Thinking Aloud with a Tutoring Robot to Enhance Learning

Aditi Ramachandran
Yale University
aditi.ramachandran@yale.edu

Chien-Ming Huang
Johns Hopkins University
cmhuang@cs.jhu.edu

Edward Gartland
Greens Farms Academy
teddygartland@gmail.com

Brian Scassellati
Yale University
brian.scassellati@yale.edu

ABSTRACT

Thinking aloud, while requiring extra mental effort, is a metacognitive technique that helps students navigate through complex problem-solving tasks. Social robots, bearing embodied immediacy that fosters engaging and compliant interactions, are a unique platform to deliver problem-solving support such as thinking aloud to young learners. In this work, we explore the effects of a robot platform and the think-aloud strategy on learning outcomes in the context of a one-on-one tutoring interaction. Results from a 2x2 between-subjects study ($n = 52$) indicate that both the robot platform and use of the think-aloud strategy promoted learning gains for children. In particular, the robot platform effectively enhanced immediate learning gains, measured right after the tutoring session, while the think-aloud strategy improved persistent gains as measured approximately one week after the interaction. Moreover, our results show that a social robot strengthened students' engagement and compliance with the think-aloud support while they performed cognitively demanding tasks. Our work indicates that robots can support metacognitive strategy use to effectively enhance learning and contributes to the growing body of research demonstrating the value of social robots in novel educational settings.

KEYWORDS

Child-robot Interaction; Education; Tutoring

1. INTRODUCTION

A growing body of research has demonstrated the efficacy of social robots as effective tutoring agents in a variety of educational settings [18, 20, 23]. The physical presence and embodiment of robots help improve cognitive learning gains during tutoring [27] and increase user engagement, enjoyment, and compliance during instructional
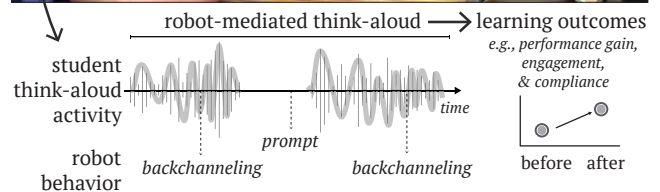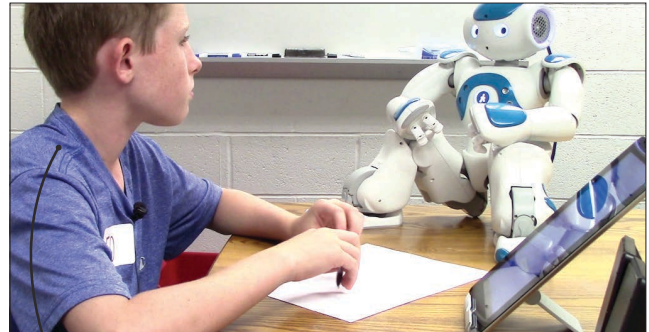
Figure 1: We studied how thinking aloud with a robot tutoring system can promote child learning during one-on-one tutoring interactions. We investigated the effects of both the think-aloud strategy and the robot platform on learning outcomes.

interactions [4, 7, 30, 39]. Moreover, social robots can personalize their support to address individual differences during learning, such as a user's affective state and cognitive capacity, to further enhance learning outcomes and experience [6, 16, 26, 34]. Despite the demonstrated potential of tutoring robots, little work has explored how these robots may support metacognitive strategies—which are particularly important for learning independence and efficiency—in learning and solving complex problems.

Thinking aloud—verbalizing one's thoughts during a cognitive task—is a metacognitve strategy that can aid students with complex reasoning tasks. Think aloud protocols are conventionally utilized as a technique for researchers to gain insight into a person's cognitive processes [8]. In tutoring interactions, thinking aloud has also been used to better understand a child's cognitive processes during educational tasks [37, 40]. More recently, teachers have been extending the use of thinking aloud as a specific problem-solving strategy for children, capitalizing on the idea that explicitly verbalizing one's thought process while trying to solve challenging problems may lead to a more deliberate and organized plan for complex reasoning [17]. However, prior research into the think-aloud method also suggests that this strategy may create additional cognitive load for the student, potentially negatively impacting performance [19, 42]. Young

students are perhaps more vulnerable as they often require close support in order to successfully utilize metacognitive strategies [2].

In this work, we investigate whether we can leverage the social presence and embodiment of a robot to foster engagement and compliance to effectively support young students' use of the thinking aloud strategy during a cognitively complex problem-solving task (Figure 1). We contextualized our investigation in solving "word" problems, requiring students to use critical reasoning skills to decide what mathematical operations to perform to arrive at an answer. We built a robot tutoring system capable of supporting children as they think aloud and conducted a 2x2 between-subjects user study to evaluate the use of both the think-aloud strategy and the robot platform on measures of learning, engagement, and compliance.

2. BACKGROUND

In this section, we review relevant work on robot tutoring interactions as well as the use of the thinking aloud strategy and its relationship to learning in various educational settings.

2.1. Robot Tutoring

Robot tutors have been successfully used to teach a variety of traditional subjects, such as math, reading, and language learning, as well as physical learning tasks, such as handwriting or physical exercises [10, 15, 18, 23, 28, 35]. Leyzberg et al. showed that the physical presence of a robot tutor can increase learning gains by demonstrating that adults who interacted with a physically embodied robot outperformed those who interacted with a video-based agent or a disembodied voice on a cognitive puzzle-solving task [27]. Moreover, Saerbeck et al. demonstrated that a robot tutor exhibiting socially supportive behaviors had a positive effect on student learning performance [38]. Furthermore, robot tutors can personalize their instruction and behavior to better support individual learners in one-on-one tutoring interactions [26]. Ramachandran et al. showed that a robot employing personalized strategies for when to provide a break to a child during a cognitively taxing math tutoring interaction could positively impact learning [34]. Leite et al. demonstrated the capabilities of a social robot that could adapt its support to the affective reactions of a child during a cognitive task [25]. Social robots have demonstrated a positive impact on a diverse set of educational applications, exhibiting the vast potential of robot tutoring to effectively promote student learning. Different from prior studies, in this work, we explore how robot tutors can support children engaging in the metacognitive learning strategy of thinking aloud during a challenging problem-solving task.

Within the complex process of learning, student engagement is a critical contributor to academic achievement [11]. Particularly for children, much of the promise of social robots as tutors is built on research indicating that embodied robots foster engagement during a learning interaction, thereby improving the potential to learn effectively. Prior HRI work has investigated the advantages of physically-embodied robots in establishing improved human-robot interactions [4, 32, 41]. Pereira et al. demonstrated increased enjoyment in an instructional game through interaction with an embodied robot as compared to an on-screen agent [30]. Studies have also demonstrated that physical embodiment can cause robots to be perceived as more engaging, credible, and informative than animated characters [24]. Kennedy et al. compared various embodiments of tutoring agents and found that robots contributed to increased social presence [21].

Furthermore, the social availability of a physical robot tutor was shown to benefit learners during a math-based tutoring scenario [22]. In this study, we seek to understand whether the social presence of a robot tutor impacts student engagement with thinking aloud during a tutoring interaction.

Though we know social robots can be effective tutors in many educational contexts, less attention has been given to exploring the use of social robots to support students with metacognitive learning strategies. Metacognitive strategies are important for effective learning and academic success [12]; however, they are difficult for young children to successfully use without support [3, 31]. In this work, we investigate the impact of a social robot tutor that supports students engaging in a novel metacognitive strategy. Furthermore, we explicitly examine the effects of the robot platform through which we deliver the strategy support to further understand the potential for embodied social robots as effective tutoring agents.

2.2. Thinking Aloud

Thinking aloud refers to verbalizing one's thoughts out loud while completing a task. Think aloud protocols were originally developed as a tool for researchers to understand a subject's cognitive processes while engaged in a cognitive activity [8, 43]. In the educational domain, the think-aloud method has been used to gain understanding of how children of different ability levels cognitively solve math problems [29, 37]. The think-aloud method has also been used to elicit reflection in a concept learning task with adults to assess their metacognitive skill use [5].

There are many online resources[1,2] to support teachers' exploration of metacognitive strategy use to help their students learn. Teachers have recently begun to explore an innovative use of thinking aloud as an explicit problem-solving strategy that may improve performance, in which students' verbalizations potentially lead to more carefully planned problem-solving steps [17]. Older adults demonstrated significant performance improvements on an abstract reasoning task while thinking aloud [13]. Thinking aloud also positively impacted performance for children who engaged in the strategy as they completed verbal and spatial analogies [40]. Furthermore, students who explained their steps while solving geometry problems demonstrated greater understanding of the material as compared to those who did not [2].

Although thinking aloud appears to be a promising metacognitive strategy to explore, prior work involving thinking aloud indicates that use of the strategy may become difficult when the task at hand is demanding [36]. The additional cognitive load for the user already engaged in a problem-solving task may slow down the user or negatively impact performance [19]. During an information search task, adults who concurrently engaged in a think-aloud task demonstrated lower task performance than those who did not think aloud [42]. Younger students might be particularly susceptible to this negative impact during an already challenging problem-solving task. In this work, we seek to explore how a social robot can provide close support to students to utilize the potential of the think-aloud strategy in solving complex problems effectively.

---

3. THINKING ALOUD WITH A ROBOT TUTORING SYSTEM

In this section, we provide a detailed description of the design of an autonomous robot tutoring system capable of supporting students in thinking aloud during a learning interaction. We also present the strategies employed by the interactive system to actively encourage and respond to think-aloud behavior during tutoring.

3.1. System Overview

Our robot tutoring system consisted of a Nao robot as a tutoring agent and several key software components including the content manager, the voice activity monitor, and the behavior planner (Figure 2). The system and each of these components were implemented as part of a ROS architecture [33]. The content manager is responsible for starting the session with a short interactive lesson activity followed by up to 12 practice questions. This component manages the content of the interaction by leveling up the difficulty of the questions after the student completes four questions of a given difficulty level. The voice activity monitor uses *openSMILE*, an open source audio feature extraction tool [9], to automatically detect voice activity, and implements a speaking binary node that outputs a stream of zeros and ones to represent binary detection of a student's voice during the tutoring session. Without advanced natural language understanding, the behavior planner uses the continuous stream of speaking binary to decide when to provide certain behaviors that support the child's think-aloud activity.

3.2. Design of Think Aloud Support

Traditional use of think-aloud protocols indicates that subjects need to be instructed, reminded, and prompted to engage in the thinking aloud exercise [43]. To support children using a think-aloud strategy during a tutoring interaction, we included the following robot behaviors in our tutoring system. The implementation of these behaviors were informed by a pilot exploration study involving seven students thinking aloud while problem solving.

Remind— As students are not typically familiar with the think-aloud strategy, the system provides a reminder to think aloud each time a new exercise in the problem-solving task is displayed. An example reminder given by the robot is "When you are doing the question, remember to say everything out loud."

Prompt— We noticed from the exploration study that students periodically forget to think aloud when concentrating on the math problems, indicating that prompting students to continue talking out loud is necessary during a challenging problem-solving task. We observed that students would talk continuously for small periods of time lasting 5.63 seconds on average ($SD = 2.22$). However, they also paused frequently as they talked out loud while doing problems, leaving gaps between speech of 4.25 seconds on average ($SD = 2.85$). Based on these observations, we designed our robot's prompting behavior to trigger after approximately 6 seconds of detected silence, according to a dynamic sampling from a normal distribution with $M = 6.00$ and $SD = 2.00$, in order to avoid frequently interrupting students pausing to think. The robot would give prompts such as "Keep talking out loud!" or "Don't forget to think aloud!".

Reflect— When students make an incorrect attempt, the system will instruct them to reflect on why their answer was wrong and to think out loud while doing this. For example, the robot would say
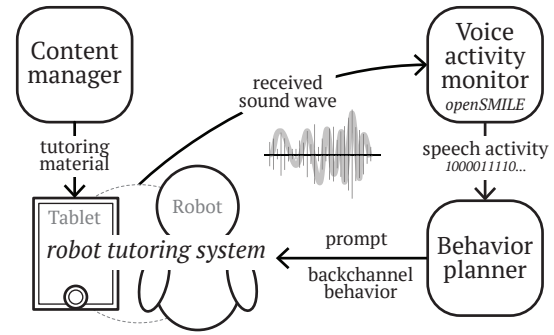

Figure 2: System architecture of our robot tutoring system, capable of supporting students thinking aloud during problem-solving.

"Reflect on why you might have gotten the problem wrong. Make sure to think aloud as you do this."

In addition to the above supporting behaviors of the think-aloud exercise, we designed a basic backchanneling behavior conveying that the tutoring system can hear whether the child is talking. The tutoring system actively tracked the child's voice activity using *openSMILE* and constrained voice activity detection to either "talking" or "not talking." Our backchanneling behaviors involved simple nodding motions that occurred regularly during continuous speech. In particular, the robot demonstrated backchanneling after detecting approximately 2.5 seconds of talking according to a dynamic sampling from a normal distribution with $M = 2.50$ and $SD = 1.00$. This design choice is to show nods approximately twice during an average length utterance.

3.3. Mechanisms of Tutoring Application

In addition to the above behaviors that support the think-aloud protocol, our tutoring system included a tablet application that provided several basic mechanisms to dictate the flow of a tutoring interaction. The tablet application displayed all the necessary information on its screen and was used as an input device to enter answers during the tutoring task. At the start of each tutoring session, students completed a short, interactive lesson on a strategy for solving certain math problems (see strategy steps in Table 1). After the lesson, the tablet displayed questions one at a time for the student to answer. Feedback about correct and incorrect answers was displayed on the tablet screen after each answer attempt. After two incorrect attempts on a given problem, students would see feedback on the tablet while the tutoring agent employed the strategy taught at the beginning of the interaction to provide an explanation of the correct answer. These mechanisms applied to all versions of the tutoring system regardless of the various experimental conditions described in Section 4.2. Aside from the basic tutoring mechanisms, the robot, serving as a tutoring agent, displayed simple interactive behaviors, including looking towards the student when talking and towards the tablet when the student was working on a problem, as well as extending its arm towards the tablet while instructing the student that the next problem would appear on the tablet screen.

4. METHODS

In this section, we describe a user study exploring the effects of a robot tutoring system that supports thinking aloud, as described in Section 3, on student learning outcomes.

Table 1: An example of a practice math problem given to the students during the tutoring session. The left column contains the steps for solving word problems presented during the initial lesson activity. Also shown are examples of participants' think-aloud utterances that align with the steps.

Steps for Solving Word Problems	Example
#1. Read the problem.	Samantha wants to put solar panels on the roof of her house. Her roof is a flat rectangle that is 8 feet long and 10 feet wide. If each solar panel is 4 square feet, how many panels will she need to cover her roof?
#2. Figure out what information the problem gives you.	"*her roof is 8 feet long and 10 feet wide and her roof is a flat rectangle... each solar panel is 4 square feet*" (P10)
#3. Ask yourself what the problem wants you to find and what strategies you can use.	"*So we have to find out how big her roof is...if each solar panel is four square feet so we have to divide*" (P13)
#4. Make a plan for what to do to find the answer.	"*So the roof is 80 ft...if each solar panel is four square feet, we have to divide eighty by four...twenty*" (P13)

4.1. Evaluation Context

Our user study involved a math-based tutoring interaction in which children learned a multi-step problem-solving approach for solving *word problems*, were guided to work out one example problem step-by-step, and then completed practice exercises. Word problems refer to math problems that require students to read the problem and apply some critical reasoning skills to determine how relevant mathematical concepts could be applied to the problem at hand (Table 1). Children typically struggle with this type of problem-solving [14]. In particular, as the number of steps required to complete each problem increases, students often feel confused and simply combine numbers mentioned in the problem to guess an answer.

We designed a total of 12 multi-step word problems on area and perimeter, which are concepts that students have learned in school but have not frequently encountered within the context of word problems. To ensure the appropriateness of the concepts and difficulty of the problems used in the study, we validated the problems with a local public school teacher who has years of experience teaching children in our targeted age range and grade level.

4.2. Experimental Design

We designed a 2x2 between-subjects study involving two independent variables that each contain two levels: the *platform* through which tutoring support is delivered (robot vs. no robot), and the use of the thinking aloud *strategy* during problem solving (think-aloud vs. no think-aloud). Students received the same educational content regardless of experimental condition. Below are the four conditions.

Robot&ThinkAloud— In this condition, we implemented the tutoring system as detailed in Section 3. This condition includes using a robot as the platform to provide tutoring intervention. Students in this condition were also explicitly instructed by the robot to think aloud and received reminders and prompts to do so throughout the tutoring interaction.

Robot-Only— In this condition, students interacted with a robot tutoring agent throughout the interaction; however, there were no think-aloud instructions, prompts, or reminders for them during the session. The robot still served as the tutoring agent and displayed the tutoring support mechanisms detailed in Section 3.3.

ThinkAloud-Only— Students in this condition received their tutoring support without the presence of the robot. Students were provided with verbal instructions, prompts, and reminders to think aloud from the tablet. To signal "backchanneling" behavior conveying listening awareness, we implemented a dynamic circle that varies its size depending on received voice activity.

Baseline— Students completed their tutoring session without the presence of the robot as well as without the use of the think-aloud strategy. This condition simulates the scenario in which students would use a typical tutoring application on a tablet.

4.3. Experimental Procedure

Prior to participation, parental and child consent forms were collected for each student. Children were informed that they were allowed to stop the experiment at any point without any repercussions. An experimenter escorted children from their classroom one at a time to participate in the study for approximately one hour. Prior to interacting with the tutoring system, students completed a pretest consisting of six word problems to assess prior knowledge. They were then randomly assigned to one of the four experimental conditions and interacted with the tutoring system for approximately 30 minutes regardless of experimental condition.

Students sat at a table facing the tutoring system that included a tablet and speakers in all conditions, and a robot in the *Robot* conditions (Figure 1). Each child participated in a completely autonomous interaction with the tutoring system, where no input from experimenter was required during the tutoring session. After the interaction, children completed a posttest assessment to measure their knowledge of the concepts learned. After the posttest, students then completed a short questionnaire about their interaction experience with the tutoring system and were given a pencil and a sticker for participating in the study. Approximately one week after the interaction, students also completed a follow-up posttest assessment to measure sustained performance after several days. The pretest, posttest, and follow-up assessments were identical and consisted of the same six questions that were each a word problem involving the concepts of area or perimeter.

4.4. Measures

To evaluate the benefits of the use of the think-aloud strategy as well as the platform through which the tutoring support was delivered on both learning outcomes and student behavior during the think-aloud activity, we employed several objective measures involving (1) learning gains, (2) engagement, and (3) compliance. To measure learning gains, we used normalized learning gain (*nlg*) between two test scores, which is defined as follows for an individual student *i*:

$$nlg(i) = \begin{cases} \dfrac{score_{\text{post}}(i) - score_{\text{pre}}(i)}{1 - score_{\text{pre}}(i)} & \text{if } score_{\text{post}} \geq score_{\text{pre}} \\[2ex] \dfrac{score_{\text{post}}(i) - score_{\text{pre}}(i)}{score_{\text{pre}}(i)} & \text{if } score_{\text{post}} < score_{\text{pre}} \end{cases}$$

Measuring *nlg*, which ranges from -1.0 to 1.0, captures normalized change between individual test scores, and provided us with a single metric of improvement for each individual that accounts for varying prior knowledge levels by measuring improvement relative to each student's pretest accuracy. The measure of *score* itself is not a normalized value, but rather a measure of accuracy for an individual student on a given test (pretest, posttest, or follow-up) calculated by dividing the number of questions answered correctly by the total number of questions on the test. We analyzed *nlg* from pretest to posttest as well as *nlg* from pretest to follow-up to understand our system's effects on immediate learning outcomes as well as those that remain several days after the tutoring session.

In addition to evaluating learning outcomes, we sought to understand how the platform through which the tutoring support (e.g., thinking aloud) is delivered can impact children's engagement in and compliance with the intended support. To quantify children's engagement in the thinking aloud exercise, we derived two measures—the percentage of time students talked during the tutoring session and the number of prompts needed to keep the students thinking aloud. The percentage of time talking, extracted automatically from the logged openSMILE voice activity data, was calculated by dividing the amount of time that talking was detected over the total time students were given the opportunity to be talking out loud during the problem-solving task. This measure excludes times when the tutoring agent was talking, prompting, or giving any instructions intermittently throughout the session. We interpret a higher percentage of talking during the tutoring sessions as higher engagement in the tutoring support, as it signifies active utilization of the think-aloud strategy over the duration of the tutoring session. On the other hand, fewer prompts needed to have the students continue talking indicates higher engagement.

To assess students' compliance with the tutoring support, we calculated the number of prompts the students ignored during the session. We define an ignored prompt as a prompt that goes unanswered due to a lack of voice activity detection and subsequently triggers an additional prompt due to the prolonged silence. Fewer ignored prompts indicates higher compliance with the support.

4.5. Participants

We recruited 53 participants from local middle schools to participate in this study. We excluded one participant from this data analysis due to a perfect pretest score, resulting in 13 participants in each experimental group. The included 52 participants were comprised of 14 females and 38 males that were gender-balanced across groups. The majority of the students in this study were in sixth grade, with the average age being 11.21 years old ($SD = .89$).

5. RESULTS

In this section, we first our present findings on how students progressed over the course of the tutoring session, which informs our further analysis. We then present results characterizing the learning outcomes of students across our experimental groups (Figures 3 and 4) as well as differences in engagement and compliance behaviors between our two *ThinkAloud* conditions (Figure 3). We used analysis of variance (ANOVA) tests when comparing all groups and t-tests when directly comparing the two *ThinkAloud* conditions. We used non-parametric statistical tests when appropriate and an α level of .05 for significance in our analysis.

5.1. Characterization of Learning Progress

Students completed up to 12 practice problems during their tutoring session, limited by an overall time limit to ensure students spent approximately the same amount of time with the system in all conditions. Students in the four experimental groups progressed through the exercises comparably, with no significant differences across groups in the number of the problems they were able to complete during their session (Robot&ThinkAloud: $M = 10.85, SD = 2.34$; Robot-Only: $M = 10.54, SD = 2.07$; ThinkAloud-Only: $M = 9.62, SD = 2.67$; Baseline: $M = 10.23, SD = 2.20$).

To explore the broader benefits of the tutoring system, we administered follow-up tests to examine students' persistent performance after several days following the session. In this exploration, we observed that many students improved their problem-solving performance on the follow-up test rather than on the posttest (Figure 4).

5.2. Learning Gains

To evaluate the effect of our two independent variables—platform and strategy—on learning outcomes, we first compared normalized learning gain, *nlg*, from pretest to posttest using a two-way ANOVA test (Robot&ThinkAloud: $M = .37, SD = .49$; Robot-Only: $M = .39, SD = .40$; ThinkAloud-Only: $M = .19, SD = .50$; Baseline: $M = -.02, SD = .31$). This analysis revealed a significant main effect of platform (robot or no robot) on *nlg* from pretest to posttest, $F(1, 48) = 6.785, p = .012, \eta^2 = .120$. Students who interacted with the robot improved from pretest to posttest (*nlg*: $M = .38$, $SD = .41$) significantly more than those who did not interact with the robot platform (*nlg*: $M = .08, SD = .42$), suggesting the benefit of using a robot as a platform to deliver tutoring support. There was no significant main effect of strategy (think-aloud or no think-aloud) on *nlg* from pretest to posttest ($F(1, 48) = .716, p = .402, \eta^2 = .013$), nor was there a significant interaction effect, $F(1, 48) = 1.139$, $p = .291, \eta^2 = .020$.

As informed by our observations of students' improvement on the follow-up test, we compared *nlg* from pretest to follow-up (Figure 3) using a two-way ANOVA to compare the learning gains from before the tutoring session to approximately one week after the session (Robot&ThinkAloud: $M = .52, SD = .39$; Robot-Only: $M = .44, SD = .32$; ThinkAloud-Only: $M = .39, SD = .36$; Baseline: $M = .09, SD = .11$). We found a significant main effect of platform on *nlg* from pretest to follow-up ($F(1, 48) = 7.350, p = .009, \eta^2 = .119$), which shows that students who interacted with the robot
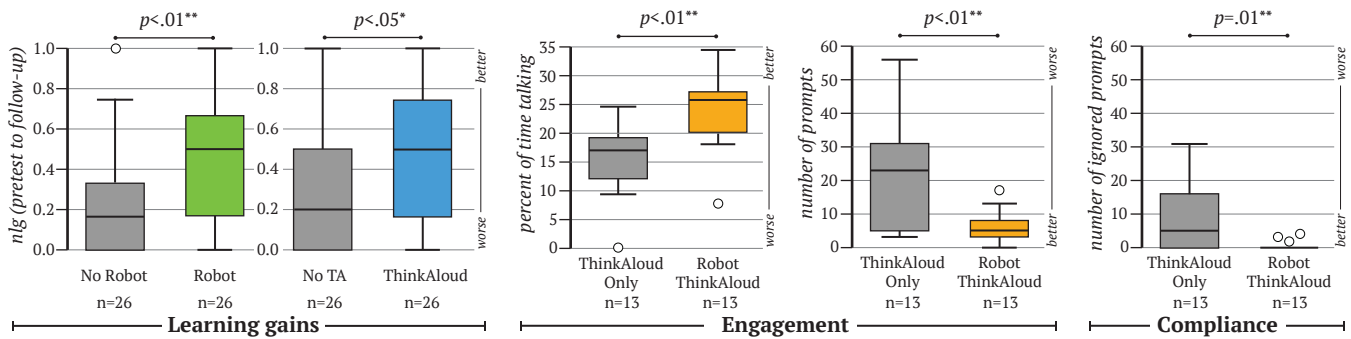
Figure 3: Our results show the benefits of thinking aloud with an interactive robot system. *Learning gains*: Both the robot and the think-aloud strategy led to improved learning from pretest to follow-up. *Engagement*: students talked significantly more and required fewer prompts to continue talking when thinking aloud with the robot. Higher engagement corresponds to more talking and fewer prompts. *Compliance*: Students ignored fewer prompts to talk out loud when thinking aloud with the robot. For all boxplots, the line inside the box represents the median and the extents of the box are the first and third quartiles.

platform ($M = .48$, $SD = .35$) improved significantly more than those who did not ($M = .24$, $SD = .30$). Additionally, we also found a significant main effect of strategy on *nlg* from pretest to follow-up ($F(1, 48) = 4.743$, $p = .034$, $\eta^2 = .077$), indicating that the mean *nlg* was significantly higher for those who engaged in the think-aloud strategy ($M = .45$, $SD = .37$) than those who did not utilize the think-aloud strategy ($M = .27$, $SD = .29$). However, there was no significant interaction effect of platform and strategy on *nlg* from pretest to follow-up, $F(1, 48) = 1.549$, $p = .219$, $\eta^2 = .025$.

We further investigated each experimental group separately to understand which groups demonstrated *immediate* learning gains from pretest to posttest and *persistent* learning gains from posttest to follow-up. We performed Wilcoxon Signed Ranks tests to evaluate the differences between consecutive pairs of test scores (see Figure 4 for a visual representation of these results). Students in the robot conditions, including both the Robot&ThinkAloud group and the Robot-Only group, significantly improved their test scores from pretest to posttest. In contrast, students in the ThinkAloud-Only group and the baseline group showed no significant difference between pretest and posttest scores. We further analyzed the differences in performance from posttest to follow-up. In this analysis, we excluded participants who achieved a perfect score on the posttest across all groups due to no improvement being possible for these students. Our analysis revealed that students who engaged in the think-aloud activity, including both the Robot&ThinkAloud group and the ThinkAloud-Only group, showed significant improvements on their test scores between the posttest and follow-up. However, the Robot-Only group did not show additional improvements between the posttest and follow-up, nor did the baseline group.

Taken together, the observed improvements from pretest to posttest for those who interacted with the robot, as well as the improvements from posttest to follow-up for those who engaged in the think-aloud activity, indicate the potential of both the robot platform and the think-aloud strategy on learning outcomes. Moreover, the Robot&ThinkAloud group demonstrated both immediate (from pretest to posttest) and persistent (from posttest to follow-up) improvements, suggesting the promise of a robot in reinforcing metacognitive strategies, particularly thinking aloud, in an educational application.

5.3. Engagement
Engagement is critical to student learning and achievement [11]. Below, we report our findings on two measures of student engagement to further explore how a robot's social presence impacts engagement during a metacognitive educational task.

5.3.1. Percent of Time Talking
The percent of time students talked during the tutoring session is an approximate measure of their engagement in the think-aloud tutoring activity. We conducted a two-way ANOVA to measure the effects of our two independent variables on the percentage of time students talked during the tutoring interaction (Robot&ThinkAloud: $M = 23.77\%$, $SD = 6.68\%$; Robot-Only: $M = 2.60\%$, $SD = 5.47\%$; ThinkAloud-Only: $M = 15.81\%$, $SD = 6.70\%$; Baseline: $M = 4.31\%$, $SD = 6.49\%$). A significant main effect of strategy demonstrated that students in the *ThinkAloud* groups ($M = 19.79\%$, $SD = 7.71\%$) talked significantly more than those who completed the tutoring interaction without the thinking aloud activity ($M = 3.46\%$, $SD = 5.94\%$), $F(1, 48) = 85.892$, $p < .001$, $\eta^2 = .594$. This is expected as students do not typically talk out loud very frequently unless explicitly instructed to do so. This result confirms that students who participated in the think-aloud exercise actively engaged in the task and talked out loud more frequently than those who were not given think-aloud instructions or support.

While no significant main effect of platform on percent time talked was found ($F(1, 48) = 3.136$, $p = .083$, $\eta^2 = .022$), the test revealed a significant interaction effect between strategy and platform on percent time talked, $F(1, 48) = 7.523$, $p = .009$, $\eta^2 = .052$. This result indicates that the effect of the think-aloud strategy on how much students talk differs based on the platform. A simple effects test between the Robot&ThinkAloud group and the ThinkAloud-Only group (Figure 3) showed that students talked out loud significantly more when thinking aloud with the robot ($M = 23.77\%$, $SD = 6.68\%$) than when thinking aloud without the robot ($M = 15.81\%$, $SD = 6.70\%$), $p = .002$.

5.3.2. Prompts to Think Aloud
In both the Robot&ThinkAloud and the ThinkAloud-Only conditions, the tutoring system prompted the students to continue thinking
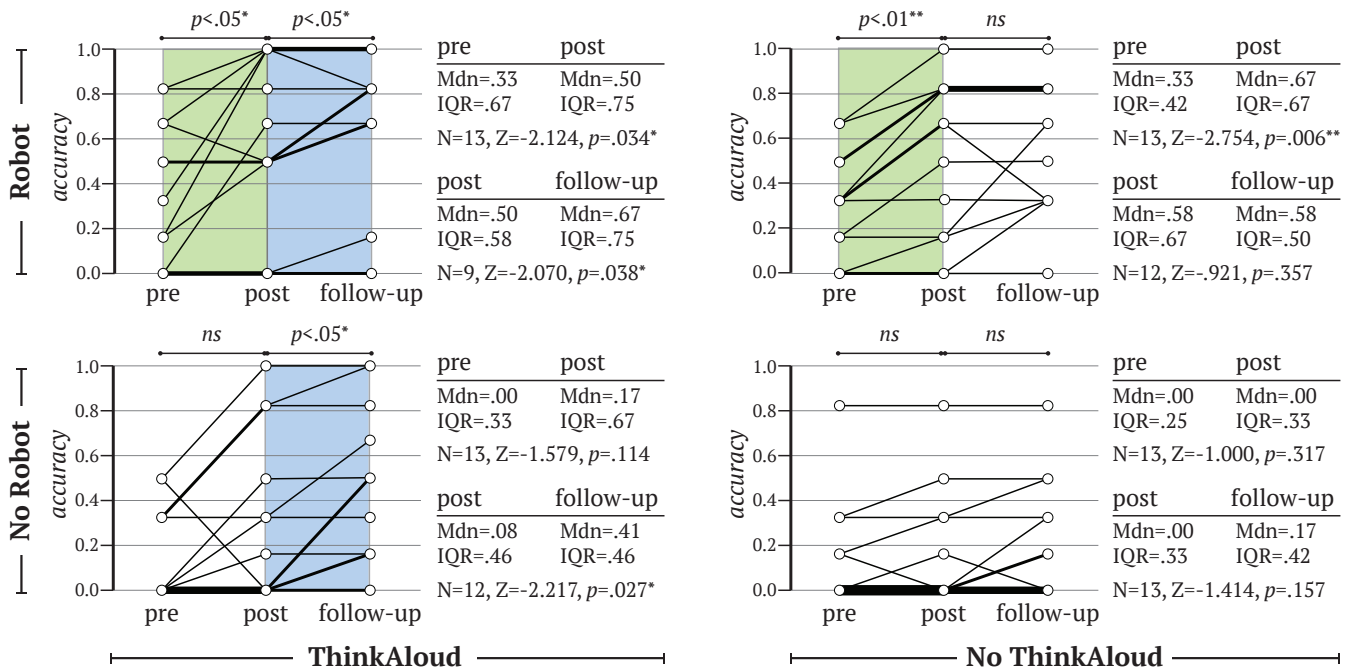
Figure 4: Pretest, posttest, and follow-up scores for each student, separated by experimental condition. Thicker lines indicate multiple participants with the same scores. Students that interacted with the robot improved their scores significantly between pretest and posttest (shaded in green) regardless of think-aloud strategy use. Students that engaged in the think-aloud strategy improved their scores from posttest to follow-up (shaded in blue) regardless of platform. The Robot&ThinkAloud group showed both immediate and persistent learning gains. (*) and (**) denote $p < .050$ and $p < .010$, respectively.

aloud when periods of silence were detected. Here, we report the comparison of how often these prompts were triggered between the two think-aloud conditions (Figure 3). An independent samples t-test showed that students who interacted with the robot triggered fewer prompts to continue thinking out loud ($M = 6.08$, $SD = 4.70$) as compared to those who completed the activity without the robot ($M = 22.23$, $SD = 17.84$), $t(24) = −3.156$, $p = .004$, $d = 1.24$. This finding indicates that students needed fewer reminders to stay actively engaged in the think-aloud exercise when interacting with the robot. We speculate that the social presence and embodiment of the robot might contribute to such active engagement.

5.4. Compliance

Effective tutoring agents need to foster student compliance with the educational strategies they deliver during the interaction. Here we report the students' compliance with the think-aloud exercise, as measured by the number of prompts that were ignored by the students (Figure 3). An independent samples t-test showed that students in the Robot&ThinkAloud condition ignored significantly fewer prompts ($M = .69$, $SD = 1.37$) than students in the ThinkAloud-Only condition ($M = 9.23$, $SD = 10.97$), $t(24) = −2.784$, $p = .010$, $d = 1.09$. Students interacting with the robot tutoring agent complied with the request to continue talking more often than those completing the think-aloud exercise without the robot. Furthermore, we see that students who did the think-aloud activity with the robot were almost fully compliant, as the average number of ignored prompts across students in this group was close to zero. This high level of

compliance with the robot's requests further indicates the effectiveness of the robot platform in supporting students' megacognitive strategy use that may be difficult for them.

6. DISCUSSION

In this work, we explore the effects of two variables—the use of a metacognitive learning strategy and the platform through which the tutoring support is delivered—on student learning outcomes during a tutoring task. We found that students benefited from both interaction with the robot tutoring platform as well as from engaging in the think-aloud strategy during problem-solving. We also observed that during the think-aloud exercise, the robot fostered increased engagement and compliance, two important ingredients for achieving effective tutoring. Our findings further highlighted two phases of learning improvements, as we observed the robot's impact on immediate learning gains and the think-aloud strategy's effect on persistent gains measured a week after the tutoring session.

6.1. Using Robots to Support Thinking Aloud

Our results showed that students who interacted with the robot tutoring platform outperformed those who did not, as indicated by normalized learning gain from pretest to follow-up (Figure 3). Moreover, students who interacted with a robot improved their performance from pretest to posttest, demonstrating immediate learning gains after a single tutoring session. We speculate that this immediate benefit may come from the embodied social presence that the tutoring robot fostered during the interaction, as empirical evidence has suggested various positive influences of the perceived social

presence of robots on human-robot interactions [1, 21, 24]. Learning improvement from posttest to follow-up, however, was not observed in the Robot-Only group, yet their performance did not drop either, suggesting that the problem-solving skills they improved during the tutoring session remained when measured several days later.

The use of the think-aloud strategy did not lead to the same immediate benefits as the robot platform, as students in the ThinkAloud-Only group did not show immediate learning gains from pretest to posttest. Due to the additional mental effort needed for the think-aloud activity, some students may have experienced increased cognitive load during the tutoring exercise. The intense cognitive burden, coming from both the think-aloud activity and the problem-solving task, could have potentially drained students' attention and patience in completing the posttest assessment. However, students who completed the think-aloud activity demonstrated learning improvements from posttest to follow-up, indicating that they were able to demonstrate improved problem-solving performance after receiving a cognitive break of a few days. Though it took longer for the benefits to become observable, these learning improvements for students in the ThinkAloud groups showed that the metacognitive strategy of thinking aloud did help students' problem-solving skills.

Students in the Robot&ThinkAloud group demonstrated both immediate and persistent learning gains, indicating that robot tutoring agents are a promising platform through which to deliver metacognitive strategy support for children. One possible explanation of these observed gains is that students in the Robot&ThinkAloud condition had a social entity to direct their thinking aloud towards, as if they were engaged in a regular talking activity. For example, one student frequently referenced the robot while thinking aloud: "*then you multiply seven times seven* [looks toward robot] *equals forty-nine* [looks at robot again]*, right?*" In contrast, when there was no social entity to talk to, students might have had to deliberately carry the cognitive burden of thinking aloud.

We also observed that students in the baseline condition did not demonstrate large improvements in either of the two phases (Figure 4). This may be because of the cognitively taxing problem-solving interaction they completed, without any added strategy or platform that increased the engagement or novelty of the task at hand. This highlights the need to continue exploring novel technological interaction paradigms for children in tutoring settings.

6.2. Quality of Thinking Aloud

Though we did not conduct a full analysis of the content of each child's speech during the think-aloud exercise, we observed variety in the quality and content of students' utterances when thinking aloud. For example, some students were clear in their ability to plan and execute their problem-solving steps. One participant quickly deduced that the problem was asking about perimeter: "*So we need to find out the perimeter, because they said we need to find out the distance he needs to walk around the building.*" Another student demonstrated organized steps to find the perimeter: "*Oh, we have to find the perimeter...so we have to multiply ten times two, which is twenty, and then thirty times two, which is sixty, then add those two together, and that would be sixty plus twenty...eighty.*" Others were less organized in their reasoning and often started talking about numbers without planning: "*Eight minus four is four, twelve minus eight...ok, it's forty four.*"

Follow-up analysis should be conducted to understand whether the content or quality of the think-aloud utterances differ as a result of the platform through which the strategy was supported. There are many factors that may contribute to differences in the quality of the think-aloud content, including prior abilities, personality traits, and academic confidence. In future robot tutoring interactions, building more comprehensive student models and responding to these differences should be explored.

6.3. Limitations and Future Work

Though this paper demonstrated the benefits of our interactive robot tutoring platform that supported children in their use of the think-aloud strategy, our work has limitations that should be addressed in future work. Firstly, we compared our embodied robotic system as an entire entity in our analysis of how platform affected outcomes of a tutoring interaction. As children likely have varying preferences for different appearances and behaviors of a robot tutor, future research should tease apart more specific design considerations that may impact learning and engagement during tutoring. Secondly, while our study included a follow-up assessment which broadened our understanding of student learning outcomes in this context, future research should investigate robots supporting these strategies over longer periods of time to assess the impact on longer-term learning gains. Thirdly, our robot tutoring system used real-time voice activity detection to interactively prompt and backchannel during tutoring; however, this voice activity detection was limited to talking and silence. To leverage the think-aloud strategy effectively, future tutoring systems should work towards intelligent understanding of the students' think-aloud dialogue to provide timely interventions that can help students to prevent mistakes or flawed solution paths. Finally, our robot tutoring system supported children using a particular metacognitive strategy—thinking aloud—in one educational domain: multi-step math word problems. To fully understand how metacognitive strategy use can benefit learning, robot tutoring systems should explore the transfer of metacognitive strategy use to other educational domains as well.

7. CONCLUSION

Our work is among the first to explore the use of robot tutors as providers of support for children engaging in a metacognitive strategy. We present empirical evidence showing the benefits of both a robot tutoring platform and use of the think-aloud strategy on student learning outcomes. Our analysis highlights two phases of learning improvements: the physically embodied robot tutor fostered immediate learning benefits, while the think-aloud strategy's positive impact on learning took longer to become observable, signifying potential longer-term benefits. We also found that students completing the think-aloud exercise engaged and complied with the support more effectively when it was delivered through the robot tutoring platform. Our work reinforces the promise of social robot tutors to support children with metacognitive strategy use in challenging learning environments.

8. ACKNOWLEDGMENTS

References

[1] Sigurdur O Adalgeirsson and Cynthia Breazeal. 2010. MeBot: a robotic platform for socially embodied presence. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*. IEEE Press, 15–22.

[2] Vincent Aleven and Kenneth R Koedinger. 2002. An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive science* 26, 2 (2002), 147–179.

[3] Vincent Aleven, Bruce Mclaren, Ido Roll, and Kenneth Koedinger. 2006. Toward meta-cognitive tutoring: A model of help seeking with a Cognitive Tutor. *International Journal of Artificial Intelligence in Education* 16, 2 (2006), 101–128.

[4] Wilma A Bainbridge, Justin W Hart, Elizabeth S Kim, and Brian Scassellati. 2011. The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics* 3, 1 (2011), 41–52.

[5] Maria Bannert and Christoph Mengelkamp. 2008. Assessment of metacognitive skills by means of instruction to think aloud and reflect when prompted. Does the verbalisation method affect learning? *Metacognition and Learning* 3, 1 (2008), 39–58.

[6] Paul Baxter, Emily Ashurst, Robin Read, James Kennedy, and Tony Belpaeme. 2017. Robot education peers in a situated primary school study: Personalisation promotes child learning. *PLoS One* 12, 5 (2017), e0178126.

[7] Vijay Chidambaram, Yueh-Hsuan Chiang, and Bilge Mutlu. 2012. Designing persuasive robots: how robots might persuade people using vocal and nonverbal cues. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. ACM, 293–300.

[8] K Anders Ericsson and Herbert A Simon. 1980. Verbal reports as data. *Psychological review* 87, 3 (1980), 215.

[9] Florian Eyben, Felix Weninger, Florian Gross, and Björn Schuller. 2013. Recent developments in opensmile, the Munich open-source multimedia feature extractor. In *Proceedings of the 21st ACM international conference on Multimedia*. ACM, 835–838.

[10] Juan Fasola and Maja Mataric. 2013. A socially assistive robot exercise coach for the elderly. *Journal of Human-Robot Interaction* 2, 2 (2013), 3–32.

[11] Jeremy D Finn and Kayla S Zimmer. 2012. Student engagement: What is it? Why does it matter? In *Handbook of research on student engagement*. Springer, 97–131.

[12] John H Flavell. 1979. Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist* 34, 10 (1979), 906.

[13] Mark C Fox and Neil Charness. 2010. How to gain eleven IQ points in ten minutes: Thinking aloud improves Raven's Matrices performance in older adults. *Aging, Neuropsychology, and cognition* 17, 2 (2010), 191–204.

[14] Mary L Gick. 1986. Problem-solving strategies. *Educational psychologist* 21, 1-2 (1986), 99–120.

[15] Goren Gordon and Cynthia Breazeal. 2015. Bayesian Active Learning-Based Robot Tutor for Children's Word-Reading Skills.. In *AAAI*. 1343–1349.

[16] 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.. In *AAAI*. 3951–3957.

[17] Lisa M Henjes. 2007. *The use of think-aloud strategies to solve word problems*. Master's thesis. University of Nebraska - Lincoln, Lincoln, NE.

[18] Deanna Hood, Séverin Lemaignan, and Pierre Dillenbourg. 2015. When children teach a robot to write: An autonomous teachable humanoid which uses simulated handwriting. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 83–90.

[19] Riitta Jääskeläinen. 2010. Think-aloud protocol. *Handbook of translation studies* 1 (2010), 371–374.

[20] Takayuki Kanda, Takayuki Hirano, Daniel Eaton, and Hiroshi Ishiguro. 2004. Interactive robots as social partners and peer tutors for children: A field trial. *Human-computer interaction* 19, 1 (2004), 61–84.

[21] James Kennedy, Paul Baxter, and Tony Belpaeme. 2015. Comparing robot embodiments in a guided discovery learning interaction with children. *International Journal of Social Robotics* 7, 2 (2015), 293–308.

[22] James Kennedy, Paul Baxter, Emmanuel Senft, and Tony Belpaeme. 2015. Higher nonverbal immediacy leads to greater learning gains in child-robot tutoring interactions. In *International conference on social robotics*. Springer, 327–336.

[23] James Kennedy, Paul Baxter, Emmanuel Senft, and Tony Belpaeme. 2016. Social robot tutoring for child second language learning. In *11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 231–238.

[24] Cory D Kidd and Cynthia Breazeal. 2004. Effect of a robot on user perceptions. In *Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004)*, Vol. 4. IEEE, 3559–3564.

[25] Iolanda Leite, André Pereira, Ginevra Castellano, Samuel Mascarenhas, Carlos Martinho, and Ana Paiva. 2011. Modelling empathy in social robotic companions. In *International Conference on User Modeling, Adaptation, and Personalization*. Springer, 135–147.

[26] 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, 423–430.

[27] Daniel Leyzberg, Samuel Spaulding, Mariya Toneva, and Brian Scassellati. 2012. The physical presence of a robot tutor increases cognitive learning gains. In *Proceedings of the Cognitive Science Society*, Vol. 34.

[28] Joseph E Michaelis and Bilge Mutlu. 2017. Someone to Read with: Design of and Experiences with an In-Home Learning Companion Robot for Reading. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 301–312.

[29] Marjorie Montague and Brooks Applegate. 1993. Middle school students' mathematical problem solving: An analysis of think-aloud protocols. *Learning Disability Quarterly* 16, 1 (1993), 19–32.

[30] André Pereira, Carlos Martinho, Iolanda Leite, and Ana Paiva. 2008. iCat, the chess player: the influence of embodiment in the enjoyment of a game. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 3*. International Foundation for Autonomous Agents and Multiagent Systems, 1253–1256.

[31] Paul R Pintrich and Elisabeth V De Groot. 1990. Motivational and self-regulated learning components of classroom academic performance. *Journal of educational psychology* 82, 1 (1990), 33.

[32] Aaron Powers, Sara Kiesler, Susan Fussell, and Cristen Torrey. 2007. Comparing a computer agent with a humanoid robot. In *2nd ACM/IEEE International Conference on Human-Robot Interaction*. IEEE, 145–152.

[33] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y Ng. 2009. ROS: an open-source Robot Operating System. In *ICRA workshop on open source software*, Vol. 3. Kobe, 5.

[34] Aditi Ramachandran, Chien-Ming Huang, and Brian Scassellati. 2017. Give Me a Break!: Personalized Timing Strategies to Promote Learning in Robot-Child Tutoring. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 146–155.

[35] Aditi Ramachandran, Alexandru Litoiu, and Brian Scassellati. 2016. Shaping productive help-seeking behavior during robot-child tutoring

interactions. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*. IEEE Press, 247–254.

[36] Yvonne Rogers, Helen Sharp, and Jenny Preece. 2011. *Interaction design: beyond human-computer interaction*. John Wiley & Sons.

[37] Carly Rosenzweig, Jennifer Krawec, and Marjorie Montague. 2011. Metacognitive strategy use of eighth-grade students with and without learning disabilities during mathematical problem solving: A think-aloud analysis. *Journal of learning disabilities* 44, 6 (2011), 508–520.

[38] Martin Saerbeck, Tom Schut, Christoph Bartneck, and Maddy D Janse. 2010. Expressive robots in education: varying the degree of social supportive behavior of a robotic tutor. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1613–1622.

[39] Elaine Short, Katelyn Swift-Spong, Jillian Greczek, Aditi Ramachandran, Alexandru Litoiu, Elena Corina Grigore, David Feil-Seifer, Samuel Shuster, Jin Joo Lee, Shaobo Huang, et al. 2014. How to train your dragonbot: Socially assistive robots for teaching children about nutrition

through play. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2014)*. IEEE, 924–929.

[40] Elizabeth J Short, Christopher Schatschneider, Cara L Cuddy, Steven W Evans, Dani M Dellick, and Laura A Basili. 1991. The effect of thinking aloud on the problem-solving performance of bright, average, learning disabled, and developmentally handicapped students. *Contemporary Educational Psychology* 16, 2 (1991), 139–153.

[41] Candace L Sidner, Christopher Lee, Cory D Kidd, Neal Lesh, and Charles Rich. 2005. Explorations in engagement for humans and robots. *Artificial Intelligence* 166, 1-2 (2005), 140–164.

[42] Maaike Van Den Haak, Menno De Jong, and Peter Jan Schellens. 2003. Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. *Behaviour & information technology* 22, 5 (2003), 339–351.

[43] MW van Someren, YF Barnard, and JAC Sandberg. 1994. The think aloud method: a practical approach to modelling cognitive processes. *Knowledge-based systems* (1994).