Shaping Productive Help-Seeking Behavior During Robot-Child Tutoring Interactions

Aditi Ramachandran, Alexandru Litoiu, Brian Scassellati
Department of Computer Science
Yale University
New Haven, CT 06511
{aditi.ramachandran, alex.litoiu, brian.scassellati}@yale.edu

***Abstract*—In intelligent tutoring systems, one fundamental problem that limits learning gains is the unproductive use of on-demand help features, namely overuse or aversion, resulting in students misusing the system rather than engaging in active learning. Social robots as tutoring agents have the potential to mitigate those behaviors by actively shaping productive help-seeking behaviors. We hypothesize that effectual help-seeking behavior is a critical contributor to learning gains in a robot-child tutoring interaction. We conduct a between-subjects study where children interacted with a social robot solving fractions problems over multiple sessions (29 children; 4 sessions per child) in one of two groups. Results showed that participants in our experimental group, who received adaptive shaping strategies from the robot targeting suboptimal help requests, reduced their suboptimal behaviors over time significantly more than a control group, as well as improved their scores from pretest to posttest significantly more than a control group.**

I. Introduction

There has been a large body of research demonstrating that students that receive one-on-one tutoring perform, on average, one to two standard deviations better than students learning via conventional classroom instruction when tested on the same material [1], [2]. Because many schools lack the resources to provide a one-on-one tutor to each student, we aim to uncover other methods of instruction that will emulate the benefits of one-on-one human tutoring. Preliminary research involving robotic agents as tutors indicates that the physical presence of a robot tutor can increase cognitive learning gains [3]. Further research shows that a robot tutor employing relatively simple personalization strategies can benefit the learner [4]. This motivates the need to more deeply investigate robotic tutoring systems as an effective method of instruction.

One salient aspect of a tutoring interaction that human tutors are well suited for is providing help to the student at the right time [5]. Responding to help-seeking behaviors, or how students request and utilize help in a tutoring environment, should be a key aspect of designing effective robot tutoring agents. One established way of providing help in a learning environment is to allow the student to utilize on-demand help. On-demand help refers to help provided by the learning environment that must be actively solicited by the learner [6]. While there has been research demonstrating that on-demand help is useful [7], students benefit more from this help when it is used productively [6]. Unproductive help-seeking includes behaviors such as "gaming the system" (rapid hint requests, for example) or help-aversion (lack of utilization of available help features) in a learning environment.

As HRI research supports increased enjoyment and compliance in participants who interacted with a physical robot as compared to similar on screen representations [8], [9], we leverage the use of a physically present social robot during a tutoring interaction to use social influence to shape productive help-seeking behavior. Because use of these unproductive help-seeking behaviors may affect learning outcomes, understanding if these behaviors can be successfully shaped by a robot tutor, and whether productive help-seeking behavior impacts learning in robot-child tutoring interactions is pivotal to the design of future robot tutoring interactions.

The role of help-seeking behaviors within robot-child tutoring interactions is a rather unexplored research direction, prompting many open questions. Can we design robots to shape more productive help-seeking behavior? Does the proper use of help features in a robot tutoring system impact learning? Are simple adaptive strategies employed by the robot tutor effective, or is it enough for children to rely on on-demand help in a tutoring scenario with a robot?

To address these questions: (1) We designed a controlled human-robot interaction study to understand the effects of a robot tutoring system aimed at shaping productive help-seeking behavior. (2) We conducted a four-session repeated interaction study where children interacted with a robot in a tutoring context, in which children either received the adaptive support strategies or used on-demand help from the robot. (3) We analyzed the tutoring sessions to measure change in help-seeking behavior across sessions and calculated pre/post test differences to assess learning gains. (4) We then presented results on these metrics and discussed implications for future robot tutoring interactions. We show that shaping help-seeking behavior is a crucial aspect of tutoring required for long-term learning in children and provide design guidelines for strategies on how to effectively increase learning gains.

II. Related Work

Previous work in tutoring presented to the HRI community has focused on demonstrating the efficacy of social robots in a tutoring setting. Kanda et. al shows the establishment of a stronger relationship between the child and a robot that exhibits social behavior, while also citing the positive effects of

978-1-4673-8370-7/16/$31.00 © 2016 IEEE

having a physically present entity [10]. Howley et. al describes the potential advantages of a robot tutor over a human tutor in some situations due to distinctions in social role and evaluation apprehension regarding seeking help within a learning interaction [11]. Finally, Mohseni-Kabir et. al reinforces the idea that a robot dialogue can promote learning gains within the context of a bidirectional coaching model between an adult and a robot [12]. They show that a robot providing suggestions at the right time can significantly increase the ability of a person to perform a task and retain knowledge, allowing the person to learn the task much more efficiently. These studies provide us with further justification for investigating robots as effective tutoring agents, specifically designed to interact with children and shape behavior over time. While much of the related work from the HRI community is relevant to the study and analysis conducted in this paper, no previous work in HRI aims to understand the effects of shaping productive help-seeking behaviors in a multi-session robot-child tutoring scenario.

There has been substantial research regarding help-seeking behavior in intelligent tutoring systems (ITS) that do not involve a robot. Studies show that on-demand help in which a tutor provides a hint whenever the child requests it, is effective [7], [13]. For example, the Geometry Cognitive Tutor, a program that offers no personalization and simply offers on-demand hints, was shown to be more effective than classroom instruction when used in conjunction with the curriculum of the interactive learning environment (ILE) [7]. Additionally, there are studies suggesting that help-seeking is a goal oriented behavior, and unsuccessful attempts at seeking help may be met with a reluctance to solve the problem, suggesting that children may prefer immediate gratification in the form of on-demand hints [13]. This provides the motivation for using a strictly on-demand model as a comparison for an adaptive model that adds the presence of shaping strategies to an on-demand model for help-seeking during a tutoring interaction.

However, Aleven et. al estimates that 72% of help requests are unproductive in on-demand tutoring, showing that help requests often come at the wrong time in these scenarios [14]. Further validating the adaptive help-seeking model, Roll et. al discusses a Help Tutor used in conjunction with the Geometry Cognitive Tutor that coached students on when to ask for hints, resulting in students making less help-seeking errors and seeking out less hints in general [7]. They identify abuses of the help-seeking system, overuse and aversion, and attempt to solve them using the Help Tutor, having students use the tutor for six sessions over three weeks. The Help Tutor focused on providing suggestions that advise against requesting a hint too often, ultimately reducing the number of hints the children ask for on average and improving help-seeking behavior. Additional research in the ITS community supports that the use of these unproductive help-seeking behaviors correlates with overall learning [14].

These studies provide a solid foundation with findings that support the possibility of a more effective tutoring model that involves robots. Because little work has been done involving help-seeking behavior within robot tutoring interactions, we do not yet know if the effects will be in accordance with what has been observed within ILEs that do not involve a physical robot. As our work involves a physical robot tutor, it is of crucial importance to understand the role of help-seeking behavior, whether we can shape this behavior using the social influence of a physical robot, and how help-seeking behavior impacts learning during robot-child tutoring interactions.

III. Suboptimal Help-Seeking Behaviors in Tutoring Interactions

As this study examines robot tutoring utilizing simple adaptive strategies for help-seeking as well as robot tutoring relying solely on on-demand help features, it is important to outline why these adaptive strategies are necessary. The existing ITS literature validates the presence of exploitative behavior ("gaming the system") and help-averse behavior (sometimes called help avoidance) in students who interact with ILEs [14], [15]. As these are the two most prevalent suboptimal help-seeking behaviors observed within tutoring environments, countering these would most likely lead to sizable learning gains due to help-seeking behavior shaping. Although psychological reasons likely play a role in understanding why these suboptimal help-seeking behaviors occur, we are most interested in understanding if they can be productively shaped. We describe each of these behaviors and provide examples from the ITS community.

A. Gaming the System

In the context of tutoring systems, gaming the system has been defined as "attempting to succeed in a learning environment by exploiting properties of the system rather than by learning the material and trying to use that knowledge to answer correctly" [16]. There are many examples of this behavior observed within intelligent tutoring systems. Two notable examples include systematic guessing and rapid hint requests [15]. Systematic guessing may involve guessing the same answer repeatedly to advance to the next question, while rapid hint requests typically involve the learner trying to acquire the answer without expending considerable effort thinking through the problem. In many of the intelligent tutoring systems in which gaming behavior has been detected, each problem will contain a series of hints, where the last hint is often called a "bottom-out hint" because it contains very specific information that is necessary to solve the problem [7]. Because the ITS community has demonstrated that gaming behavior is associated with poorer learning gains [17], we identify this is as one type of help-seeking behavior that should be productively shaped during a robot tutoring interaction.

B. Help-Aversion

The other noteworthy help-seeking behavior identified in tutoring environments that impacts learning is help-aversion. Help-averse behavior typically involves the lack of use of help features in a learning environment when it is likely to benefit the learner. In most ILEs, there are help features built into the

	Session 1	Session 2	Session 3	Session 4
Control:	Hint provided when requested	Hint provided when requested	Hint provided when requested	Hint provided when requested
Adaptive:	Hint provided when requested	Support strategies triggered	Support strategies triggered	Support strategies triggered

Fig. 1. Experimental design for four-session robot-child tutoring interaction study.

system, for example, in the form of a button on the screen, where the user can request help when needed. Help-aversion is typically observed when the student makes many incorrect attempts but ignores the help button altogether, ultimately failing to utilize the help available to them from the tutoring system. As ineffective use of the help features in a tutoring system may also impact learning gains in a robot tutoring interaction, we identify help-aversion as the other help-seeking behavior that we aim to productively shape through adaptive strategies with a robot tutor.

IV. Methodology

This section gives a detailed account of an experiment in which a robot tutor assisted children with math problems over four separate sessions. The study employed a between-subjects design in which participants were randomized into one of two conditions: control (participants utilize on-demand help) and adaptive (participants received adaptive support strategies from the robot). Each student participated in four one-on-one tutoring sessions with the robot, over the course of approximately two weeks. We chose to use a repeated measures design to examine behavior change and learning gains of the participants over time.

A. Participants

The participants in this study were fifth and sixth grade students from local public schools. A total of 33 students were recruited in the schools where the study was conducted; however, four participants were excluded (three for not completing the study due to school absences, and one for non-compliance). For this analysis, we considered a total of 29 students, with 15 participants in the control condition and 14 participants in the adaptive condition. In the control group, there were eight males and seven females with a mean age of 10.9 years (SD = .80). The ethnicity of each participant was reported by parents: 13.3% Asian, 60.0% Caucasian, 13.3% Hispanic, 6.7% reported more than one ethnicity, and 6.7% did not report. This group's average pretest score was .51 (SD = .27). In the adaptive group, there were eight males and six females with a mean age of 10.68 years (SD = .54). The ethnicities of the participants as reported by parents were: 7.1% Asian, 78.6% Caucasian, 7.1% reported more than one ethnicity, and 7.1% did not report. The average pretest score for the adaptive group was .31 (SD = .29). There were no major differences in the distribution of the participants across groups.

B. Tutoring Scenario

Participants were escorted from their classrooms by the experimenter. Before tutoring session one with the robot, each child was introduced to the robot. The robot greeted the participant, saying "Hello! My name is Nao, your personal robot tutor. I'm really excited to meet you and work on some problems together." The robot was in a seated position and also waved during the introduction. After this introduction, the participants completed a pretest. They then completed four distinct tutoring sessions with the robot, spaced over approximately two weeks. They also completed a posttest after session four of the tutoring interaction. The tests and interview were not completed in the presence of the robot.

Children in the study were asked to complete math problems using a tablet device. For both adaptive and control groups, a robot acted as a tutoring agent throughout each interaction by introducing each math problem, providing hints on how to solve the problems, and informing the student when an answer was correct or incorrect. In both conditions, the participants could request help from the robot through the use of buttons on the tablet interface.

In the adaptive condition, in sessions two through four, the robot employed two simple strategies to shape help-seeking behavior when suboptimal help-seeking took place. In the control condition, no such strategies were triggered. We compared behavior and performance by group to understand the effectiveness of the robot's use of adaptive strategies for productive help-seeking behavior. A visual representation of the study design can be seen in Figure 1. Because we wanted to focus on each child's behavior and not solely on learning outcomes, we needed to assess each child's baseline behavior during the first interaction, without introducting the confound of the robot's shaping strategies. While this may set up some expectation from the child, we did not believe that starting the simple strategies in session two for the adaptive group was enough to violate their expectations for the robot behavior in a significant way. Since we planned to compare the groups directly in our analysis, it was important to assess this behavior in precisely the same way for both groups of participants during session one. This is the motivation for why the adaptive group received the shaping strategies from the robot starting in session two.

C. Robot

We used a NAO robot to act as a tutoring agent in each of the four sessions. Over the course of the tutoring inter-

Question #	Problem Type
1	equivalent fractions
2	equivalent fractions
3	finding a common denominator
4	converting mixed numbers
5	adding fractions
6	subtracting fractions
7	adding fractions
8	subtracting fractions

TABLE I
PROBLEM TYPES WITHIN EACH TUTORING SESSION

actions, the robot also acted as a social agent by speaking to the participant. Each time a new session started, the robot greeted the participant, utilizing phrases such as "nice to see you again" in sessions after the first. The robot also said "Congratulations! You have completed this session" at the end of each of the four sessions. The robot gave the impression that it was observing the interaction, as it was in a seated position to the left of the participant and alternated between looking at the participant during speech and at the tablet while the participant was working on a problem (see Figure 2). At the start of each problem, the robot introduced the type of question. For example, it would say "now here is a question about adding fractions" while performing simple gestures to accompany its speech.

Each time a child entered an answer on the tablet, the robot provided feedback on whether it was correct or incorrect, and used phrases such as "great job!" for a correct answer and "give it another shot!" for an incorrect answer. At the beginning of each interaction, the robot informed the child that they could press buttons on the bottom of the tablet to ask for help if needed. The robot reacted exclusively to tablet input from the child throughout the course of the tutoring sessions. Additionally, the robot provided hints when the child requested help by using these buttons on the tablet device. These robot behaviors occurred consistently across all participants in both groups over all sessions. The following robot behavior occured for only the adaptive group: the robot automatically provided a hint or denied a hint when specific conditions were met by the participant. The robot operated entirely autonomously in real-time with each child, with no input from the experimenter present throughout the duration of an interaction.

D. Tablet Application

The students in the study completed math problems on a tablet device positioned in front of them. Each of the four sessions contained eight math problems, specifically dealing with fractions concepts. All problems followed state curriculum standards, and were designed for students in fifth or sixth grade. The concepts covered were consistent across the four sessions. Table I shows which concepts each of the eight questions per session covered. In each session, the same problems with different numbers were used.

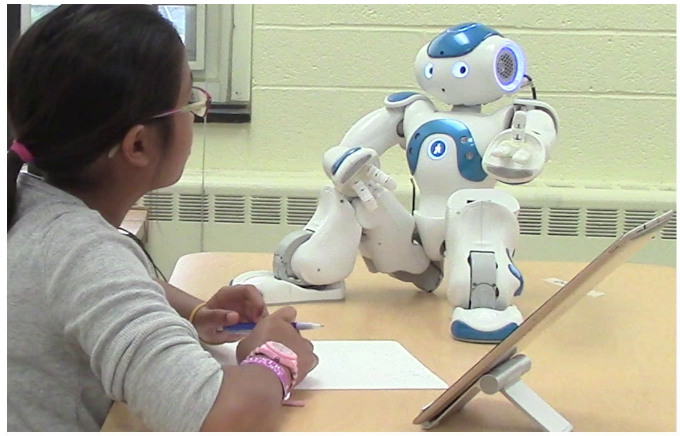


Fig. 2. Child interacting with the NAO robot in the fractions tutoring scenario.

The tablet interface was simple and served as an input device for the tutoring session. The screen displayed the question after it was introduced by the robot. Students could enter their answer on the screen using the number pad available to them. After pressing the submit button, the robot would inform them if their answer was correct or incorrect. The tablet also displayed feedback showing the words 'correct' or 'incorrect' as necessary for each attempt. If the answer was correct, the student could initiate the application to move on to the next question. If the answer was incorrect, the current question would remain on screen. After an incorrect attempt, participants could also see their most recent attempt displayed on the screen, as well as how many remaining attempts they had for that question. Participants were given five attempts per problem. If a participant made five incorrect attempts on a problem, the correct answer would be displayed after the fifth attempt, and the student would then initiate the application to display the next question.

Buttons towards the bottom of the tablet application served as a the method for a participant to ask the robot for help. Each problem had exactly three hints associated with it and the hints had to be requested in order. Because the robot was verbally saying the hint, the hint could be repeated at any time while working on that problem once it was originally requested. Each successive hint provided more information, meaning that the third hint contained the most information relevant to the given problem. A screenshot of the tablet application can be seen in Figure 3.

E. Adaptive Strategies

Participants interacted with the robot in four sessions as part of either the control group or the adaptive group. As mentioned in the previous section, there were three hints per problem, which had to be requested in sequential order. Participants in the control group relied on the buttons on the tablet interface to make help requests, thereby utilizing on-demand help features of the application.

Participants in the adaptive group also followed this same method of requesting help; however, the robot employed two simple strategies aimed at countering suboptimal help-seeking behavior. The two strategies in use were designed to combat help-averse behavior and exploitative behavior, two established help-seeking behaviors in this particular tutoring context. The robot executed the following strategies for participants in the adaptive group:

- $\mathbf{S_1}$: If the participant makes two consecutive incorrect attempts on a problem without asking for any hints, the robot will automatically provide the participant with the next hint they have not yet requested.
- $\mathbf{S_2}$: If the participant makes three consecutive hint requests on a problem without an attempt in between, the robot will deny the participant the third hint, and request that the participant attempts the problem before asking for more help.

We chose our adaptive strategies in the context of the tutoring system we designed, and we wanted to be conservative in what we defined as "too much" or "too little" help. Because we did not know how all children would use the system in advance, we did not want the robot triggering the strategies too frequently throughout the sessions. To accomplish this, we erred on the side of automatically providing a hint only when children made multiple incorrect attempts without requesting help on their own, which indicated having difficulty with the problem without using the provided help features. Additionally, we chose to deny a hint request only when children requested all three hints in a row, which clearly indicated requesting too much help when there were only three hints for each problem.

While triggering $\mathbf{S_1}$ may not indicate that the participant was completely help-averse, the behavior involved was still considered suboptimal in this context. Therefore, if a participant triggered $\mathbf{S_1}$, they were not using the help features of the tutoring system in the most productive way, due to lack of use of the help available. Similarly, $\mathbf{S_2}$ is a strategy used to counter another expected pattern of suboptimal help usage, namely making successive hint requests to receive the most information before attempting the problem. This behavior indicated that the participant was not trying to utilize the information presented in previous hints. Asking the student to make an attempt before requesting more help can encourage the participant to make a bigger effort to understand the previous hints and utilize that information.

While individuals may require different levels of help based on the challenges the questions pose to them, our goal in this study was not to adapt to this difference, but rather to shape extreme help-seeking behaviors because of their prevalence in ITS research. Our aim in designing a robot tutor that employs these rules was to understand whether simple strategies can be effective in shaping more productive help-seeking behavior over several sessions, as well as understanding how it affects learning outcomes as compared to a robot tutoring system that only offers on-demand help.

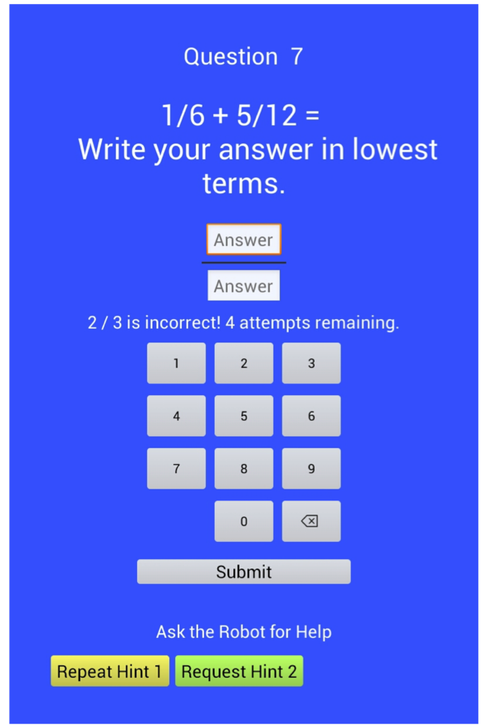


Fig. 3. A screenshot of the tablet application used by participants in the study. This screen shows an example question answered by participants. In this example, the buttons at the bottom of the screen show that the first hint has already been requested, and can now be repeated as needed. The second hint is available and can be requested by pressing the green button.

V. Data Analysis

In this section, we describe the metrics we use to evaluate the effectiveness of the robot employing simple adaptive support strategies over several sessions.

A. Help-seeking Behavior Change

Though strategies $\mathbf{S_1}$ and $\mathbf{S_2}$ were only employed for participants in the adaptive group during sessions two through four, we counted the number of times both strategies would have been triggered for each participant for a given session. This is simply the sum of the number of times a hint would be automatically given (num_auto_hints) and the number of hints that would be denied (num_denied_hints) during a given tutoring session. For participant i during session S, the following formula was used:

$$num\_triggers(i,S) = num\_auto\_hints(i,S) + num\_denied\_hints(i,S)$$

We calculate $num\_triggers(i, S)$ for each participant where $S = 1$ and $S = 4$. This metric represents the number of suboptimal help-seeking behaviors observed for a participant in session one and in session four. Collecting this metric from session one allows us to assess one aspect of baseline help-seeking behavior. We can then compare this number for each participant to the number of triggers observed in session four. As we expect children to vary in the number of triggers they exhibit in session one, we are interested in whether the number of triggers decreases for each participant across sessions, and whether this decrease is significant for the control group, the adaptive group, or both groups.

In order to directly compare the control and adaptive groups based on our between subjects design, we define $\Delta_{triggers}$ as a metric that captures the difference between number of triggers from session one to session four for participant i:

$$\Delta_{triggers}(i) = num\_triggers(i,1) - num\_triggers(i,4)$$

We are interested in understanding whether this metric differs significantly between groups.

B. Learning Gains

Participants were asked to complete a pretest before session one and a posttest after session four in order to measure learning gains. Both the pretest and posttest contained eight questions, containing similar content to what was presented during the four tutoring sessions. The questions on both pretest and posttest were almost identical, with differing numbers within each problem. We scored both tests by awarding one point for each correct answer and dividing the number of correct answers by the total number of questions, resulting in scores ranging from zero to one. The difference in test scores between the pretest and the posttest is a within-subjects measure of each participant's learning improvement over the course of the entire experiment. While it seems likely that working through fractions problems would result in an improvement from pretest to posttest involving the same type of fractions problems, we are interested in understanding if learning gains increase statistically significantly for the control group, the adaptive group, or both groups. This will allow us to understand whether participants who receive adaptive help-seeking strategies from the robot improve their test scores before and after the repeated interactions, as well as whether participants who are limited to on-demand help from the robot are able to improve their scores over the sessions.

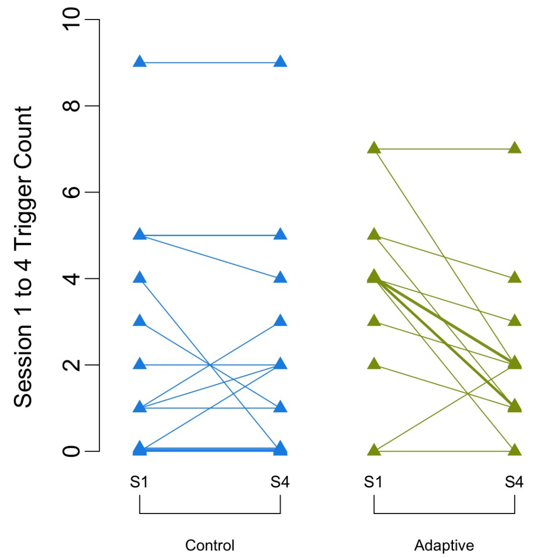
In accordance with our between-subjects design, we again define a metric, Δ_{score}, which captures the change between pretest and posttest score for each participant i. Rather than using an absolute difference between pretest and posttest score, we use normalized learning gain, which is an established metric that allows us to control for individuals starting at different levels of expertise.

$$\Delta_{score}(i) = \frac{score_{\text{post}}(i) - score_{\text{pre}}(i)}{1 - score_{\text{pre}}(i)}$$

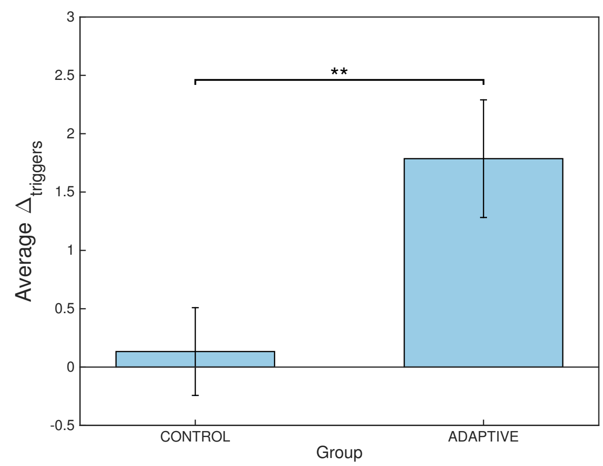
We are interested in understanding whether learning gains as defined by this metric differs significantly between groups.

VI. Results

As we are interested in understanding how the number of suboptimal help-seeking behaviors and test scores change for each individual participant over time, we first analyze these using within-subjects measures. We then utilize the change metrics, $\Delta_{triggers}$ and Δ_{score}, to directly compare between experimental conditions. For within-subjects measures, we separate our data by group (control and adaptive) so that we can understand if there is significant change in suboptimal help-seeking behavior and learning gains over time for either of the groups. Between-subjects measures allow us to understand whether the change over time for a given measure is significantly different between the adaptive and control groups. We used an alpha level of .05 for all statistical tests.



(a) Number of suboptimal triggers from session one (S1) to session four (S4) for individual participants, separated by group. Thicker lines represent multiple participants with the same values in sessions one and four.



(b) Average change in trigger count ($\Delta_{triggers}$) for participants in each group. (**) denotes $p < .01$, where the p-value is based on the Mann-Whitney test statistic. Means and error bars are depicted to visualize results for this data. Error bars represent standard error.

Fig. 4. Results for behavior change indicate that participants in the adaptive group were able to significantly decrease their suboptimal help-seeking behaviors from session one to four, as well as decrease these behaviors significantly more than the control group.

A. Help-seeking Behavior Change

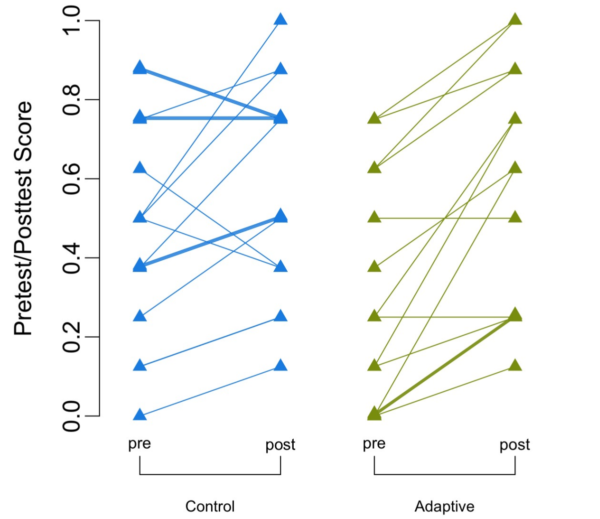
In order to assess behavior change, we calculate the number of suboptimal help-seeking behaviors as defined above both in session one and in session four for each participant. For participants in the adaptive group, the number of suboptimal triggers in session four (Mdn = 2.0, IQR = 1) was statistically significantly lower than the number of suboptimal triggers in session one (Mdn = 4.0, IQR = 2), Z = -2.605, $p = .009$ (Wilcoxon Signed-ranks test). The same test was run for participants in the control group, and it showed that the tutoring sessions limiting participants to using on-demand help did not elicit a statistically significant change in number of suboptimal help-seeking behaviors observed from session one (Mdn = 1.0, IQR = 4) to session four (Mdn = 1.0, IQR = 3), Z = -.213, $p < .832$ (Wilcoxon Signed-ranks test). Figure 4a shows the change in number of triggers from session one and session four for each participant separated by group. These results demonstrate that the adaptive robot strategies aimed at shaping productive help-seeking behavior were successful in mitigating the occurrences of suboptimal help-seeking behaviors over time. Participants in the adaptive condition significantly decreased their number of suboptimal help-seeking behaviors over time while participants in the control condition did not.

The decrease in number of triggers was significantly greater for the adaptive group (Mdn = 1.5, IQR = 3) than for the control group (Mdn = 0.0, IQR = 0), indicated by a Mann-Whitney test, U = 45.000, $p = .008$. These results show that participants receiving the adaptive strategies from the robot were able to decrease the number of suboptimal help-seeking behaviors they exhibit over time statistically significantly more than those in the control group (see Figure 4b). This result further validates the effectiveness of the adaptive strategies employed by the robot.

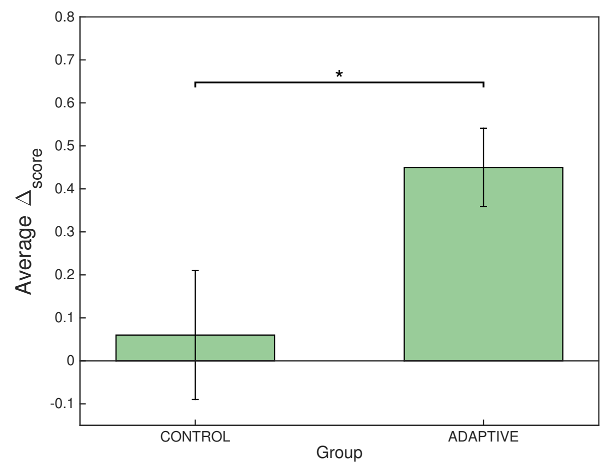
B. Learning Gains

Because we assess learning gains by examining differences in score from pretest to posttest, we again have two related samples for each participant. For participants that received the adaptive robot strategies, posttest scores (Mdn = .62, IQR = .63) were statistically significantly higher than pretest scores (Mdn = .25, IQR = .63), Z = 3.089, $p = .002$ (Wilcoxon Signed-ranks test). The same test was run for the control group, and the output indicated that there was no statistically significant change in score from pretest (Mdn = .50, IQR = .38) to posttest (Mdn = .75, IQR = .38), Z = 1.615, $p = .106$ (Wilcoxon Signed-ranks test). Figure 5a shows learning gains over time for each participant, separated by group. These results indicate that the group that received adaptive strategies from the robot tutor over four sessions were able to significantly improve their test scores, while the group that relied on using on-demand help from the robot over the four sessions were not. This suggests that a robot employing strategies aimed at shaping productive help-seeking behavior can impact learning gains over time.

Because the data measuring Δ_{score} for each group are normally distributed, we use an independent samples t-test to understand differences in normalized learning gains between groups. Our results indicate that participants in the adaptive group (M = .45, SD = .34) improved their score from pretest to posttest statistically significantly more than those in the control group (M = .06, SD = .59), $t(27) = -2.169$, $p = .039$ (see Figure 5b). This result demonstrates that participants receiving adaptive strategies from the robot were able to increase their learning gains more than those in the control group, indicating that the shaping strategies employed by the robot improved help-seeking behaviors, which thereby impacted learning outcomes between groups.



(a) Pretest/posttest scores for individual participants, separated by group. Pretest and posttest scores were measured in number of questions answered correctly divided by eight total questions. Thicker lines represent multiple participants with the same scores.



(b) Average normalized learning gains (Δ_{score}) for participants in each group. (*) denotes $p < .05$. Error bars represent standard error.

Fig. 5. Results for learning gains demonstrate that participants in the adaptive group were able to significantly increase their scores from pretest to posttest, as well as improve their score significantly more than those in the control group.

VII. Discussion

This study assesses whether robot tutors employing adaptive strategies aimed at shaping productive help-seeking behavior provides benefits to learners when compared to robot tutoring systems that provide on-demand help. Our results demonstrate the effectiveness of the simple adaptive strategies employed by the robot for participants in the adaptive condition, as those in the adaptive condition were able to reduce the number of times they engaged in suboptimal help-seeking significantly more than those in the control condition.

Furthermore, the shaping strategies were not only successful in causing participants to trigger them less frequently, but also may have caused participants to expend more effort on the problem at hand. When participants were forced to make an attempt before receiving an additional hint, they were expected to utilize the hints they had already received to think through the problem. When participants were automatically provided with a hint after making multiple wrong attempts without explicitly requesting help, they were expected to utilize the information in the hint to better understand how to do the problem rather than resort to guessing. We believe that these strategies were useful in causing participants to engage more effectively with the learning environment, as participants who received this adaptive robot behavior improved their test scores over time statistically significantly more than those who did not receive the adaptive strategies from the robot.

This leads us to believe that even relatively simple behavior-shaping strategies can impact learning outcomes during robot tutoring interactions, and must be explored further as a promising method of one-on-one tutoring. These results combined with the fact that the adaptive strategies were aimed at shaping productive help-seeking behavior further confirms the importance of thoroughly examining help-seeking behaviors in children within learning environments. We conclude that the help features of a robot-child tutoring system must play a crucial role in the design of future tutoring systems. Additional work is needed to understand whether the improvement in suboptimal help-seeking will persist over a longer course of time, as well as whether it will transfer to other learning environments.

VIII. Conclusion

In this paper, we investigated the role of a robot tutor that employed adaptive strategies designed to mitigate suboptimal help-seeking behaviors in children. We compared both number of suboptimal behaviors and learning gains over time across two conditions, one in which the robot utilized adaptive shaping strategies, and one in which the robot provided on-demand help. We found that participants who received the adaptive strategies from the robot improved their help-seeking behavior significantly more than those who solely relied on on-demand help from the robot. Additionally, the participants in the adaptive group improved their test scores significantly more than those in the control group, indicating that productive help-seeking behavior during robot tutoring interactions does impact learning outcomes. We conclude that even simple strategies aimed at shaping productive help-seeking behaviors can provide significant benefits for children engaging in learning interactions with a social robot.

Acknowledgments

This work was supported by the National Science Foundation award #1139078 (Socially Assistive Robots). We thank Priyanka Krishnamurthi, Eric Ho, and Nicole Salomons for helping with the experiment and sections of the paper. We also thank the students and staff at the schools where data was collected.

References

[1] B. S. Bloom, "The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring." *Educational Researcher*, vol. 13, no. 6, pp. 4–16, 1984. [Online]. Available: http://www.eric.ed.gov/ERICWebPortal/

[2] K. VanLehn, "The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems," *Educational Psychologist*, vol. 46, no. 4, pp. 197–221, 2011.

[3] D. Leyzberg, S. Spaulding, M. Toneva, and B. Scassellati, "The physical presence of a robot tutor increases cognitive learning gains," in *34th Annual Conference of the Cognitive Science Society (Cog. Sci.'12)*, 2010.

[4] D. Leyzberg, S. Spaulding, and B. Scassellati, "Personalizing robot tutors to individuals' learning differences," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2014, pp. 423–430.

[5] D. C. Merrill, B. J. Reiser, M. Ranney, and J. G. Trafton, "Effective tutoring techniques: A comparison of human tutors and intelligent tutoring systems," *The Journal of the Learning Sciences*, vol. 2, no. 3, pp. 277–305, 1992.

[6] V. Aleven, E. Stahl, S. Schworm, F. Fischer, and R. Wallace, "Help seeking and help design in interactive learning environments," *Review of Educational Research*, vol. 73, no. 3, pp. 277–320, 2003.

[7] I. Roll, V. Aleven, B. M. McLaren, and K. R. Koedinger, "Improving students help-seeking skills using metacognitive feedback in an intelligent tutoring system," *Learning and Instruction*, vol. 21, no. 2, pp. 267–280, 2011.

[8] A. Pereira, C. Martinho, I. Leite, and A. Paiva, "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, 2008, pp. 1253–1256.

[9] W. A. Bainbridge, J. Hart, E. S. Kim, and B. Scassellati, "The effect of presence on human-robot interaction," in *The 17th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2008)*. IEEE, 2008, pp. 701–706.

[10] T. Kanda, M. Shimada, and S. Koizumi, "Children learning with a social robot," in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2012, pp. 351–358.

[11] I. Howley, T. Kanda, K. Hayashi, and C. Rosé, "Effects of social presence and social role on help-seeking and learning," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2014, pp. 415–422.

[12] A. Mohseni-Kabir, C. Rich, S. Chernova, C. L. Sidner, and D. Miller, "Interactive hierarchical task learning from a single demonstration," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2015, pp. 205–212.

[13] S. N.-L. Gall, "Help-seeking behavior in learning," *Review of Research in Education*, pp. 55–90, 1985.

[14] V. Aleven, B. McLaren, I. Roll, and K. Koedinger, "Toward tutoring help seeking," in *Intelligent Tutoring Systems*. Springer, 2004, pp. 227–239.

[15] R. S. Baker, A. T. Corbett, I. Roll, and K. R. Koedinger, "Developing a generalizable detector of when students game the system," *User Modeling and User-Adapted Interaction*, vol. 18, no. 3, pp. 287–314, 2008.

[16] R. S. Baker, A. T. Corbett, K. R. Koedinger, S. Evenson, I. Roll, A. Z. Wagner, M. Naim, J. Raspat, D. J. Baker, and J. E. Beck, "Adapting to when students game an intelligent tutoring system," in *Intelligent Tutoring Systems*. Springer, 2006, pp. 392–401.

[17] R. S. Baker, A. T. Corbett, K. R. Koedinger, and A. Z. Wagner, "Off-task behavior in the cognitive tutor classroom: when students game the system," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2004, pp. 383–390.