Design Specifications for a Social Robot Math Tutor

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

Kees Hoogland
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
Amsterdam, The Netherlands

ABSTRACT

To benefit from the social capabilities of a robot math tutor, instead of being distracted by them, a novel approach is needed where the math task and the robot's social behaviors are better intertwined. We present concrete design specifications of how children can practice math via a personal conversation with a social robot and how the robot can scaffold instructions. We evaluated the designs with a three-session experimental user study (n = 130, 8-11 y.o.). Participants got better at math over time when the robot scaffolded instructions. Furthermore, the robot felt more as a friend when it personalized the conversation.

CCS CONCEPTS

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

KEYWORDS

Child-Robot Interaction, Mathematics Education, Social Robotics, Personalization

ACM Reference Format:
Mike E.U. Ligthart, Simone M. de Droog, 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 (HRI '23), March 13–16, 2023, Stockholm, Sweden.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3568162.3576957

1 INTRODUCTION

The COVID-19 pandemic disrupted education and increased inequalities in the classroom worldwide [59]. It is not easy for schools to support students to catch up. With the SOROCOVA project[1] we aim to contribute to reducing math deficiencies in primary schools by utilizing a social robot math companion.

Although the efficacy of social robots in education has been well established, there are still a number of key challenges that prevent a wide spread adoption of this technology [4]. Belpaeme et al. (2018) identified the technical challenge of facilitating a robust autonomous social interaction and the logistical challenge of content creation. Furthermore, using the social capabilities of a social robot does not automatically result in a better learning outcome [31, 33]. More research is needed to address the identified challenges and determine *how* to best use the robot's social capabilities.

In previous work we presented a conversational social robot that is able to autonomously and robustly facilitate [40] and personalize a multi-session interaction [39]. A new math module for the robot has been co-designed with children, teachers, and math education experts. The results of a series of focus groups is presented in [19]. In this paper we present the robot behavior design of the math module and provide concrete guidelines for creating the necessary math-related interaction content.

At the core of our design lies the operationalization of two social constructivist principles. Firstly, learning is inherently a socially interactive process [1, 32, 53, 68]. Secondly, it is more effective to sustainably acquire math problem solving skills in a socially grounded "realistic" context [13]. The latter is the basis of the instruction theory for mathematics we follow, called Realistic Mathematics Education (RME) [61].

Instead of adopting the classical paradigm of using a tablet (or other aid) as focal point for the math task and use the robot to provide feedback or intervene in another way (e.g. [5, 12, 31]), we propose a novel approach where the learning task is fully integrated in the social interaction. The core of this interaction is a conversation. We created a storyworld for the robot to exist in, like a character in a book. This storyworld helped us to create an extensive collection of connected dialogs. In these dialogs the child is presented with a math problem that the robot could use help with to solve. The math module consists of three components.

[1] http://www.sorocova.nl/

HRI '23, March 13–16, 2023, Stockholm, Sweden

ACM ISBN 978-1-4503-9964-7/23/03.
https://doi.org/10.1145/3568162.3576957

The first component personalizes the dialogs based on the preferences and interests the child self-disclosed during the conversation. The remaining components are to keep the learning task in the Zone of Proximal Development (ZPD) [42]. Specifically, the second component adapts the difficulty of the math problem based on the child's past performance. The third component scaffolds the learning by offering guidance when a child is unable to solve the problem correctly [3, 72].

We are interested in the effect of the first and last component, that directly influence the social interaction, on how social the robot is perceived (robot sociability) and the learning outcome (math performance). The goal of the personalization is to increase the robot's sociability, foster the child-robot relationship, and thus improve the learning outcome. Furthermore, scaffolding is provided using the RME method of progressive schematization [24]. Children are offered guidance to procedurally break down the math problem. We not only expect that scaffolding will directly improve the learning outcome, but that the act of offering guidance will also increase the robot's sociability. We evaluated the effect of personalization and scaffolding on the robot's sociability and math performance with a three-session user study (n = 130 children, 8-11 years old) at six primary schools.

With this paper, we contribute by providing concrete and well grounded solutions for the identified challenges, by offering a novel approach of utilizing the social capabilities of robots in math education, and by thoroughly evaluating our designs with an extensive multi-session in-the-wild user study.

2 RELATED WORK

2.1 Social Constructivism in Education

Learning tends to be more of a shared, social process than an individual experience. Within social constructivism, knowledge development is the result of social interaction and language use [41]. Teachers (and others) can challenge children to share their own ideas, provide realistic contexts in which they can test their ideas, and help them distil rules and guidelines from these ideas [41, 47, 69]. Learning together may also increase a deeper understanding of what is being learned, the ability to solve problems, and the motivation to learn [41, 47, 48]. Social constructivism is a motivation for using the social capabilities of a robot in an education context.

While children can learn from both human and non-human agents, it is important that these potential teachers are socially relevant, also referred to as "meaningful" [16, 28, 36]. We tend to distrust information presented by unfamiliar agents that we have not learned from in the past [36]. For this reason, children learn mathematics better from their mother and familiar onscreen characters than from an unfamiliar agent [36]. Literature suggests that the way children are related to a non-human agent determines whether they are motivated to actively attend to, encode, retain, and ultimately imitate and learn from the agent [16]. Children relate to non-human agents "parasocially", meaning they have developed a one-sided (from the perspective of the child) friendship with the agent that can instill emotional feelings for and a greater motivation to interact with the agent [16, 27, 36]. From this perspective, children may be more motivated to learn long-term from a robot math tutor when they have developed feelings of friendship for it.

Robots can stimulate feelings of friendship through strategic child-robot interactions [39, 64]. Human interactions become increasingly familiar over time due to a shared history and personal common ground [35]. Likewise, children have a need for familiar content and patterns of interaction [39]. Studies have demonstrated that children felt more close to the robot, perceived the robot as a friend, and wanted to continue the conversation when the robot used the child's name, referred to things the child had shared in previous sessions, and talked about topics that interested the child [35, 39].

Non-human agents are not only socially relevant through feelings of friendship, but can also be through their human-like responsiveness [8, 36, 64]. The more a child perceives an agent to appropriately respond to his/her actions (social contingency, e.g. eye-contact, feedback, guidance), the more the interaction and agent are perceived to be realistic, leading to greater trust in the educational information provided [36]. In other words, tuning in on the specific interaction needs of a child, whether in chitchat or tutoring dialogs, may increase a robot's social presence, which is the degree to which a robot is perceived as a "real" person [10]. Social presence is important for the acceptance of and engagement with robots [55] and can thereby stimulate children's willingness to interact with the robot math tutor long-term.

2.2 Realistic Mathematics Education

Realistic Mathematics Education is an instruction theory for mathematics education developed as an alternative to the more mechanistic teaching approaches [61]. RME has six main characteristics [58, 61, 62].

- *Active.* Following the principles of the active learning paradigm [44], math is best learned by doing it [61].
- *Reality.* Math problems are connected to a realistic context [62]. Note that realism is broader than real-world, it can also stem from a fictional reality [61].
- *Levels.* Understanding moves from informal context-related solutions to more general insights about the relation between numbers, more formal concepts, and solution strategies [61]. This transition happens by creating shortcuts and schematizations [24, 29].
- *Holistic.* The different math disciplines should not be treated in isolation from each other.
- *Interactive.* Being a social constructivist approach RME also emphasizes that learning math is best done in interaction with one another. It makes thinking processes more explicit [62] and stimulates reflection [61].
- *Scaffolding.* Teachers have a pro-active role to guide students through the different levels of understanding. Preferably by linking relations and schematizations they already discovered to new learning goals. This is called guided reinvention [21], which is a scaffolding strategy [3].

Using a social robot in math education fits in a RME approach. It inherently makes it active and interactive. Acting as a guide to improve a child's level of understanding is a challenging, but fitting, role for the robot. The robot is also an interesting starting point for a fictional reality, about who the robot is and what it wants [14, 15, 34, 39, 57], to ground math problems in. Making the math more holistic has more to do with the content of the math activity than with the robot.

We adopt one additional non-RME instruction component because of its great fit with using a robot. Experiencing success while learning math is a key predictor for a positive learning outcome [23, 30]. In a study where they artificially manipulated the success rate, they found that the more success children experienced, the more math problems were ultimately attempted and the larger the improvement in performance was [30]. The more confident children are, the better they are at math [23]. Friendly robots are often perceived as non-judgemental [25, 56, 60], which is a way the robot can contribute to the experience of success and children's confidence [22, 52, 54].

2.3 Social Robots in Math Education

For educational robots in (second) language learning, intertwining social interaction and learning has been proven beneficial. Several studies showed that child-robot conversations can increase learning gains, when using a meaningful context that is familiar to, and liked by children [4, 67]. In mathematics education however, the opportunities of integrating child-robot interaction and learning have remained largely unexplored.

An often used approach is to use an aid, like tablets or touch tables to perform a math task [5, 12, 31, 50]. Those that integrate the math task into a direct interaction with the robot not always include social behaviors (e.g. [26]) or not couple the social behaviors directly to doing math (e.g. [33]). A consequence of this disconnect is that there is a risk that the social behaviors of the robot distract the child from the math task instead of support it [31, 33].

An alternative strategy is to intertwine the social behaviors with the math task. Not having to switch between the task and the robot (e.g. caused by social cues and other sounds and movements of the robot) saves cognitive resources [20]. This means less distraction and a more deep processing of the educational content [17].

3 DESIGN RATIONALE AND SPECIFICATIONS

In this section, we present the rationale and specifications of the design of a child-robot math interaction. Based on the input from domain experts [19] we decided to focus on multiplication for this first iteration. The core of the interaction is a conversation. From the related work we identified five requirements the conversation needs to fulfil. The conversation needs to:

(1) contribute to relationship formation [34, 36, 39];
(2) be intertwined with the math task [17, 20, 62];
(3) provide a grounded reality for the math problems [61, 62];
(4) scaffold the learning process by providing guidance at the right time [3, 21] and
(5) provide children with an experience of success [23, 30].

The first step is to have a solid architecture that facilitates a robust social conversation. This is provided by an artificial cognitive agent. The second step is to create dialogs for the conversation. We created a storyworld to help the dialog writing process. The technical and creative details of the cognitive agent[2] and the storyworld are provided in the supplemental materials. The third step is to include robot behaviors that fulfill the five identified requirements. This is provided by three modules: math level adaptation, personalization, and scaffolding.

3.1 Math Dialogs

We distinguish between three types of dialogs. Chitchat dialogs to create a personal conversation, functional dialogs for greetings, and math dialogs. A math dialog is a short anecdote about one of the jobs the robot has had before. In this dialog the robot introduces a multiplication problem that it could use help with to solve. For example, "I used to work as a dishwasher in a restaurant. After a busy night there were X piles of dirty plates and each pile contained Y plates. How many dirty plates did I have to clean?". Depending on the current math level of the child, different values would be inserted for the X and Y.

The robot has had a wide variety of jobs and has multiple anecdotes for each job, creating a vast reality to ground the math problems in. The child and the robot are simply chatting about each others interests, triggering the robot to remember one anecdote after the other, intertwining the math task into the conversation. The robot presents a math problem not as an assignment children have to get right, but rather as something it forgot the answer to and would like to have solved. The child helps the robot by solving the problem. In case they cannot provide a correct answer, the robot does not judge, suddenly remembers the actual answer, and moves on. This is a strategy to steer the focus away from right and wrong, and focus more on practicing and experiencing success [30].

3.2 Math Level Adaptation

12 different difficulty levels of multiplication sums were defined ranging from 2x[2-10] to [11-100]x[11-100][3]. To facilitate children with an experience of success, as suggested by [30], each child started with a low difficulty level (lvl 2). After each correct answer a point is added to a counter. After each incorrect answer a point is subtracted from the counter. If the counter is at 0 when an incorrect answer is given, the difficulty level is lowered by 1. If the counter is at 2 (threshold value) when a correct answer is given, the difficulty level is increased by 1. After each level change the counter is reset to 0. This buffered approach also facilities an experience of success by providing a gradual increase of difficulty when children perform well and a fast decrease in difficulty when they perform poorly. This is reflected by participants' perception of difficulty. The majority (82%) thought the math problems were easy (33%) or at their level (49%). 13% felt they were too easy and only 4% thought they were (too) hard.

3.3 Personalization

Personalizing different aspects of the conversation has been shown to be an effective strategy to foster the child-robot relationship [11, 37, 39, 45, 66]. We implemented the memory-based personalization strategy presented by [39]. The chitchat dialogs contain a lot of questions about the child's interests, hobbies, and preferences. This information is stored in a persistent database.

[2]Code: https://bitbucket.org/socialroboticshub/sorocovagoalagent/src/main/

The memory-based personalization strategy consists of three sub-strategies. The first is memory references where the robot uses the name of the child and refers to things the child has previously shared with the robot. The second is content augmentation. It includes co-creating a secret handshake the robot displays at the start and end of every session. Finally, there is content selection based on the stored interests and preferences of the child. To more directly couple personalization to the math task, the content selection strategy specifically targets the math dialogs. For example, during the chitchat the robot tells the child it worked on a farm and asks which farm animal they like most. If they like horses the most, the robot recalls the time it has worked as a stable help and needs help figuring out how many horses were at the ranch. If they liked cows, the same dialog will be used but with cows instead.

The manipulation checks included in the user study (see Section 4) not only showed that children do notice these memory references ($r = .399$, $p < .001$). They also thought the math stories were more interesting when personalized ($r = .251$, $p = .016$).

3.4 Scaffolding

Progressive schematization is a Realistic Mathematics Education method that starts with helping children to organically develop models for mathematical concepts (e.g. 4 is double as much as 2) and informal strategies (e.g. 4x13 is double as much as 2x13). This is followed by helping them discover patterns and procedural rules within their informal strategies. It helps them increase their level of understanding in such a way they can more easily generalize strategies to solve similar or more difficult problems [24, 29, 58].

The scaffolding module will include an additional step after the child answers incorrectly and before the robot tells the answer and moves on. The robot will state it remembered a different answer and wants to be sure. The robot will go through an informal strategy[3] that the child could use to solve the current problem, without providing an answer. The robot subsequently asks the child to double check their solution and provide a new answer. For example, "To solve 9x12 you could start with 2x12 and double it to get 4x12. Then double it again to 8x12 and this is only 1x12 away from the solution". When receiving guidance children felt they were helped more by the robot ($r = .316$, $p = .002$) and felt the problems were easier to solve ($r = -.429$, $p = .001$) then without guidance.

3.5 Child-Robot Math Interaction

We developed three child-robot math interaction sessions for our prototype. The duration of each session is approximately 15 minutes. The first session starts with a getting acquainted introduction (1 min), where names are exchanged and the robot introduces its fictional and real-world goals. This is followed by a how-to-talk-to-me tutorial (2 min), where the robot and child practice with the mechanics of the speech recognition, using the tablet as fallback, and solving a math problem. This is the same in all conditions. The following session blocks have personalization (NP vs. P) and scaffolding (NS vs. S) manipulations.

The second and third session start with a greeting (0.5 min). The robot says "Hi [name], nice to see you again" either with (P) or without (NP) the child's name inserted, followed by either a generic wave (NP) or the personalized secret handshake (P). After the start, the child and robot engage in a chitchat conversation (1: 2 min, 2: 2.5 min, 3: 0.5 min), followed by the math conversation (1: 8 problems, 10 problems, 12 problems). After session 1 and 2 another chitchat block is included (1 min) followed by a goodbye (1 min). After the third session only a, bigger, goodbye block (2 min) is included.

Memory references (P) or a non-personal alternative of the same length (NP) are included in the chitchat and math conversations. For example, "Let's talk about your favorite animal, [lions]" versus "Let's talk about the amazing animal, otters". The math conversation consist of a prespecified amount of math dialogs with either a random topic (NP) or a topic that matches with the child's collected interests (P). No dialog is included more than once.

During each math dialog the robot presents a math problem in story form. The robot transforms the problem to the A x B format verbally and visually on the tablet. Then the robot waits for the child to act or for a time limit of 90 seconds to expire. After the 90 seconds the robot asks if the child needs more time (resetting the timer) or if they want to get help (S) or move to the next math problem (NS).

Each child is provided with a paper and pencil to do calculations. The Nao robot has buttons on its feet with a led above them. Those leds are turned green (right) and purple (left). The child can press a button to signal the robot they are ready to give an answer (green button), or that they do not know the answer (purple button)[4]. In case they do not know the answer, the robot will give guidance (S) or move to the next problem (NS).

When the child does answer, the robot repeats the answer. The leds above the feet buttons turn green (right) and red (left) for three seconds. If the child presses the red button, they can answer again. This can be used, for example, if the robot misheard the answer or if the child wants to correct a mistake last-minute. After three seconds or if the child pressed the green button, the robot evaluates the answer and adapts the math level for the next problem accordingly. After a correct answer the robot praises the child and after an incorrect answer the robot either moves on (NS) or provides guidance (S).

4 METHODS

4.1 Hypotheses

Our core design principle is that by positioning the practicing of math in the social space of the child-robot interaction, we facilitate learning. Providing *personalized* multiplication problems and *scaffolding* for children to schematize the problem are the two central strategies for the robot to achieve this. Following the rationale provided in the previous section, we expected that personalization and scaffolding both positively affect performance in math and the robot's sociability.

4.1.1 Math Performance. We measured the ratio of correct answers and the time it takes to provide an answer. *We hypothesised that over time participants will give more correct answers (a) and solve it faster (b) with personalization (H1a and H1b) and scaffolding (H2a and H2b).*

[3] See supplemental materials for full overview

[4] An implementation of the Touch-Based Speech Activation pattern from [26]

4.1.2 Sociability. We established that increasing the robot's sociability is important for two reasons. Firstly, if we want to stimulate learning via social mechanisms, the robot needs to be perceived as a social relevant entity (social presence). Secondly, being able to experience feelings of friendship with the robot increases children's willingness to continue interacting with the robot [39]. Both aspects are important for the long-term adherence and effectiveness of the math activity with the robot as children imitate and learn more from socially relevant entities with which they have meaningful connections [28, 36].

We hypothesised that after three sessions participants in the personalization condition would perceive the robot as more socially present (H3a) and experience more feelings of friendship (H3b). Furthermore, we hypothesised that in the scaffolding condition the robot would be perceived as more socially present (H4a). We did not expect a direct effect of providing scaffolding on feelings of friendship (H4b).

4.2 Participants

130 children aged 8-11 (97% 9-10; 63 boys and 67 girls) completed the experiment. The participants were recruited from six different Dutch primary schools, all from grade 6. The respective teachers provided a centralized national math level, ranging from E (lowest) to A (highest), for each child. The levels were spread fairly evenly between levels, with level E being the smallest group (14%) and level A and B being the biggest (both 23%). Participants with the same age, gender, and math level where randomly split over the experimental conditions. Participants and their legal guardians signed an informed consent form before participating. This study was approved by the ethical committee of the institution of the last author (ref. number: 2022-054032).

4.3 Experimental Design

The study had a mixed factorial design, namely with 2 (personalization: without vs. with) x 2 (scaffolding: without vs. with) between-subjects factors, and sessions (1, 2, 3) as within-subjects factor. The four conditions are abbreviated as P-S (n = 33), P-NS (n = 31), NP-S (n = 33), NP-NS (n = 33).

In the personalization (P) condition the robot used the interests and preferences shared by the child to personalize the topic and content of the math stories. In the non-personalization (NP) condition the robot used math stories with a random topic and fixed content.

In the scaffolding (S) condition the robot offered guidance after an incorrect answer. In the no scaffolding (NS) condition the robot moved on to the next problem after an incorrect answer.

4.4 Measures and Instruments

4.4.1 Biographical Information. The teachers provided the age, gender, and general math level for each child.

4.4.2 Math Performance. The system logged for each multiplication problem whether it was answered correctly, and the time it took to provide an answer. For each session, the ratio of correct answers (M = .81, SD = .16, range = $[.2, 1.0]$) and the average time it took to solve a problem (M = 27.7, SD = 9.4, range = $[16.9, 86.0]$) was calculated per participant.

4.4.3 Sociability. Social presence and feelings of friendship were measured using two adapted self-report questionnaires. The social presence questionnaire was based on the questionnaires developed by [2] and [70] and contained five items. The feelings of friendship questionnaire was based on the questionnaires developed by [63] and [43] and contained six items[3]. To accommodate children a 4-point Likert scale was used: No, definitely not; No, not so; Yes, a little; Yes, definitely so [18].

Three pairs of research assistants were running the study in parallel at multiple locations. To minimize the differences between the groups, a digital Qualtrics questionnaire was developed that required no actions from the research assistant. It included instructions, a practice question, and both questionnaires. Each questionnaire item together with the four-point Likert scale were displayed on the screen. The question was read out loud by a researcher in a prerecorded video. Reading the questions out loud one-by-one helps children to process and actively deliberate before giving an answer [38]. Participants could click on one of the four options.

Separate principal components analyses with varimax rotation were performed for the social presence and feelings of friendship scales. While the 5 items of social presence formed 1 factor, the reliability of this scale was insufficient (α = .577). For this reason, the individual items were included in the analyses. The 6 items of feelings of friendship formed a scale with sufficient reliability, and were averaged to form a single measure (α = .778, M = 3.708, SD = .345).

4.4.4 Manipulation Check. To check whether participants observed the implemented manipulations (personalization, scaffolding, and math level adaptation), five manipulation check items were added to the digital questionnaire.

4.4.5 Covariates. Children's gender (ranging from: r = .173 to .402, p =< .001 to .049) and math level (ranging from: $r = -.180$ to $-.296$, p =< .001 to .042) correlated significantly with social presence and feelings of friendship, and were thus included as covariates in the analyses.

4.5 Set-up and Procedure

The study took place in an 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 Lenova Tab4 10 tablet was placed in a tablet stand. On the other side a BRÄDA lap table was placed with paper and pencil. A rug was placed in front of the robot to seat the participants 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 and was positioned far behind the participant to avoid unnecessary contact. The research assistant only intervened in the case of a system crash. After a reboot, the participant could continue the interaction where they left off.

Participants came to the room one by one. There were three sessions on separate days within one week. At the start of the first session, participants received general instructions about the study and the robot and were reminded that they could stop at any moment without reason or consequences. The interaction with the robot started with a tutorial on how to talk to it and how the math exercises worked. The remainder of the first session and the two other sessions consisted of the math activity as specified in Section 3.5. After the third session, participants could say goodbye to the robot and were escorted to a separate room where they filled in the digital questionnaire and were interviewed by a different research assistant unaware of the child's experimental condition.

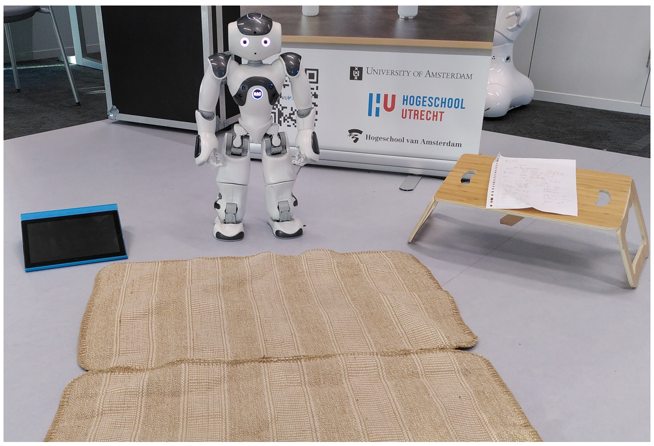

Figure 1: Experimental set-up with the robot in the middle, a small lap table with a paper and pencil on the right, and a tablet on the left.

5 RESULTS

5.1 Math Performance

The method of Brunner et al. (2002) [6], and the nparLD R-package [46], for non-parametric analysis of longitudinal data in factorial experiments (with the Wald-Type Statistic [WTS]) was used to perform a three-way analysis to investigate if there were interaction effects between the scaffolding and personalization conditions and session on the ratio of correct answers and the answer time respectively (H1, H2). Post hoc tests were run with a Bonferroni correction. Data points are median [quartiles].

5.1.1 Correct Ratio. The three-way interaction was not statistically significant, $WTS = 1.6$, $p = .45$. There was a significant two-way interaction between scaffolding and session, $WTS = 27.6$, $p \leq .0001$. All other two-way interactions were not statistically significant ($p's \geq .27$). Furthermore, significant main effects of scaffolding, $WTS = 76.4$, $p \leq .0001$, and session, $WTS = 14.4$, $p = .0007$, on the ratio of correct answers were found. The main effect of personalization was not statistically significant, $WTS = .21$, $p = .64$.

The ratios in the scaffolding conditions are shown in Figure 2 (top). Mann-Whitney U tests revealed a statistical significant difference between no scaffolding ratio (2: .79 [.70 .89] ;3: .70 [.67 .75]) and scaffolding ratio (2: 1.0 [.83 1.0];3: 1.0 [.83 1.0]) in session 2, $U = 3307.0$, $z = 5.9$, $p \leq .0001$, Cohen's $d = 1.18$, and 3, $U = 3623.5$, $z = 7.9$, $p \leq .0001$, Cohen's $d = 1.89$. The difference in session 1 (.75 [.62 .88] vs. .88 [.67 1.0]) was not statistically significant, $U = 2387.0$, $z = 2.2$, $p = .03$.

Friedman's tests revealed that for both scaffolding conditions there was a statistically significant effect of session on the correct ratio, χ^2's ≥ 10.9, $p's \leq .004$. Pairwise comparisons showed that in the no scaffolding conditions performance dropped significantly between the second (.79 [.70 .89]) and third session (.70 [.67 .75]), $p = .002$. In the scaffolding condition on the other hand performance increased after the first session (.88 [.67 1.0]) and stabilised at 1.0 [.83 1.0], $p = .014$ (session 1-2) and $p = .002$ (session 1-3). No other pairwise comparisons were statistically significant, $p's \geq .021$

5.1.2 Answer Time. The three-way interaction was not statistically significant, $WTS = 1.3$, $p = .52$. There was a significant two-way interaction between scaffolding and session, $WTS = 44.3$, $p \leq .0001$. All other two-way interactions were not statistically significant ($p's \geq .279$). Furthermore, a significant main effect of scaffolding on the time it took to answer a question was found, $WTS = 59.9$, $p \leq .0001$. The main effect of personalization and session were not statistically significant ($p's \geq .08$).

The answer times in the scaffolding conditions are shown in Figure 2 (bottom). Mann-Whitney U tests revealed a statistical significant difference between no scaffolding ratio (2: 30 [25 35];3: 31 [25 40]) and scaffolding ratio (2: 21 [20 24];3: 21 [20 24]) in session 2, $U = 696.5$, $z = -6.5$, $p \leq .0001$, Cohen's $d = 1.40$, and 3, $U = 479.5$, $z = -7.4$, $p \leq .0001$, Cohen's $d = 1.74$. The difference in session 1 (25 [22 28] vs. 26 [23 33]) was not statistically significant, $U = 1636.5$, $z = -1.56$, $p = .12$.

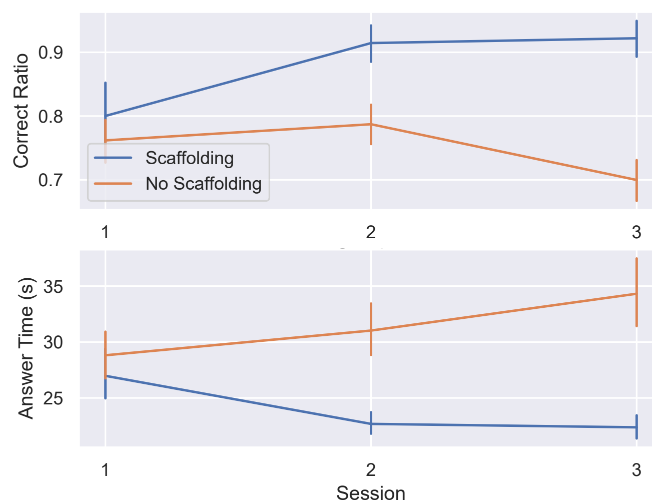

Figure 2: Ratio of correct answers (Top) and average time in seconds participants took to provide a solution to a math problem (bottom) with 95% CI for all three sessions in the scaffolding and no scaffolding conditions.

5.2 Robot Sociability

To investigate that after three sessions personalization increases social presence and feelings of friendship while scaffolding increases solely social presence (H3, H4), two separate analyses of (co-)variance were performed with personalization and scaffolding as between-subject factors, either social presence or feelings of friendship as dependent variables, and gender and math level as covariates. Because of the expected direction of the relations, all the reported significance levels in this results section are based on one-tailed tests (all means and standard deviations are reported in the supplemental material).

5.2.1 Social Presence. From all the individual social presence items, the analysis yielded a significant personalization x scaffolding interaction effect solely for item SP4 *"Do you think the robot can have feelings (e.g. sadness or joy)?"*, $F(1, 122) = 2.79$, $p = .049$. The estimated marginal means indicated that only in the non-personalized conditions did children who were guided by the robot more often thought that robots can have feelings ($M = 2.87$, $SD = .15$) than children who received no guidance from the robot ($M = 2.41$, $SD = .15$). In the personalized conditions, the means between guided and non-guided children were both around 2.70.

Furthermore, the analysis yielded a significant main effect of scaffolding for item SP5 *"Do you think the robot is a living creature?"*, $F(1, 122) = 3.88$, $p = .026$. The estimated marginal means indicated that children who were guided by the robot more often thought that the robot is a living creature ($M = 2.27$, $SD = .10$) than children who received no guidance from the robot ($M = 1.97$, $SD = .11$). No other main effects of scaffolding were significant, nor were there any significant main effects of personalization.

5.2.2 Feelings of Friendship. No significant effects were found when the feelings of friendship scale was used as dependent variable. For this reason, similar to social presence, the individual items of feelings of friendship were included in the analysis. This yielded a main effect of personalization on item F3 *"Does the robot feel like a friend to you?"*, $F(1, 122) = 3.20$, $p = .038$. The estimated marginal means indicated that children in the personalization condition more often felt that the robot was a friend to them ($M = 3.64$, $SD = .07$) than children in the non-personalized condition ($M = 3.47$, $SD = .07$). No other main effects of personalization were significant, nor were there any significant main effects of scaffolding or interaction effects.

6 DISCUSSION

6.1 Math Performance

Creating a more personal conversation and personalizing the math problems did not lead to a better math performance (rejection of hypotheses H1a and H1b). Scaffolding on the other hand did result in more correct and faster answers over time (acceptance of H2a and H2b). After session 1, participants who received guidance after a mistake outperformed those who did not.

An important question is what aspect of the scaffolding caused the increase in performance. There is the instructional part of the guidance, the progressive schematization. And there is the behavior of the robot, lending a helping hand and providing a second attempt to solve the problem. In the no scaffolding condition the robot provides the correct solution, without an explicit judgement, but also without help or second chance. Looking at the overall performance provides relevant context to address the question.

Performance overall was high. Participants solved on average 81% of the math problems correctly during the first attempt. This is likely the result of the calibration of the difficulty of the math problems. It was initially set on a low difficulty and only gradually became more difficult. This was done to provide children the experience of success. Because the robot only provided guidance after an incorrect answer, a consequence of the high success rate is a low exposure to the guidance of the robot. It could be that participants quickly picked up schematization skills as a result of the instruction. Although typically this takes longer than a few exposures [7]. Thus, it is unlikely that the instructional part is the strongest contributor of the found effect.

Getting a second chance reduces the risk of giving an answer you are unsure about, because you are able to correct it when you do get it wrong. This could explain why participants answer faster in the scaffolding condition. Furthermore, one of the guiding principles of the storyworld we created is that the child and the robot are doing math together. Proving a helping hand, in the form of helpful instructions, is better in line with that narrative. They can count on the robot helping out whenever they make a mistake or get stuck. This could explain why aspects of the social presence of the robot was higher in the scaffolding condition. Although this evidence is circumstantial, it could point to a mediating factor that the scaffolding behaviors supported children's math self-efficacy. Math self-efficacy is well established as a strong predictor for math performance [49, 51, 71].

The personalization behaviors did not affect the performance directly. Perhaps we were too optimistic with our expectations that it could influence math performance. At least it also did not inhibit performance as reported by [31, 33]. The math stories were overall well appreciated and this appreciation was stronger when they were personalized. The more the stories were appreciated, the more positive the children were about doing math with the robot. The pay-off of this effect – children remaining interested in practicing math with the robot – is likely only noticeable after more sessions [39]. Personalization could lead to more math adherence and thus a better performance over a longer period of time.

The set-up and measures we used were not sufficient to definitively explain what caused the performance increase in the scaffolding condition and to explore the long-term effects of personalization on math performance. Those questions are left for future research, with more exposures and a more longitudinal character. What remains though, is that scaffolding improves math performance.

6.2 Robot Sociability

Personalizing the interaction and math problems did not convincingly increase the robot's social presence after three sessions (H3a). It did, however, increase children's feelings of friendship toward the robot (H3b). Specifically, children in the personalization condition more often felt that the robot was a friend to them. Furthermore, providing guidance when children gave a wrong answer did indeed increase the robot's social presence (H4a) in two ways: children more often thought that the robot could have feelings like sadness and joy, and believed that the robot was a living creature. As anticipated, scaffolding did not affect children's feelings of friendship toward the robot (H4b).

In line with [36, 64], scaffolding made the robot more socially relevant by responding appropriately to children's need for guidance when they did not know the answer to the problem. That tuning into children's needs makes a robot feel more like a real person, was also observed on the 4th manipulation question where children more often believed the robot to be a living creature when the robot adjusted the level of the problems to their own math level (r = .344, p < .001). Although an interaction effect was found of personalization and scaffolding on the perception that the robot could have feelings, the impact in the personalization conditions was equal, and only a significant difference on social presence was found in the non-personalized conditions between the two scaffolding conditions. Further inspection indicated that personalization was positively related to children's awareness that the robot used things they said while chatting to create a math story (r = .399, p < .001) and, in turn, this 2nd manipulation question was also positively related to the perception that the robot could have feelings (r = .301, p = .004) (while not related to any of the other social perception items). Thus, it seems that only when children are aware of the fact that the robot uses memory-based personalization (similar to human interactions) do they believe that the robot can have feelings. However, further research is needed to convincingly support this theory.

A reason why the other items of social presence were not affected, could be because SP1-2-3 (i.e. "feels" like a real person, can see and understand me) can still be answered in the affirmative when the robot is seen as a (well-programmed) machine, while SP4-5 explicitly refer to humans as having feelings and being a living creature. Children varied more in their answers to SP4-5, as resembled in the lower mean scores on SP4-5 (compared to the scores on SP1-2-3). During the survey, we also observed that children were more skeptical about SP4-5, unsure whether to answer them affirmatively, and took longer to answer them. Some children even had difficulty understanding SP5 and asked the research assistant for further clarification, indicating that "living creature" can still be a challenging concept for some 8-11-year olds.

Consistent with [35, 64], memory-based personalization mimics human interactions, creating the feeling of having a conversation with someone who could be your friend. Using personal information makes the interaction partner more likable as humans enjoy talking about themselves, stimulating a willingness to continue the conversation [65]. Children in our study also mentioned out loud that next time they would also like to know more about the robot, thus, showing an interest in the background story of the robot, which is an indication of initial friendship formation [9].

A reason why the other items of feelings of friendship were not affected, is their high mean scores (M > 3.7 on a 4-point-scale). Feeling more comfortable around the robot, wanting to chat further with the robot, do more activities with the robot, and wanting to see the robot more often could all be the result of a novelty effect that is still present after 3 short exposures (within 1 week). In a follow-up study, we will test whether these scores remain high after longer exposure. Given the still developing cognitive abilities of middle childhood and the parasocial nature of child-robot relationships, our feelings of friendship scale contained more concrete behavioral measures. It would be interesting to add measures of trust and closeness, which are more commonly used in child-robot relationship formation studies [64].

Finally, an interesting overall effect was observed that when children appreciated the stories, they also evaluated the robot more positively on almost all social presence and feelings of friendships items. Strong positive emotions evoked by media content are often by children transferred to everything associated with that content (affect transfer) [17]. However, it remains unclear whether engaging stories turn a robot into a more likable and realistic storyteller or whether children more actively attend to stories told by a more socially relevant robot. Further research is needed to disentangle this process.

6.3 Limitations

The core focus of this study was to get insight about if and how the positioning of math in the social space of child-robot interaction can improve math performance. The results we found should be primarily treated as formative in nature rather than summative. Especially given the fact that this was a short-term study, but with multiple interactions. Another limitation was the lacking construct validity of social presence and not measuring math self-efficacy. In future work we plan to report more behavioral measures and have a more long-term perspective.

Furthermore, we did not compare our intervention with a non-robot control condition, making it hard to compare it to the current day practice. Our intervention was primarily based on a verbal interaction. Although a tablet was present to visually display the math problem, this might not be sufficient for every child. Exploring a more multimodal social learning set-up and comparing it to the current situation is left for future work.

7 CONCLUSION

We presented a novel, social constructivist, design for using an autonomous social robot in math education. The math is part of a personal conversation with the robot. We provide concrete design specifications of *how* the robot can personalize the conversation and keep the learning task in the Zone of Proximal Development by scaffolding. We furthermore provide a clear strategy, by using a storyworld, for creating a collection of connected minidialogs to populate the conversation. The results of a three-session experimental user study (n = 130, 8-11 y.o.) show that participants got better at math over time when the robot provided guidance. It is likely that not the educational content of the guidance, but rather the social support it offers explains the performance increase. This is best illustrated by children feeling the robot was more alive and more capable of having emotions when it provided guidance. Furthermore, the robot felt more as a friend when it personalized the conversation. The math dialogs were appreciated by the children, and this was even stronger when the dialogs were personalized. Collectively the results confirm that more strongly intertwining the robot's social behaviors with the math task and scaffolding leads to more effective and enjoyable learning.

ACKNOWLEDGMENTS

We would like to thank all the consortium partners of SOROCOVA, namely the schools, their teachers, and most importantly the children for contributing to this study. This research 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] Jeremy N. Bailenson, Jim Blascovich, Andrew C. Beall, and Jack M. Loomis. 2001. Equilibrium Theory Revisited: Mutual Gaze and Personal Space in Virtual Environments. *Presence: Teleoperators and Virtual Environments* 10, 6 (Dec. 2001), 583–598. https://doi.org/10.1162/105474601753272844

[3] Brian R. Belland. 2016. *Instructional scaffolding in STEM education: Strategies and efficacy evidence.* Springer International Publishing. https://doi.org/10.1007/978-3-319-02565-0

[4] 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.

[5] LaVonda N. Brown and Ayanna M. Howard. 2014. The positive effects of verbal encouragement in mathematics education using a social robot. In *2014 IEEE Integrated STEM Education Conference.* IEEE, Princeton, NJ, USA, 1–5. https://doi.org/10.1109/ISECon.2014.6891009

[6] Edgar Brunner, Sebastian Domhof, and Frank Langer. 2002. *Nonparametric analysis of longitudinal data in factorial experiments.* Vol. 373. Wiley-Interscience.

[7] Marilyn Burns and Robyn Silbey. 2000. *So you have to teach math?: Sound advice for K-6 teachers.* Math Solutions.

[8] Sandra L. Calvert, Bonnie L. Strong, Eliza L. Jacobs, and Emily E. Conger. 2007. Interaction and Participation for Young Hispanic and Caucasian Girls' and Boys' Learning of Media Content. *Media Psychology* 9, 2 (April 2007), 431–445. https://doi.org/10.1080/15213260701291379 Publisher: Routledge _eprint: https://doi.org/10.1080/15213260701291379.

[9] Lola Cañamero and Matthew Lewis. 2016. Making New "New AI" Friends: Designing a Social Robot for Diabetic Children from an Embodied AI Perspective. *International Journal of Social Robotics* 8, 4 (Aug. 2016), 523–537. https://doi.org/10.1007/s12369-016-0364-9

[10] Yi-Chen Chen, Edwinn Gamborino, Li-Chen Fu, Hsiu-Ping Yueh, and Su-Ling Yeh. 2021. Social Presence in Evaluations for a Humanoid Robot and Its Effect on Children-Robot Relationship. In *HCI International 2021 - Posters,* Constantine Stephanidis, Margherita Antona, and Stavroula Ntoa (Eds.). Springer International Publishing, Cham, 191–199.

[11] Caitlyn Clabaugh, Kartik Mahajan, Shomik Jain, Roxanna Pakkar, David Becerra, Zhonghao Shi, Eric Deng, Rhianna Lee, Gisele Ragusa, and Maja Matarić. 2019. Long-Term Personalization of an In-Home Socially Assistive Robot for Children With Autism Spectrum Disorders. *Frontiers in Robotics and AI* 6 (2019). https://www.frontiersin.org/articles/10.3389/frobt.2019.00110

[12] Caitlyn Clabaugh, Fei Sha, Gisele Ragusa, and Maja Mataric. 2015. Towards a Personalized Model of Number Concepts Learning in Preschool Children. Seattle, WA, USA, 26–30.

[13] Paul Cobb, Erna Yackel, and Terry Wood. 1992. A constructivist alternative to the representational view of mind in mathematics education. *Journal for Research in Mathematics education* 23, 1 (1992), 2–33. Publisher: National Council of Teachers of Mathematics.

[14] Kate Darling. 2015. 'Who's Johnny?' Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy. https://doi.org/10.2139/ssrn.2588669

[15] Kate Darling, Palash Nandy, and Cynthia Breazeal. 2015. Empathic concern and the effect of stories in human-robot interaction. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN).* 770–775. https://doi.org/10.1109/ROMAN.2015.7333675

[16] Simone M. de Droog and Moniek Buijzen. 2014. Presenting validated self-report character involvement scales for 4- to 6-year-olds. ICA, Seattle, WA.

[17] Simone M. de Droog, Moniek Buijzen, and Patti M. Valkenburg. 2014. Enhancing children's vegetable consumption using vegetable-promoting picture books. The impact of interactive shared reading and character–product congruence. *Appetite* 73 (Feb. 2014), 73–80. https://doi.org/10.1016/j.appet.2013.10.018

[18] 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.

[19] Lamia Elloumi, Marianne Bossema, Simone M. de Droog, Matthijs Smakman, Stan van Ginkel, Mike E.U. Ligthart, Kees Hoogland, Koen V. Hindriks, and Somaya Ben Allouch. 2022. Exploring Requirements and Opportunities for Social Robots in Primary Mathematics Education. In *31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN).* IEEE Press.

[20] Shalom M. Fisch. 2000. A Capacity Model of Children's Comprehension of Educational Content on Television. *Media Psychology* 2, 1 (Feb. 2000), 63–91. https://doi.org/10.1207/S1532785XMEP0201_4 Publisher: Routledge _eprint: https://doi.org/10.1207/S1532785XMEP0201_4.

[21] Hans Freudenthal. 2006. *Revisiting mathematics education: China lectures.* Vol. 9. Springer Science & Business Media.

[22] Ana Garcia-Melgar, Julianne East, and Noel Meyers. 2021. Peer assisted academic support: a comparison of mentors' and mentees' experiences of a drop-in programme. *Journal of Further and Higher Education* 45, 9 (Oct. 2021), 1163–1176. https://doi.org/10.1080/0309877X.2020.1851665 Publisher: Routledge _eprint: https://doi.org/10.1080/0309877X.2020.1851665.

[23] Eugene Geist. 2015. Math Anxiety and the "Math Gap": How Attitudes Toward Mathematics Disadvantages Students as Early as Preschool. *Education* 135, 3 (March 2015), 328–336.

[24] 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

[25] Charlotte Goddard. 2020. Hello, Mr Robot. *Nursery World Select* 2020, Sup6 (March 2020), 29–30. https://doi.org/10.12968/nuwa.2020.Sup6.29 Publisher: Mark Allen Group.

[26] Koen V. Hindriks and Sander Liebens. 2019. A Robot Math Tutor that Gives Feedback. In *Social Robotics (Lecture Notes in Computer Science),* Miguel A. Salichs, Shuzhi Sam Ge, Emilia Ivanova Barakova, John-John Cabibihan, Alan R. Wagner, Álvaro Castro-González, and Hongsheng He (Eds.). Springer International Publishing, Cham, 601–610. https://doi.org/10.1007/978-3-030-35888-4_56

[27] Cynthia Hoffner. 1996. Children's wishful identification and parasocial interaction with favorite television characters. *Journal of Broadcasting & Electronic Media* 40, 3 (June 1996), 389–402. https://doi.org/10.1080/08838159609364360 Publisher: Routledge _eprint: https://doi.org/10.1080/08838159609364360.

[28] 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.

[29] 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

[30] Brenda R. J. Jansen, Jolien Louwerse, Marthe Straatemeier, Sanne H. G. Van der Ven, Sharon Klinkenberg, and Han L. J. Van der Maas. 2013. The influence of experiencing success in math on math anxiety, perceived math competence, and math performance. *Learning and Individual Differences* 24 (April 2013), 190–197. https://doi.org/10.1016/j.lindif.2012.12.014

[31] 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

[32] Beaumie Kim. 2001. Social constructivism. *Emerging perspectives on learning, teaching, and technology* 1, 1 (2001), 16. Publisher: Beskikbaar.

[33] 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

[34] Jacqueline M. Kory-Westlund and Cynthia Breazeal. 2019. Exploring the Effects of a Social Robot's Speech Entrainment and Backstory on Young Children's Emotion, Rapport, Relationship, and Learning. *Frontiers in Robotics and AI* 6 (2019). https://www.frontiersin.org/articles/10.3389/frobt.2019.00054

[35] 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,* Adriana Tapus, Elisabeth André, Jean-Claude Martin, François Ferland, and Mehdi Ammi (Eds.). Springer International Publishing, Cham, 380–389.

[36] 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.

[37] 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

[38] Mike Ligthart and Rifca Peters. 2018. The Challenges of Evaluating Child-Robot Interaction with Questionnaires. In *What Could Go Wrong?! (HRI'18 Workshop).*

[39] 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.

[40] 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 (2022). https://www.frontiersin.org/articles/10.3389/frobt.2022.853665

[41] Matthew Lynch. 2016. Social constructivism in education. *The Edvocate* (2016), 14–32.

[42] Mariane Hedegaard. 2002. *The zone of proximal development as basis for instruction.* Routledge. https://doi.org/10.4324/9780203434185-13 Pages: 183-207 Publication

Title: An Introduction to Vygotsky.

[43] James C. McCroskey and Thomas A. McCain. 1974. The measurement of interpersonal attraction. *Speech Monographs* 41, 3 (Aug. 1974), 261–266. https://doi.org/10.1080/03637757409375845 Publisher: Routledge _eprint: https://doi.org/10.1080/03637757409375845.

[44] Chet Meyers and Thomas B Jones. 1993. *Promoting Active Learning. Strategies for the College Classroom.* ERIC.

[45] Terran Mott, Joslyne Lovelace, and Bennett Steward. 2021. Design Considerations for Child-Robot Interaction in Pediatric Contexts. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21 Companion)*. Association for Computing Machinery, New York, NY, USA, 252–256. https://doi.org/10.1145/3434074.3447170

[46] Kimihiro Noguchi, Yulia R. Gel, Edgar Brunner, and Frank Konietschke. 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

[47] Martin Nystrand, Adam Gamoran, Robert Kachur, and Catherine Prendergast. 1998. Opening Dialogue: Understanding the Dynamics of Language and Learning in the English Classroom. *Language* 74 (June 1998), 444. https://doi.org/10.2307/417942

[48] Penny Oldfather, Jane West, Jennifer White, and Jill Wilmarth. 1999. *Learning through children's eyes: Social constructivism and the desire to learn.* American Psychological Association, Washington, DC, US. https://doi.org/10.1037/10328-000 Pages: x, 125.

[49] Frank Pajares and M. David Miller. 1995. Mathematics self-efficacy and mathematics performances: The need for specificity of assessment. *Journal of Counseling Psychology* 42 (1995), 190–198. https://doi.org/10.1037/0022-0167.42.2.190 Place: US Publisher: American Psychological Association.

[50] Lena Pareto, Sara Ekström, and Wolmet Barendregt. 2019. Augmenting Game-Based Learning With a Robot Tutee. In *Proceedings of the 12th European Conference on Game Based Learning*. ACPI, 65. https://doi.org/10.34190/GBL.19.138

[51] James Pietsch, Richard Walker, and Elaine Chapman. 2003. The relationship among self-concept, self-efficacy, and performance in mathematics during secondary school. *Journal of Educational Psychology* 95 (2003), 589–603. https://doi.org/10.1037/0022-0663.95.3.589 Place: US Publisher: American Psychological Association.

[52] Sharon Pinkney and Gary Walker. 2020. 'It was me, but it was them that helped me': Exploring the issues for care experienced young people within higher education. *Children and Youth Services Review* 108 (Jan. 2020), 104576. https://doi.org/10.1016/j.childyouth.2019.104576

[53] Wolff-Michael Roth. 1999. Authentic School Science: Intellectual. *Learning & Knowledge* 3 (1999), 6. Publisher: Sage.

[54] Scott A. Schartel. 2012. Giving feedback – An integral part of education. *Best Practice & Research Clinical Anaesthesiology* 26, 1 (March 2012), 77–87. https://doi.org/10.1016/j.bpa.2012.02.003

[55] Dong-Hee Shin and Hyungseung Choo. 2011. Modeling the acceptance of socially interactive robotics: Social presence in human–robot interaction. *Interaction Studies* 12, 3 (Jan. 2011), 430–460. https://doi.org/10.1075/is.12.3.04shi Publisher: John Benjamins.

[56] Joana Galvão Gomes da Silva, David J. Kavanagh, Tony Belpaeme, Lloyd Taylor, Konna Beeson, and Jackie Andrade. 2018. Experiences of a Motivational Interview Delivered by a Robot: Qualitative Study. *Journal of Medical Internet Research* 20, 5 (May 2018), e7737. https://doi.org/10.2196/jmir.7737 Company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada.

[57] Reid Simmons, Maxim Makatchev, Rachel Kirby, Min Kyung Lee, Imran Fanaswala, Brett Browning, Jodi Forlizzi, and Majd Sakr. 2011. Believable Robot Characters. *AI Magazine* 32, 4 (Dec. 2011), 39–52. https://doi.org/10.1609/aimag.v32i4.2383 Number: 4.

[58] Adrian Treffers. 1987. *Three dimensions: A model of goal and theory description in mathematics instruction—The Wiskobas Project*. Vol. 3. Springer Science & Business Media.

[59] UNSDG. 2020. UNSDG | Policy Brief: Education during COVID-19 and beyond. https://unsdg.un.org/resources/policy-brief-education-during-covid-19-and-beyond,https://unsdg.un.org/resources/policy-brief-education-during-covid-19-and-beyond

[60] Dina Utami and Timothy Bickmore. 2019. Collaborative User Responses in Multiparty Interaction with a Couples Counselor Robot. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 294–303. https://doi.org/10.1109/HRI.2019.8673177 ISSN: 2167-2148.

[61] Marja Van den Heuvel-Panhuizen and Paul Drijvers. 2014. Realistic Mathematics Education. In *Encyclopedia of Mathematics Education*, Stephen Lerman (Ed.). Springer Netherlands, Dordrecht, 521–525. https://doi.org/10.1007/978-94-007-4978-8_170

[62] M Van Groenestijn, C Borghouts, and C Janssen. 2011. Protocol Ernstige RekenWiskunde-problemen en Dyscalculie (Protocol ERWD),(BAO, SBO, SO). (2011). Publisher: Assen: Koninklijke Van Gorcum/NVORWO.

[63] 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.

[64] Caroline L. van Straten, Jochen Peter, and Rinaldo Kühne. 2020. Child–Robot Relationship Formation: A Narrative Review of Empirical Research. *International Journal of Social Robotics* 12, 2 (2020), 325–344. https://doi.org/10.1007/s12369-019-00569-0

[65] Caroline L. van Straten, Jochen Peter, Rinaldo Kühne, and Alex Barco. 2022. On sharing and caring: Investigating the effects of a robot's self-disclosure and question- asking on children's robot perceptions and child-robot relationship formation. *Computers in Human Behavior* 129 (April 2022), 107135. https://doi.org/10.1016/j.chb.2021.107135

[66] Caroline L. van Straten, Jochen Peter, Rinaldo Kühne, and Alex Barco. 2022. The wizard and I: How transparent teleoperation and self-description (do not) affect children's robot perceptions and child-robot relationship formation. *AI & SOCIETY* 37, 1 (March 2022), 383–399. https://doi.org/10.1007/s00146-021-01202-3

[67] Paul Vogt, Rianne van den Berghe, Mirjam de Haas, Laura Hoffman, Junko Kanero, Ezgi Mamus, Jean-Marc Montanier, Cansu Oranç, Ora Oudgenoeg-Paz, Daniel Hernández García, Fotios Papadopoulos, Thorsten Schodde, Josje Verhagen, Christopher D. Wallbridge, Bram Willemsen, Jan de Wit, Tony Belpaeme, Tilbe Göksun, Stefan Kopp, Emiel Krahmer, Aylin C. Küntay, Paul Leseman, and Amit Kumar Pandey. 2019. Second language tutoring using social robots: a large-scale study. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI '19)*. IEEE Press, Daegu, Republic of Korea, 497–505.

[68] Lev Semenovich Vygotsky and Michael Cole. 1978. *Mind in society: Development of higher psychological processes.* Harvard university press.

[69] Keith Weber, Carolyn Maher, Arthur Powell, and Hollylynne Stohl Lee. 2008. Learning opportunities from group discussions: warrants become the objects of debate. *Educational Studies in Mathematics* 68, 3 (July 2008), 247–261. https://doi.org/10.1007/s10649-008-9114-8

[70] Astrid Weiss, Daniela Wurhofer, and Manfred Tscheligi. 2009. "I Love This Dog"—Children's Emotional Attachment to the Robotic Dog AIBO. *International Journal of Social Robotics* 1, 3 (Aug. 2009), 243–248. https://doi.org/10.1007/s12369-009-0024-4

[71] Trevor Williams and Kitty Williams. 2010. Self-efficacy and performance in mathematics: Reciprocal determinism in 33 nations. *Journal of Educational Psychology* 102 (2010), 453–466. https://doi.org/10.1037/a0017271 Place: US Publisher: American Psychological Association.

[72] David Wood, Jerome S. Bruner, and Gail Ross. 1976. The role of tutoring in problem solving. *Child Psychology & Psychiatry & Allied Disciplines* 17 (1976), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x Place: United Kingdom Publisher: Pergamon Press.