect of personalization was found on relationship formation [39]. Besides the lack of a break, another important difference between studies is that ours centered around an educational task, while theirs was purely a social conversation.

The findings contribute in the following five ways. Firstly, we replicate the results of [39] and add to the evidence that memory-based personalization is an effective strategy for relationship building. Secondly, we strengthen the evidence for the hypothesis that children first rate their relationship with the robot as what could be, and after more exposure switch to a more genuine evaluation. This stresses the importance of running more longitudinal studies, because especially effects that rely on self-report might be inflated by unrealistic expectations.

Thirdly, we demonstrate that relationship building also works when the personalization is integrated in the educational task. This is important, because if the educational task and social behaviors are separated, the social behaviors can actually distract from the education task [27, 29]. Fourthly, the findings suggest that the amount of time passed between interactions is not as important as the amount of exposure to the robot for relationship formation and assessment. The break did not seem to affect the relationship scores, when comparing the scores with a similar study that did not have a break [39]. Finally, the break also did not seem to reduce the effectiveness of the personalization strategy. Thus, memory-based personalization is a robust strategy for relationship building even if there are breaks in the interaction.

## 6.2 Contribution of Memory References

When the robot personalizes, children perceived the robot to include more memorable information that they had previously shared with the robot. This was significantly the case in both phases. We can therefore accept hypothesis  $H2_a$ . Interestingly, after phase 1 participants in the no-personalization condition were reasonably sure that the robot made personal memory references, while in fact the dialogs were all scripted and did not use stored information. This might be caused by a form-function attribution bias [17], where the social behaviors, and perhaps the humanoid appearance, of the Nao robot made children assume it made memory references [2]. In the no-personalization condition participants perceived significantly less memory references after phase 2 than phase 1. In the personalization condition this perception remained equivalently high between both phases. We can therefore accept hypothesis  $H2_b$  and  $H2_c$  respectively. Looking at the scores after phase 2, participants were fairly sure whether or not the robot used memory

references, indicating that the form function bias reduces with increasing exposure to the robot.

As the differences between conditions and between phases 1 and 2 grow, the importance of including memory references becomes clearer. After phase 2, there is a significant but weak positive relationship between perceiving the memory references and relationship formation ( $R_s = .25$ ). We can therefore accept hypothesis  $H2_d$ . By looking at the correlations between the perception of memory references and the individual items of the relationship formation scale, we can gain more detailed insights about which parts of relationship formation are probably supported by the memory references. Our study shows that children who perceive more memory references feel the robot is a better fit ( $R_s = .34$ ) and the robot feels more like a friend ( $R_s = .23$ ). This is confirmed by the interviews. Memory references however do not significantly relate to children's willingness to interact more with the robot.

Finally, the use of memory references in earlier sessions does increase children's expectations towards the robot doing so in future sessions even after a big break. It also becomes more important for relationship formation, especially to feel more similar to the robot. After session 4 more participants indicated that memory references were a key reason for why they felt similar to the robot, compared to session 3.

## 6.3 Contribution of Personalized Math Dialogs

The differences in interest scores of the math dialogs with or without personalization were not statistically significant. We therefore have to reject hypothesis  $H3_a$ . Where after phase 1 children did find personalized math dialogs more interesting, this difference dissipated in phase 2. In both conditions the scores after phases 1 and 2 were equivalent. We can therefore reject hypothesis  $H3_b$  and accept hypothesis  $H3_c$ .

The math dialogs were generally rated as interesting irregardless of personalization and phase. It confirms we succeeded in creating appealing math dialogs that are generally aligned with children's interests. It is possible that personalizing them only marginally improves them. However, we would like to suggest two possible factors that could have reduced the effectiveness of the personalization in phase 2.

The robot used information from both previous sessions as well as the current session to personalize the math dialogs. In phase 2, about 60% of the math dialogs were tailored with information from phase 1, nine months prior. It could be that the information had become outdated, because children's interests had changed or certain topics had become less relevant to them.

Another plausible explanation is that children were never interested in one or more of the topics anyway. They provided an answer when the robot asked what their favorite wild animal was for example, while they actually do not have a strong preference. Regardless of their lack of preference, they appreciated the robot using it to personalize a math dialog during phase 1. After the break either this appreciation faded or they simply forgot what answer they gave, resulting in them not recognizing a personalization attempt by the robot.

Furthermore we found that the more interesting the math dialogs were to the child, the stronger the relationship formation



( $R_s = .40$ ). We can therefore accept hypothesis  $H3_d$ . Looking at the correlations between story interest and the individual relationship formation items, we found that children felt more comfortable around the robot, the robot was a better fit, felt more like a friend, and children were more willing to continue interacting with the robot. The correlations became stronger between phases 1 and 2, signaling the importance of investing effort in keeping the math dialogs interesting as children keep interacting with the robot.

To improve our design we need to better account for changing or inaccurate interests and preferences. Including verification of the relevance and accuracy of collected information about the child could be a key improvement. This verification should be integrated in the conversation in a clever way, i.e. to not be obtuse or repetitive, when new information is received and after a significant break. The weak positive correlation in phase 2 between perceiving memory references and dialog interest signals another opportunity for (minimally) improving the personalization strategy. If the robot more clearly communicates to the child that it weaves information previously shared by the child into the math dialogs, it might make them more interesting.

#### 6.4 Contribution of Getting Reacquainted

Children felt significantly more remembered by the robot in the personalization condition. We can therefore accept  $H4_a$ . During the interview participants indicated that the robot's summary of their preferences was a key reason they felt remembered by the robot. 1 out of 5 participants in the no-personalization condition explicitly mentioned the lack of memory references reduced their feeling of being remembered. Memory references are an important mechanism for feeling remembered ( $R_s = .56$ ). Feeling remembered was strongly appreciated by the participants.

Having a recap, regardless of whether children actually (fully) recognize themselves in it, might also contribute to this effect as well. Furthermore, in the no-personalization condition, the robot offered no recap, but did say "nice to see you again". Some children indicated that this statement alone was enough for them to feel remembered by the robot. Investigating how much the robot must share and how accurate it must be for children to feel remembered will be an interesting question for future work.

There was a weak positive relationship between feeling remembered and relationship formation ( $R_s = .19$ ). We can therefore tentatively accept  $H4_b$ . The robot felt to be a better fit and more like a friend when children felt it remembered them more. When children felt more remembered it did not mean they felt more comfortable or were more willing to continue interacting as well.

#### 6.5 Limitations

Although the study included multiple sessions and took place in a school setting, it was still relatively short (in an educational context) and controlled. For example, there was a researcher in the room during the interactions. Even though the robot operated autonomously and had some ability to adapt to each child, the majority of the interaction was scripted. More work is to be done to verify how much holds up unsupervised in a real class room (e.g. like [7]) with more sessions over a longer time.

The investigation of the contribution of the individual design components is based on single item manipulation checks. Although this does not necessarily diminish the value of children's responses, the conclusions do depend on the interpretation of this one item. Although we are confident that the children understood these items well, this needs to be kept in mind when considering the findings.

Finally, the significant correlations we reported were mostly weak to moderate. This reemphasizes the complexity of factors influencing child-robot interaction. We must not oversimplify. Yes, we believe our designs are effective, but that does not mean they are the best or indeed the only way to achieve the intended effects.

#### 6.6 Future Work

A key finding of the discussed work regarding personalizing interactions, is the importance of signalling that the robot is using information previously shared by the child. How to achieve this effectively, without becoming repetitive, gimmicky, or outright obnoxious over time will be an important question for future research.

In this paper we took a deep dive into the social design elements of the robot math tutor. A similar investigation focusing on the math design elements themselves would also be valuable.

Finally, we are currently analysing task and robot engagement on the basis of the collected video data. The results will be reported in future work.

### 7 CONCLUSION

Memory-based personalization that makes tutoring by a robot more personal is an effective strategy to make the tutoring more sustainable. We have shown that our personalization strategy robustly fosters the child-robot relationship past a nine-month break.

Enabling the robot to refer to information that the child previously shared with it is a key component of the personalization strategy. Conditional for the effectiveness of these memory references is that children notice the robot making them. Feeling remembered by the robot also contributes to relationship formation, albeit to a lesser extent. Including a moment to get reacquainted after a longer break between sessions supports the feeling of being remembered. However, personalizing the math dialogs seems to be less successful. Although children's interest in the math dialog is positively related to relationship formation, the way in which they were personalized did not increase children's interest. It seems likely that the information used to tailor the math dialogs was not sufficiently relevant or meaningful. Validating the social relevancy of memory information is an important next step, especially when there are breaks in the interaction.

By making robot tutoring more personal, a social robot tutor can increase its long-term impact and finally fulfil its purpose to support children in their learning around the world.

### ACKNOWLEDGMENTS

We would like to thank all the consortium partners of SOROCOVA and in particular the schools, their teachers, the students who supported us in collecting the data and most importantly the children for contributing to this research. The SOROCOVA project was funded by the Taskforce Applied research of the Dutch Scientific Organization under number 01.006.

## REFERENCES

- [1] Roya Jafari Amineh and Hanieh Davatgari Asl. 2015. Review of constructivism and social constructivism. *Journal of Social Sciences, Literature and Languages* 1, 1 (2015), 9–16.
- [2] Alex Barco, Chiara de Jong, Jochen Peter, Rinaldo Kühne, and Caroline L. van Straten. 2020. Robot Morphology and Children's Perception of Social Robots: An Exploratory Study. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20)*. Association for Computing Machinery, New York, NY, USA, 125–127. <https://doi.org/10.1145/3371382.3378348>
- [3] Tony Belpaeme, James Kennedy, Aditi Ramachandran, Brian Scassellati, and Fumihide Tanaka. 2018. Social robots for education: A review. *Science Robotics* 3, 21 (Aug. 2018), eaat5954. <https://doi.org/10.1126/scirobotics.aat5954> Publisher: American Association for the Advancement of Science.
- [4] Edgar Brunner, Sebastian Domhof, and Frank Langer. 2002. *Nonparametric analysis of longitudinal data in factorial experiments*. Vol. 373. Wiley-Interscience.
- [5] Victoria Clarke and Virginia Braun. 2014. Thematic Analysis. In *Encyclopedia of Critical Psychology*, Thomas Teo (Ed.). Springer New York, New York, NY, 1947–1952. [https://doi.org/10.1007/978-1-4614-5583-7\\_311](https://doi.org/10.1007/978-1-4614-5583-7_311)
- [6] Alexandre Coninx, Paul Baxter, Elettra Oleari, Sara Bellini, Bert Bierman, Olivier Blanson Henkemans, Lola Cañamero, Piero Cosi, Valentin Enescu, Raquel Ros Espinoza, Antoine Hiole, Rémi Humbert, Bernd Kiefer, Ivana Kruijff-Korbayová, Rosmarijn Looije, Marco Mosconi, Mark Neerincx, Giulio Paci, Georgios Patsis, Clara Pozzi, Francesca Sacchitelli, Hichem Sahli, Alberto Sanna, Giacomo Somnavilla, Fabio Tesser, Yiannis Demiris, and Tony Belpaeme. 2015. Towards Long-Term Social Child-Robot Interaction: Using Multi-Activity Switching to Engage Young Users. *Journal of Human-Robot Interaction* 5, 1 (Aug. 2015), 32. <https://doi.org/10.5898/JHRI.5.1.Coninx>
- [7] Daniel P. Davison, Frances M. Wijnen, Vicky Charisi, Jan Van Der Meij, Vanessa Evers, and Dennis Reidsma. 2020. Working with a Social Robot in School: A Long-Term Real-World Unsupervised Deployment. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Cambridge United Kingdom, 63–72. <https://doi.org/10.1145/3319502.3374803>
- [8] Simone M. de Drog and Moniek Buijzen. 2014. Presenting validated self-report character involvement scales for 4- to 6-year-olds. ICA, Seattle, WA.
- [9] Edith D De Leeuw. 2011. Improving data quality when surveying children and adolescents: Cognitive and social development and its role in questionnaire construction and pretesting. In *Report prepared for the Annual Meeting of the Academy of Finland: Research programs public health challenges and health and welfare of children and young people*. 10–12.
- [10] Edward L Deci and Richard M Ryan. 2012. Self-determination theory. *Handbook of theories of social psychology* 1, 20 (2012), 416–436.
- [11] Jessica T. DeCuir-Gunby, Patricia L. Marshall, and Allison W. McCulloch. 2011. Developing and Using a Codebook for the Analysis of Interview Data: An Example from a Professional Development Research Project. *Field Methods* 23, 2 (May 2011), 136–155. <https://doi.org/10.1177/1525822X10388468> Publisher: SAGE Publications Inc.
- [12] Helen Demetriou and Elaine Wilson. 2010. Children should be seen and heard: the power of student voice in sustaining new teachers. *Improving Schools* 13, 1 (March 2010), 54–69. <https://doi.org/10.1177/1365480209352545>
- [13] Matthias Glade and Susanne Prediger. 2017. Students' individual schematization pathways - empirical reconstructions for the case of part-of-part determination for fractions. *Educational Studies in Mathematics* 94, 2 (Feb. 2017), 185–203. <https://doi.org/10.1007/s10649-016-9716-5>
- [14] Goren Gordon, Samuel Spaulding, Jacqueline Kory Westlund, Jin Joo Lee, Luke Plummer, Marayna Martinez, Madhurima Das, and Cynthia Breazeal. 2016. Affective Personalization of a Social Robot Tutor for Children's Second Language Skills. *Proceedings of the AAAI Conference on Artificial Intelligence* 30, 1 (March 2016). <https://doi.org/10.1609/aaai.v30i1.9914>
- [15] Frederic Guay, Robert Stupnisky, Michel Boivin, Christa Japel, and Ginette Dionne. 2019. Teachers' relatedness with students as a predictor of students' intrinsic motivation, self-concept, and reading achievement. *Early Childhood Research Quarterly* 48 (2019), 215–225. <https://doi.org/10.1016/j.ecresq.2019.03.005>
- [16] Mirjam De Haas, Peta Baxter, Chiara De Jong, Emiel Krahmer, and Paul Vogt. 2017. Exploring Different Types of Feedback in Preschooler and Robot Interaction. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Vienna Austria, 127–128. <https://doi.org/10.1145/3029798.3038433>
- [17] Kerstin S. Haring, Katsumi Watanabe, Mari Velonaki, Chad C. Tossell, and Victor Finomore. 2018. FFAB—The Form Function Attribution Bias in Human-Robot Interaction. *IEEE Transactions on Cognitive and Developmental Systems* 10, 4 (Dec. 2018), 843–851. <https://doi.org/10.1109/TCDS.2018.2851569> Conference Name: IEEE Transactions on Cognitive and Developmental Systems.
- [18] Robert V. Hogg, Elliot A. Tanis, and Dale L. Zimmerman. 2020. *Probability and statistical inference* (tenth edition ed.). Pearson, Hoboken.
- [19] Alice Ann Howard Gola, Melissa N. Richards, Alexis R. Lauricella, and Sandra L. Calvert. 2013. Building Meaningful Parasocial Relationships Between Toddlers and Media Characters to Teach Early Mathematical Skills. *Media Psychology* 16, 4 (Oct. 2013), 390–411. <https://doi.org/10.1080/15213269.2013.783774> Publisher: Routledge \_eprint: <https://doi.org/10.1080/15213269.2013.783774>.
- [20] Karel Hurts. 2008. Building cognitive support for the learning of long division skills using progressive schematization: Design and empirical validation. *Computers & Education* 50, 4 (May 2008), 1141–1156. <https://doi.org/10.1016/j.compedu.2006.11.008>
- [21] Bahar Irfan, Mehdi Hellou, and Tony Belpaeme. 2021. Coffee With a Hint of Data: Towards Using Data-Driven Approaches in Personalised Long-Term Interactions. *Frontiers in Robotics and AI* 8 (2021), 300. <https://doi.org/10.3389/frobt.2021.676814>
- [22] Wafa Johal. 2020. Research Trends in Social Robots for Learning. *Current Robotics Reports* 1, 3 (Sept. 2020), 75–83. <https://doi.org/10.1007/s43154-020-00008-3>
- [23] T. Kanda, R. Sato, N. Saiwaki, and H. Ishiguro. 2007. A Two-Month Field Trial in an Elementary School for Long-Term Human-Robot Interaction. *IEEE Transactions on Robotics* 23, 5 (Oct. 2007), 962–971. <https://doi.org/10.1109/TRO.2007.904904>
- [24] Sean H. K. Kang. 2016. Spaced Repetition Promotes Efficient and Effective Learning: Policy Implications for Instruction. *Policy Insights from the Behavioral and Brain Sciences* 3, 1 (March 2016), 12–19. <https://doi.org/10.1177/2372732215624708>
- [25] Zerrin Kasap and Nadia Magnenat-Thalmann. 2012. Building long-term relationships with virtual and robotic characters: the role of remembering. *The Visual Computer* 28, 1 (Jan. 2012), 87–97. <https://doi.org/10.1007/s00371-011-0630-7>
- [26] Elihu Katz, Jay G. Blumler, and Michael Gurevitch. 1973. Uses and Gratifications Research. *Public Opinion Quarterly* 37, 4 (1973), 509. <https://doi.org/10.1086/268109>
- [27] James Kennedy, Paul Baxter, and Tony Belpaeme. 2015. The Robot Who Tried Too Hard: Social Behaviour of a Robot Tutor Can Negatively Affect Child Learning. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '15)*. Association for Computing Machinery, New York, NY, USA, 67–74. <https://doi.org/10.1145/2696454.2696457>
- [28] Beaumie Kim. 2001. Social constructivism. *Emerging perspectives on learning, teaching, and technology* 1, 1 (2001), 16. Publisher: Besikbaar.
- [29] Elly A. Konijn and Johan F. Hoorn. 2020. Robot tutor and pupils' educational ability: Teaching the times tables. *Computers & Education* 157 (Nov. 2020), 103970. <https://doi.org/10.1016/j.compedu.2020.103970>
- [30] Ivana Kruijff-Korbayová, Elettra Oleari, Anahita Bagherzadhalimi, Francesca Sacchitelli, Bernd Kiefer, Stefania Racioppa, Clara Pozzi, and Alberto Sanna. 2015. Young Users' Perception of a Social Robot Displaying Familiarity and Eliciting Disclosure. In *Social Robotics: 7th International Conference, ICSR 2015, Paris, France, October 26–30, 2015, Proceedings*, Adriana Tapus, Elisabeth André, Jean-Claude Martin, François Ferland, and Mehdi Ammi (Eds.). Springer International Publishing, Cham, 380–389. [https://doi.org/10.1007/978-3-319-25554-5\\_38](https://doi.org/10.1007/978-3-319-25554-5_38)
- [31] Daniel Lakens, Anne M. Scheel, and Peder M. Isager. 2018. Equivalence Testing for Psychological Research: A Tutorial. *Advances in Methods and Practices in Psychological Science* 1, 2 (June 2018), 259–269. <https://doi.org/10.1177/2515245918770963>
- [32] Alexis R. Lauricella, Alice Ann Howard Gola, and Sandra L. Calvert. 2011. Toddlers' Learning From Socially Meaningful Video Characters. *Media Psychology* 14, 2 (May 2011), 216–232. <https://doi.org/10.1080/15213269.2011.573465> Publisher: Routledge \_eprint: <https://doi.org/10.1080/15213269.2011.573465>.
- [33] Iolanda Leite, Carlos Martinho, and Ana Paiva. 2013. Social Robots for Long-Term Interaction: A Survey. *International Journal of Social Robotics* 5, 2 (April 2013), 291–308. <https://doi.org/10.1007/s12369-013-0178-y>
- [34] Iolanda Leite, André Pereira, and Jill Fain Lehman. 2017. Persistent Memory in Repeated Child-Robot Conversations. In *Proceedings of the 2017 Conference on Interaction Design and Children (IDC '17)*. Association for Computing Machinery, New York, NY, USA, 238–247. <https://doi.org/10.1145/3078072.3079728>
- [35] Daniel Leyzberg, Aditi Ramachandran, and Brian Scassellati. 2018. The Effect of Personalization in Longer-Term Robot Tutoring. *ACM Transactions on Human-Robot Interaction* 7, 3 (Oct. 2018), 1–19. <https://doi.org/10.1145/3283453>
- [36] Daniel Leyzberg, Samuel Spaulding, and Brian Scassellati. 2014. Personalizing robot tutors to individuals' learning differences. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. ACM, Bielefeld Germany, 423–430. <https://doi.org/10.1145/2559636.2559671>
- [37] Mike E.U. Ligthart, Simone M. De Drog, Marianne Bossema, Lamia Elloumi, Kees Hoogland, Matthijs H.J. Smakman, Koen V. Hindriks, and Somaya Ben Allouch. 2023. Design Specifications for a Social Robot Math Tutor. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Stockholm Sweden, 321–330. <https://doi.org/10.1145/3568162.3576957>
- [38] Mike E.U. Ligthart, Mark A. Neerincx, and Koen V. Hindriks. 2022. Memory-Based Personalization for Fostering a Long-Term Child-Robot Relationship. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction (HRI '22)*. IEEE Press, Sapporo, Hokkaido, Japan, 80–89.
- [39] Mike E. U. Ligthart, Mark A. Neerincx, and Koen V. Hindriks. 2022. Getting acquainted: First steps for child-robot relationship formation. *Frontiers in Robotics and AI* 9 (Sept. 2022), 853665. <https://doi.org/10.3389/frobt.2022.853665>
- [40] Mike E. U. Ligthart, Mark A. Neerincx, and Koen V. Hindriks. 2023. It Takes Two: Using Co-creation to Facilitate Child-Robot Co-regulation. *ACM Transactions on Human-Robot Interaction* 12, 4 (Dec. 2023), 1–32. <https://doi.org/10.1145/3593812>

- [41] Diana Murdoch, Andrea R. English, Allison Hintz, and Kersti Tyson. 2020. *Feeling Heard : Inclusive Education, Transformative Learning, and Productive Struggle*. *Educational Theory* 70, 5 (Oct. 2020), 653–679. <https://doi.org/10.1111/edth.12449>
- [42] Kimihiro Noguchi, Yulia R. Gel, Edgar Brunner, and Frank Konietzschke. 2012. nparLD: An R Software Package for the Nonparametric Analysis of Longitudinal Data in Factorial Experiments. *Journal of Statistical Software, Articles* 50, 12 (2012), 1–23. <https://doi.org/10.18637/jss.v050.i12>
- [43] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine Mcleavey, and Ilya Sutskever. 2023. Robust Speech Recognition via Large-Scale Weak Supervision. In *Proceedings of the 40th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 202)*. Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (Eds.). PMLR, 28492–28518. <https://proceedings.mlr.press/v202/radford23a.html>
- [44] Kathryn M. Rich, Carla Strickland, T. Andrew Binkowski, Cheryl Moran, and Diana Franklin. 2017. K-8 Learning Trajectories Derived from Research Literature: Sequence, Repetition, Conditionals. In *Proceedings of the 2017 ACM Conference on International Computing Education Research*. ACM, Tacoma Washington USA, 182–190. <https://doi.org/10.1145/3105726.3106166>
- [45] Carla Anne Roos, Tom Postmes, and Namkje Koudenburg. 2021. Feeling heard: Operationalizing a key concept for social relations. (2021). Publisher: PsyArXiv.
- [46] Wolff-Michael Roth. 1999. *Authentic School Science: Intellectual. Learning & Knowledge* 3 (1999), 6. Publisher: Sage.
- [47] Johannes Schult, Nicole Mahler, Benjamin Fauth, and Marlit A. Lindner. 2022. Did students learn less during the COVID-19 pandemic? Reading and mathematics competencies before and after the first pandemic wave. *School Effectiveness and School Improvement* 33, 4 (Oct. 2022), 544–563. <https://doi.org/10.1080/09243453.2022.2061014>
- [48] Tessa M. Schuurman, Lotte F. Henrichs, Noémi K. Schuurman, Simone Polderdijk, and Lisette Hornstra. 2023. Learning Loss in Vulnerable Student Populations After the First Covid-19 School Closure in the Netherlands. *Scandinavian Journal of Educational Research* 67, 2 (Feb. 2023), 309–326. <https://doi.org/10.1080/00313831.2021.2006307>
- [49] Claudia Sinoo, Sylvia Van Der Pal, Olivier A. Blanson Henkemans, Anouk Keizer, Bert P.B. Bierman, Rosemarijn Looije, and Mark A. Neerincx. 2018. Friendship with a robot: Children's perception of similarity between a robot's physical and virtual embodiment that supports diabetes self-management. *Patient Education and Counseling* 101, 7 (July 2018), 1248–1255. <https://doi.org/10.1016/j.pec.2018.02.008>
- [50] Adrian Treffers. 1987. *Three dimensions: A model of goal and theory description in mathematics instruction—The Wiskobas Project*. Vol. 3. Springer Science & Business Media.
- [51] Patti M Valkenburg and Jessica Taylor Piotrowski. 2017. *Plugged in: How media attract and affect youth*. Yale University Press.
- [52] Gijs Van Ewijk, Matthijs Smakman, and Elly A. Konijn. 2020. Teachers' perspectives on social robots in education: an exploratory case study. In *Proceedings of the Interaction Design and Children Conference*. ACM, London United Kingdom, 273–280. <https://doi.org/10.1145/3392063.3394397>
- [53] Caroline L. van Straten, Rinaldo Kühne, Jochen Peter, Chiara de Jong, and Alex Barco. 2020. Closeness, trust, and perceived social support in child-robot relationship formation: Development and validation of three self-report scales. *Interaction Studies* 21, 1 (2020), 57–84. <https://www.jbe-platform.com/content/journals/10.1075/is.18052.str> ISBN: 1572-0373 Publisher: John Benjamins Type: <https://doi.org/10.1075/is.18052.str>
- [54] Lev Semenovich Vygotsky and Michael Cole. 1978. *Mind in society: Development of higher psychological processes*. Harvard university press.
- [55] Hansol Woo, Gerald K. LeTendre, Trang Pham-Shouse, and Yuhon Xiong. 2021. The use of social robots in classrooms: A review of field-based studies. *Educational Research Review* 33 (June 2021), 100388. <https://doi.org/10.1016/j.edurev.2021.100388>