Give Me a Break! Personalized Timing Strategies to Promote Learning in Robot-Child Tutoring

Aditi Ramachandran, Chien-Ming Huang, Brian Scassellati
Department of Computer Science
Yale University
{aditi.ramachandran, chien-ming.huang, brian.scassellati}@yale.edu

ABSTRACT

A common practice in education to accommodate the short attention spans of children during learning is to provide them with non-task breaks for cognitive rest. Holding great promise to promote learning, robots can provide these breaks at times personalized to individual children. In this work, we investigate personalized timing strategies for providing breaks to young learners during a robot tutoring interaction. We build an autonomous robot tutoring system that monitors student performance and provides break activities based on a personalized schedule according to performance. We conduct a field study to explore the effects of different strategies for providing breaks during tutoring. By comparing a fixed timing strategy with a *reward* strategy (break timing personalized to performance gains) and a *refocus* strategy (break timing personalized to performance drops), we show that the personalized strategies promote learning gains for children more effectively than the fixed strategy. Our results also reveal immediate benefits in enhancing efficiency and accuracy in completing educational problems after personalized breaks, showing the restorative effects of the breaks when administered at the right time.

Keywords

Child-robot Interaction; Personalization; Education

1. INTRODUCTION

Among the large variety of learning technologies available, social robots have shown great promise as tutoring agents [9, 10, 29]. The advantages of physical embodiment especially help establish effective human-robot interactions [25, 33]. Prior human-robot interaction (HRI) research has also demonstrated that embodied tutoring agents can increase cognitive learning gains [15]. Moreover, robots can increase enjoyment and can provide social and emotional support [12, 22, 27] to facilitate effective learning. In addition, robots can personalize their tutoring strategies to meet various needs, such as different learning capacities and readiness levels, and
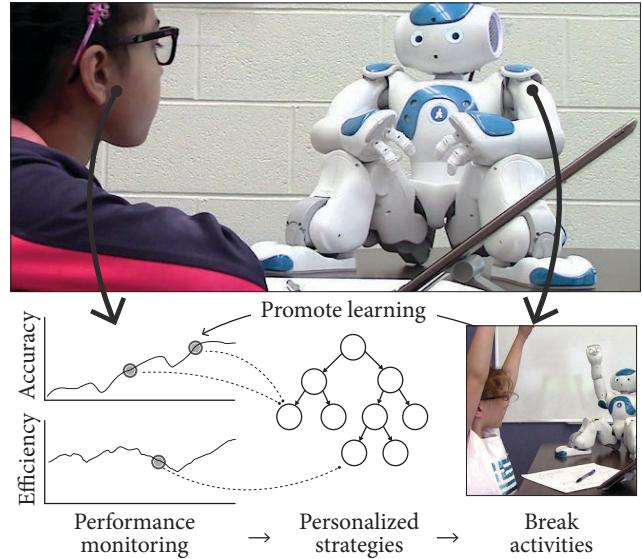
Figure 1: We studied personalized strategies for providing breaks to promote children's learning during one-on-one tutoring interactions with a robot.

can provide individualized experiences to maximize learning outcomes for students. As informed by the intelligent tutoring systems (ITS) community (e.g., [2]), prior HRI work has explored the impact of personalization on student learning and has demonstrated that personalized learning interactions can greatly benefit student learning outcomes [14, 26]. However, prior work has mainly studied content personalization, focusing on providing educational content that suits the student's ability and progress.

Aside from educational content, one crucial aspect in a learning interaction is the student's engagement and attention, particularly for younger learners who often have short attention spans. The attention spans of children can be as short as five minutes in a learning interaction [23], and the capacity for sustained attention only develops significantly between age 11 and adulthood [17]. Accommodating this need in learning interactions for young students is extremely important. Research has suggested that breaks are beneficial and provide needed cognitive rest during learning [1, 16]. Moreover, interspersing non-task breaks into an extended learning task may reduce interference and strengthen the learner's performance [3, 32]. Therefore, to effectively promote learning for younger students, robots need to intelli-

gently provide necessary breaks over the course of a learning interaction (Figure 1).

Break timing—when to provide a break—is particularly important, as it allows students to have needed cognitive rest at the right time. In a traditional learning environment, such as a classroom, breaks are usually taken at regular intervals. On the other hand, in a tutoring interaction, breaks can take place at times personalized to an individual. Breaks can be positive reinforcers for desired behavior (e.g., improved learning performance). Alternatively, breaks can be an opportunity for students to refocus after experiencing a decline in learning performance. In this work, we explore how a tutoring robot can provide breaks following a personalized schedule based on learning performance and seek to understand how such personalization might influence student learning outcomes.

2. BACKGROUND

In this section, we present literature showing that breaks can contribute to effective learning, and offer evidence about different ways to provide these breaks. We then review prior HRI work involving personalization in learning.

2.1 Non-task Breaks and Learning Gains

Breaks—pausing the current task to rest or to work on a different task—are beneficial for cognitive- and attention-loaded tasks. For example, during cognitive work involving recognition memory, breaks can foster achievement gains [32]. Furthermore, breaks of varying lengths can have restorative effects on reaction time during an auditory response task [16]. Accordingly, breaks play an important role in learning, a complex activity demanding the learner's cognitive attention. Learning can be particularly challenging for children, as their attention spans can be quite short, making them susceptible to distractions during learning [3, 23]. In support of this need, a common educational practice is to provide breaks throughout a school day at scheduled times. Primary schools in various countries that allow children to take 10- to 15-minute breaks every 40 to 45 minutes of classroom instruction report increased attentiveness after the breaks, indicating the importance of breaks during learning [20, 21]. However, limited time and resources during a school day necessitate that all children receive these breaks at the same time. Each individual has different learning needs and varied attention spans, suggesting that these non-task breaks may be most useful if provided at the "right" time.

Alternative educational practices suggest that non-task breaks could be provided in ways more personalized to each individual. For example, *success-based rewards* can enhance performance [24]. This provides evidence for one design of a personalized timing strategy for breaks in which breaks are used as success-based rewards for students demonstrating learning improvement. This strategy uses these breaks as a positive reinforcer for desired performance improvements. Another classroom practice, *positive time-out*, informs a second design of a personalized timing strategy to prevent or mitigate negative emotions. "Positive time-out" allows a child to take a brief break to avoid outbursts caused by affective reactions such as frustration [18]. Students can experience a wide variety of negative emotions during learning, often due to poor performance, and these can have a further negative impact on learning gains [19]. Thus, this strategy provides breaks during moments of potential negative emotion to enhance learning gains.

2.2 Personalized Robot Tutoring

Social robots have been shown to be a promising educational technology, largely because of their capabilities to provide personalized instruction for individual students [5]. Prior work done in this domain has focused on content-based personalization where the robot's personalized behavior directly involves the content of the learning task at hand. For example, Leyzberg et al. showed that providing personalized lessons based on an assessment of a person's skills on a cognitive task could significantly improve performance as compared to those who received non-personalized lessons or no lessons at all [14]. Westlund and Breazeal demonstrated that children showed vocabulary improvements in a story-telling task over time when the robot personalized the difficulty level of the words to match the child's ability [34]. Another relevant study by Schadenberg et al. explored the personalization of the difficulty level of a game that children played with a social robot, demonstrating that the level could be effectively adapted to each individual according to an ongoing assessment of the child's performance [28]. Different from prior studies, in this work, we explore how non-content personalization can impact learning outcomes.

Learning is a complex process involving a variety of supportive mechanisms that are often not directly related to the cognitive task at hand, such as help-seeking behavior, affective support, and rapport-building. Prior work in HRI has explored how robot tutors might personalize their behavior based on some of these supporting mechanisms. For example, Ramachandran et al. showed that a robot employing adaptive strategies that regulate a child's use of help while solving math problems can improve learning gains [26]. Gordon et al. studied a robot employing affective personalization and used reinforcement learning to discern which affective states each child preferred for the robot during a learning interaction [4]. Similarly, Henkemans et al. demonstrated the potential of a health education robot that personalized its behavior towards a child by referring to the child's name, favorite activity, and color during conversation [6]. Our work explores the personalization of the timing of when to provide breaks during tutoring to enhance learning.

The HRI community has further explored computational methods to estimate the engagement or attention of a user and has studied how robots can utilize the estimation to create effective learning. Leite et al. built data-driven classifiers to detect disengagement in groups of children versus individual children [11]. Other recent work by Lemaignan et al. defined an online method of assessing a child's attention during a learning interaction [13]. Szafir and Mutlu showed that a robot can monitor attention in real-time based on EEG sensor data and use this information adaptively to improve student recall ability [30]. Rather than elaborated estimation of attentional state, our work uses performance-based features during learning to guide tutoring interactions.

3. ROBOT TUTORING SYSTEM

In this section, we provide an overview of our design of an autonomous robot tutoring system (Figure 2). We also present the personalized strategies and support mechanisms implemented for our investigation of how the robot tutor may personalize break timing to promote learning.
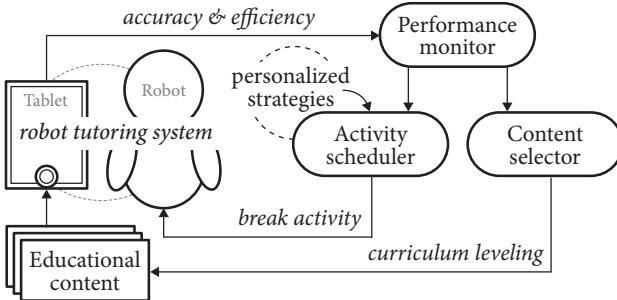
Figure 2: System architecture of our robot tutoring system.

3.1 System Overview

Our robot tutoring system consists of three main software components—*performance monitor*, *activity scheduler*, and *content selector.* The performance monitor is responsible for continuously tracking the student's learning performance, particularly his or her accuracy and efficiency at solving the educational problems presented on a tablet. The activity scheduler utilizes the collected performance information as well as personalized strategies to decide when to provide a non-task break activity during the student's learning interaction. The content selector uses the student's accuracy performance to selectively pick subsequent educational content that matches the student's mastery of the learning material. Specifically, the content selector decides the suitable difficulty level of the problems presented to the student. We use the Nao robot in our system to provide tutoring support.

3.2 Personalized Strategies

In this work, we explore three strategies for choosing when to provide a break during a tutoring interaction. The implementation of these strategies will be discussed in Section 4.2. Below, we describe the design rationale of these strategies.

Fixed strategy—This strategy provides breaks to students on a fixed schedule at regular intervals, reflecting the classroom practice that all children receive breaks at specified times rather than times particular to the individual.

Reward strategy—This strategy provides breaks as a reward after good performance as informed by the educational practice of "success-based rewards" [24]. This strategy seeks to postively reinforce desired learning improvement.

Refocus strategy—This strategy seeks to interrupt negative behaviors, such as distraction, during learning by providing a break when a drop in performance is detected. This design is informed by the educational practice of "positive time-out" [18], providing a student with the opportunity to refocus by taking a break from the task at hand.

3.3 Support Mechanisms for Tutoring

In addition to the personalized strategies, our system also implements several basic support mechanisms, including providing necessary information on the tablet application and exhibiting engaging robot behaviors, to facilitate tutoring interactions with young learners. All behaviors described in this section apply to all students regardless of the assigned conditions for the user study described in Section 4.2. At the start of a tutoring session, the robot greets each student and conducts a small, interactive lesson on an educational topic. After the lesson, the robot presents a series of questions on the taught topic for the student to practice. We carefully

designed robot behaviors to give the students the impression that the robot was responsible for facilitating the learning interaction. For instance, the robot looks at the students when speaking and looks at the tablet while they work on practice questions. The robot also uses gestures throughout a session, often extending an arm towards the tablet to invite students to direct their attention towards it.

The content selector chooses each practice question from a bank of problems with multiple difficulty levels to accommodate a variety of learners. All students start with the lowest level of questions available and can only advance to subsequent levels after demonstrating a specified level of mastery. The robot provides feedback on whether an answer is correct or incorrect after each individual problem. If the answer is incorrect, the robot also provides general feedback about how to solve the problem, while the tablet displays the associated worked-out solution. Moreover, the tutoring application helps manage the flow of the tutoring interaction by providing buttons on the screen that allow the student to initiate the presentation of the next problem and disabling buttons displayed on the tablet while the robot verbally delivers tutoring information. Upon reaching the allotted time, the robot congratulates the student for completing a session.

4. METHODS

In this section, we describe a user study investigating the effects of different personalized strategies for determining break timing, as employed by the robot tutoring system described in Section 3, on students' learning outcomes.

4.1 Evaluation Context

The user study was contextualized in a tutoring interaction in which children learned about mathematical concepts and then practiced these concepts by completing problems with the robot, thereby creating a repetitive learning interaction. Students completed a 40-minute learning interaction to approximate the length of a scheduled class period during an elementary school day. We present the educational content below, and provide a description of the non-task break activities that were used across all experimental conditions.

4.1.1 Educational Content

We chose to teach two math concepts involved in "order of operations" that the students had not previously learned in their classrooms. Specifically, the students learned that multiplication comes before addition and subtraction (C_1: *multiplication*), as well as the concept that parentheses come before all other operations, including multiplication (C_2: *parentheses*). We designed practice problems for each of the two concepts for three difficulty levels; examples are provided in Table 1. Students had to complete a minimum of ten questions per difficulty level. Moreover, they needed to

Table 1: Examples of practice math problems for each of the concepts and difficulty levels given to the students.

	C_1: Multiplication	C_2: Parentheses
Level 1	$2 + 8 \times 2$	$5 \times (2 + 4)$
Level 2	$5 + 6 \times 1 + 6 \times 4$	$6 + (2 + 6) \times 7$
Level 3	$31 + 5 \times 9 - 7 \times 4$	$8 + 8 \times (1 + 5) + 3$

(a) Game play: tic-tac-toe   (b) Physical exercise: stretch activity   (c) Refocusing activity: spot-it   (d) Relaxation: breathing activity
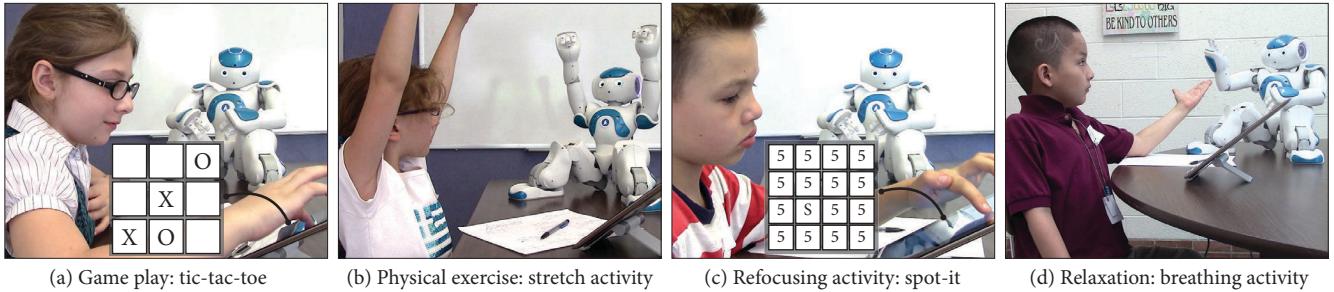
Figure 3: Four break activities that were received during the tutoring session. Each break activity lasted about two minutes.

achieve 70% accuracy to be considered to have mastery of that difficulty level before advancing to the next level.

4.1.2 Break Activities

Throughout the tutoring session, the robot provided the students with brief breaks from the learning activity. We designed four break activities that leveraged the physical and social nature of the robot, including game play, physical exercise, a refocusing activity, and relaxation, which all students could receive in the same order (Figure 3). Each activity lasted approximately two minutes and aimed to provide mental "rest" from the math-based task. The stretch and relaxation breaks were specifically designed to be two minutes long, whereas children completed as many rounds of tic-tac-toe and the visual focus activity as they could within two minutes. The robot engaged with the child during each activity; the robot played tic-tac-toe against the child, led the child in the stretch and relaxation exercises, and facilitated each round of the refocusing activity. Students were not informed in advance that they would be receiving breaks, thereby eliminating any initial expectation for breaks.

4.2 Experimental Design

We designed a between-subjects study involving three experimental conditions—*fixed*, *reward*, and *refocus*—that realized the three strategies described in Section 3.2. The only independent variable in this study was the timing of the breaks. Table 2 summarizes the implementation of the activity scheduler, listing the triggers that initiated breaks in the experimental conditions. Below, we provide detailed descriptions of the conditions and our implementation of the triggering mechanisms used in all conditions.

4.2.1 Fixed Condition

In the *fixed* condition, the robot provided a break at regular intervals for each student regardless of their real-time performance on the learning task. This design reflects the classroom practice that all students get breaks at the same time as everyone else, regardless of an individual's need for a break. Acknowledging the short attention spans of children [23], we implemented the fixed strategy by providing a break every six minutes, allowing most participating students to receive the four distinct breaks over the 40-minute session.

4.2.2 Reward Condition

The *reward* condition, as informed by the educational practice of "success-based rewards" [24], implemented the reward strategy as described in Section 3.2. In this condition, the robot provided a break to the user upon detection of substantial improvement during the session. The performance monitor measured learning performance using two quantities: the accuracy of the student in answering questions correctly, and the time it took for the student to complete each question. A local *window* of recent history (five questions) of user performance (see Figure 4) was kept for both accuracy data and timing data as the session progressed. After each question, the local window data was compared to the entire *history* of data to understand whether there had been a performance increase in accuracy and efficiency (timing). The history data was reset when the difficulty level changed.

The implementation of the reward strategy can be represented conceptually as a decision tree (Table 2). First, the strategy considered increases in accuracy (type 2). Subsequently, the strategy would consider improvements in timing (type 3). Based on whether sizable changes (20%) occurred with the local windows of accuracy and timing as compared to the whole history, a break was given to the participant. If no substantial performance changes occurred for ten consecutive questions, but the participant's overall history of accuracy remained high ($\geq$70%), the student received a break for performing consistently well (type 4).

4.2.3 Refocus Condition

The *refocus* condition, as informed by "positive time-out" [18], provided a break upon detecting performance drops. The implementation of this condition also relied on the performance monitor calculating a local window of accuracy data as well as timing data in the same way as previously described for the reward condition. The conceptual decision tree structure applies here, as well: we first considered
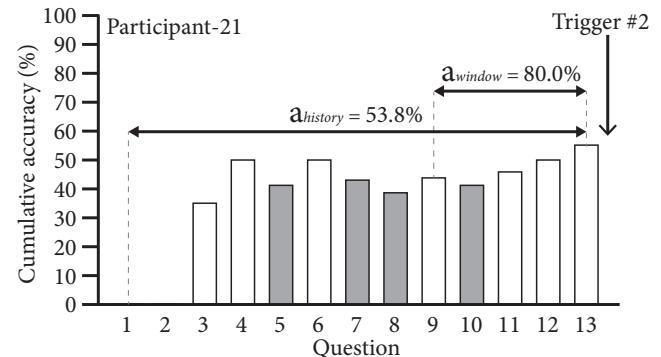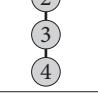

Figure 4: Sample user accuracy data showing how *window* and *history* values were calculated. The window size was five questions. Here, trigger type 2 was initiated because the percent change between $\mathbf{a}_{history}$ and $\mathbf{a}_{window}$ was greater than 20%. Gray bars represent incorrect answers.

Table 2: Break triggering mechanisms for each condition. Percent changes from history to window data for accuracy and timing are represented by $\Delta Accuracy$ and $\Delta Time$, respectively. Overall history accuracy is denoted by $\mathbf{a}_{history}$. N is the number of times each type was triggered. The conceptual decision tree shows the order in which the triggers were considered.

Condition	Trigger type	Description	Implementation	N	Conceptual decision tree
Fixed	1	Fixed schedule	Six minutes elapsed	39	N/A
Reward	2	Accuracy improvement	$\Delta Accuracy \geq 20\%$	43	
	3	Efficiency improvement	$\Delta Time \leq -20\%$, $\mathbf{a}_{history} \geq 70\%$, $\|\Delta Accuracy\| \leq 20\%$	0	
	4	Good performance over time	$\{\|\Delta Accuracy\|, \|\Delta Time\|\} \leq 20\%$ for 10 questions, $\mathbf{a}_{history} \geq 70\%$	1	
Refocus	5	Efficiency drop	$\Delta Time \geq 20\%$, $\|\Delta Accuracy\| \leq 20\%$	14	
	6	Low performance over time	$\{\|\Delta Accuracy\|, \|\Delta Time\|\} \leq 20\%$ for 10 questions, $\mathbf{a}_{history} < 70\%$	2	
	7	Timing drop, indicating guessing	$\Delta Time \leq -20\%$, $\mathbf{a}_{history} < 70\%$, $\|\Delta Accuracy\| \leq 20\%$	15	
	8	Accuracy drop	$\Delta Accuracy \leq -20\%$	13	

drops in accuracy (type 8), followed by changes in efficiency as measured by time to complete each problem (types 5 and 7). If the drop in performance between the local window and the entire history was sufficient (20%), a break was triggered. If no sizable performance drops (20%) occurred but the participant's overall history of accuracy remained low (under 70%) for ten questions in a row, the student received a break for low overall performance over time (type 6).

4.3 Experimental Procedure and Setup

Both parental and child consent for each student was obtained prior to conducting this study. Additionally, children were informed that there were no negative consequences for stopping the interaction at any time. Participating students were removed from their classrooms one at a time for the duration of approximately one hour each. Students were first asked to complete a pretest, consisting of 12 questions, to assess their knowledge of the learning concepts. Free-response questions were used to prevent students from answering correctly due to guessing. After the pretest, students engaged with the robot in a 40-minute tutoring session. This session consisted of a short lesson from the robot, followed by a series of practice problems for students to complete. According to the experimental conditions, the robot provided corresponding breaks throughout the session. If more than four breaks were triggered in a single tutoring session, the break activities would repeat starting from the first one.

During the tutoring interaction, students sat at a table in front of the robot and the tablet (Figure 1). Each child interacted exclusively with the autonomous robot tutoring system during the session, requiring no input from the experimenter in the study room. After the tutoring session, students completed a posttest to assess their knowledge of the learning concepts. Both the pretest and posttest were the same length, including four questions of each of the three difficulty levels, and were identical except the order of the questions. Students also completed a brief questionnaire about their experience with the robot. Students were given pencils and stickers after completing the entire study and returned to their classrooms.

4.4 Measures

To explore how personalized break timing may impact learning outcomes, we employ three objective measures: (1) *learning gains*, (2) *efficiency* in problem solving, and (3) *accuracy* in problem solving. We define normalized learning gain (nlg) that captures the normalized difference between pretest and posttest scores for each student i:

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

Both the pretest and posttest scores are represented as accuracy scores calculated by dividing the number of questions answered correctly by the total number of questions. This nlg metric, ranging from -1.0 to 1.0, provides an index of improvement for each student, accounting for differing incoming knowledge levels.

In addition to learning gains, we seek to understand whether the break activities have any immediate effects on student performance in completing each problem. To this end, we calculate average efficiency and accuracy in solving problems, using a window of five problems, before and after each break. As breaks were initiated by different trigger types (Table 2), we assess the difference between performance before and after the breaks separately for each trigger type.

4.5 Participants

Forty students were recruited from elementary schools to participate in this study and were randomly assigned to one of three experimental conditions. Two participants were excluded from this data analysis due to non-compliance and technical problems during data collection, resulting in a total of 38 participants (13 females). Among the 38 participants, there were 12, 14, and 12 participants in the fixed, reward, and refocus conditions, respectively. The participating students were in third grade; the average age was 8.53 years old ($SD=.60$). The groups were gender balanced, and there were no major differences found between the three conditions regarding age. Pretest scores for the three groups were: Fixed ($M = .33$, $SD = .33$); Reward ($M = .18$, $SD = .28$); Refocus ($M = .25$, $SD = .18$). A one-way ANOVA showed no statistical differences between the three groups, $F(2, 35) = 1.04$, $p = .363$, regarding the pretest.

5. RESULTS

In this section, we first present findings characterizing how the robot tutoring system was used by students, to provide a
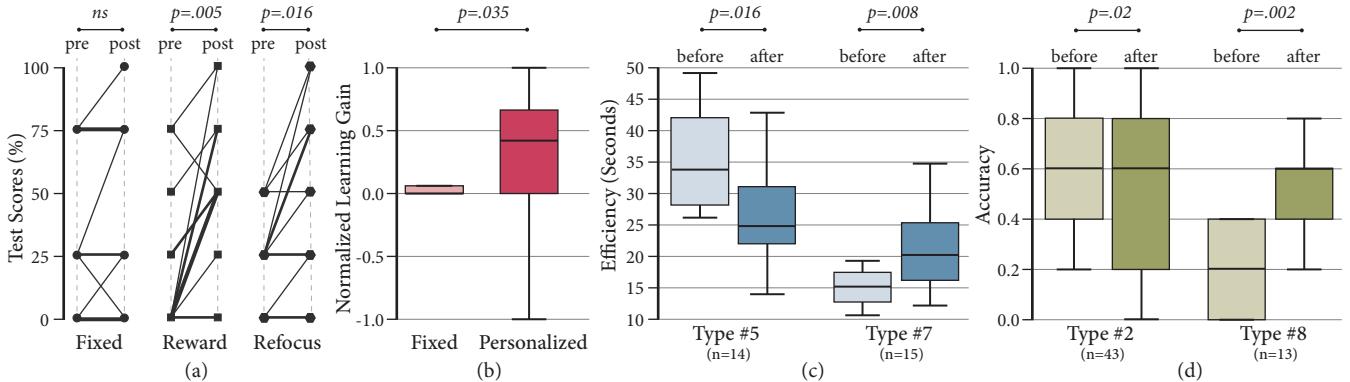
Figure 5: Results show the benefits of personalized break timing strategies: (a) Both the reward and refocus groups significantly improved in learning gains from pretest to posttest. Thicker lines indicate multiple participants with the same scores. (b) The personalization group (reward and refocus groups combined) improved significantly more than the fixed group, as measured by *nlg*. (c) Trigger types 5 and 7 significantly affected how much time students spent on problems before and after breaks. (d) Trigger types 2 and 8 significantly affected accuracy scores before and after breaks. For all boxplots, the darker line inside the box represents the median, and the extents of the box represent the first and third quartiles.

basis for our further data analyses. We then present results on student learning gains and performance during problem solving (summarized in Figure 5). For all the statistical tests reported below, we used an α level of .05 for significance. We used non-parametric statistical tests when appropriate according to the distribution of the analyzed data.

5.1 Characterization of Tutoring Sessions

Participating students were from differing backgrounds and were not able to make homogeneous progress throughout the sessions. Across the three conditions, 52.6% of the students remained in level one for the entire session, 26.3% progressed to level two, and 21.1% were able to progress to level three. Due to such diversity, in the following section, we focus our analyses on student performance on level one assessment questions to draw fair comparisons. While students received a varying number of break activities according to the timing strategies, the average number of breaks provided per student was 3.74, and this was not significantly different across conditions, $F(2, 35) = .749$, $p = .480$. Moreover, the number of breaks received and normalized learning gain were not correlated for participants in all groups, $r(36) = .10$, $p = .540$ (Pearson correlation). Furthermore, not all types of triggers were initiated equally during the sessions. The number of times that triggers were initiated is summarized in Table 2. In the reward condition, trigger type 2 (accuracy improvement) was initiated most frequently, whereas type 3 was never initiated and type 4 was only initiated once. In the refocus condition, trigger types 5, 7, and 8 occurred at a comparable rate, while type 6 was only observed twice. Accordingly, we considered trigger types that were initiated more than twice, namely types 1, 2, 5, 7, and 8, in our analyses of the effects of each trigger type.

5.2 Learning Gains

As less than 50% of the students advanced past the level-one difficulty, we focused our analysis on level-one questions to understand learning gains on content that all students spent time practicing with the robot. To assess whether students improved their scores from pretest to posttest, we used Wilcoxon Signed-ranks tests, treating the test score as a within-subjects measure, to assess each student's learn-

ing gains over the course of the tutoring session. Figure 5 (a) shows each student's score on the level-one difficulty questions on both the pretest and the posttest, separated according to experimental condition. Students in the fixed condition had posttest scores (Mdn = 1.0, $IQR = 3.0$) that did not differ significantly from their pretest scores (Mdn = 1.0, $IQR = 3.0$), $Z = -1.134$, $p = .257$. For students in the reward condition, posttest scores (Mdn = 2.0, $IQR = 1.25$) were significantly higher than pretest scores (Mdn = 0.0, $IQR = 1.25$), $Z = -2.829$, $p = .005$. Posttest scores (Mdn = 2.0, $IQR = 2.0$) were significantly higher than pretest scores (Mdn = 1.0, $IQR = 1.5$) for students in the refocus condition as well, $Z = -2.401$, $p = .016$. These results together show that the students who received personalized break timing strategies, either *reward* or *refocus*, significantly improved their scores, while the fixed group did not. These results provide evidence indicating that the personalization of when to provide breaks during a tutoring interaction can positively impact learning.

Next, we compared normalized learning gains, *nlg*, between groups using a Kruskal-Wallis test. This analysis showed marginal difference in *nlg* between the different conditions, $H(2) = 5.086$, $p = .079$. To further understand the potential benefits of personalized timing strategies, we created a *personalized* group by combining the reward and refocus groups. We then ran a Mann-Whitney test comparing *nlg* between the fixed and personalized groups (Figure 5 (b)). This comparison showed that the normalized learning gain was significantly greater for the personalized group (Mdn = .41, $IQR = .69$) than for the fixed group (Mdn = 0.0, $IQR = .19$), $U = 89.000$, $p = .035$.

5.3 Immediate Break Effects

Besides overall learning gains, we were interested in whether certain trigger types provided immediate effects on efficiency or accuracy during the tutoring interaction. Thus, we compared average performance (both efficiency and accuracy) for the window of five problems before and after each break separately for each trigger type using paired t-tests. Results for this analysis are in Table 3. These comparisons were carried out for each distinct trigger type as we sought to un-

Table 3: Results of t-tests used to assess immediate effects of breaks on accuracy and efficiency by trigger type. (*) and (**) denote $p < .050$ and $p < .010$, respectively. Significant results are shaded in green.

Trigger type	Description	N	Accuracy			Efficiency (Seconds)		
1	Fixed schedule	39	Before	After		Before	After	
			M=0.50 SD=0.27	M=0.52 SD=0.35	$t(38)=-0.387$ $p=.701$	M=45.42 SD=26.68	M=42.39 SD=26.21	$t(38)=1.212$ $p=.233$
2	Accuracy improvement	43	Before	After		Before	After	
			M=0.62 SD=0.24	M=0.50 SD=0.33	$t(42)=2.412$ $p=.02*$	M=24.92 SD=10.39	M=25.04 SD=10.79	$t(42)=-0.089$ $p=.929$
5	Efficiency drop	14	Before	After		Before	After	
			M=0.54 SD=0.28	M=0.50 SD=0.37	$t(13)=0.763$ $p=.459$	M=35.33 SD=8.10	M=26.66 SD=8.38	$t(13)=2.765$ $p=.016*$
7	Timing drop, indicating guessing	15	Before	After		Before	After	
			M=0.51 SD=0.33	M=0.39 SD=0.28	$t(14)=1.890$ $p=.080$	M=16.85 SD=6.77	M=21.65 SD=7.56	$t(14)=-3.064$ $p=.008**$
8	Accuracy drop	13	Before	After		Before	After	
			M=0.22 SD=0.17	M=0.52 SD=0.28	$t(12)=-3.987$ $p=.002**$	M=38.28 SD=20.05	M=34.96 SD=16.90	$t(12)=1.307$ $p=.216$

derstand how different timings of breaks (i.e., different types of trigger) might shape student efficiency and accuracy.

5.3.1 Efficiency in Problem Solving

Our analysis revealed that trigger types 5 and 7 had significant effects on how much time students spent on problems before and after breaks, as summarized in Figure 5 (c). Trigger type 5 was initiated when students' efficiency dropped while there were no sizable changes in their accuracy. Breaks triggered by type 5 improved students' efficiency significantly, as students spent significantly less time solving problems after the breaks as compared to the time they spent before the breaks. This result indicates that providing a break after this trigger may refocus the students, thus leading to improved efficiency in solving problems.

Trigger type 7 represented the situation where students' overall accuracy was not desirable, yet they spent less time on problems at hand, suggesting guessing on answers without investing time into each problem. Our analysis revealed that students spent significantly more time after breaks initiated by trigger type 7 than they did before. This increase in time spent on problems may suggest that after the breaks students were able to refocus their attention on the math task. Finally, for trigger types 1, 2, and 8, there was no significant difference in efficiency before and after the breaks.

5.3.2 Accuracy in Problem Solving

Results showed that trigger types 2 and 8 had significant effects on average accuracy before and after the breaks (Figure 5 (d)). Trigger type 2 was initiated upon detection of a local increase in accuracy, indicating that a student received a break based on this trigger while improving performance. The results of the t-test showed that average accuracy decreased after trigger type 2 was initiated. While the causes of this drop were not certain, we speculated that these breaks may have distracted some students in the short-term as they received them when they were in the "flow" of improving.

In the refocus condition, trigger type 8 was initiated specifically after a local performance drop was detected. Our analysis showed that students significantly improved their accuracy from before to after the breaks triggered by type 8.

This improvement in accuracy following these breaks further suggests the restorative effects non-task breaks may have on performance during learning when triggered effectively. For trigger types 1, 5, and 7, there was no significant difference in accuracy before and after the breaks.

5.4 Additional Observations

In addition to the statistical analyses on student learning gains and perfomance in problem solving, we also made several observations and formed a preliminary understanding of students' experience of interacting with our tutoring system. Overall, we observed that the students were very engaged with the robot. They glanced periodically toward the robot and ocassionally touched the robot during the sessions. We also observed students enjoying interacting with the robot by expressing smiles and laughing. Additionally, most students followed the robot's instructions during breaks to stretch their bodies and participated in the relaxation activity.

Students rated their experience with the robot tutoring system positively on 5-point scales. There were no significant differences on these ratings between the study groups. In particular, students in the fixed ($M = 4.83$, $SD = 0.39$), reward ($M = 4.50$, $SD = 0.94$), and refocus ($M = 4.42$, $SD = 0.79$) groups felt refreshed after the provided breaks, $F(2, 35) = 1.028$, $p = .368$. Students also wanted to have a similar robot tutor to help with their math homework (fixed: $M = 4.42$, $SD = 1.24$; reward: $M = 4.50$, $SD = 1.02$; refocus: $M = 4.75$, $SD = 0.62$), $F(2, 35) = .368$, $p = .695$.

6. DISCUSSION

In this paper, we explore when a tutoring robot should provide a break to children to promote their learning. To this end, we developed an autonomous robot system to support tutoring interactions. We then conducted a field study to use our robot system to engage young students in learning math concepts. Our results showed that personalized timing strategies for providing breaks during learning can benefit students' learning gains. We further found that some strategies can lead to immediate performance benefits. Below, we discuss challenges in robot-child tutoring and the benefits

of personalization we observed in our user study. We also discuss the limitations of this work that inform future work aiming to enable effective, personalized robot-child tutoring.

6.1 Challenges in Robot Tutoring

The major challenges in autonomous robot-child tutoring center around the vast individual differences between children during learning. In our study, less than half of the students progressed past the questions of the starting difficulty level. This observation indicates the diverse learning abilities present among children. In addition, children are likely to have differing preferences for teaching strategies of the robot and break activities they enjoy. For example, *learning by teaching* is an alternative teaching strategy that has been explored in HRI and may benefit learning for children [7, 31]. An effective robot tutor should therefore accommodate these potential differences in learners. Moreover, these robot tutors can employ nonverbal behaviors selectively to facilitate learning. For instance, Huang and Mutlu demonstrated that a robot emphasizing different types of gestures can shape how well the users remembered the information presented by the robot [8]. These challenges in diverse backgrounds and varying preferences provide rich opportunities for designing personalized interaction to promote learning.

6.2 Benefits of Personalization

Contributing to the increasing evidence showing the benefits of personalization in human-robot tutoring (e.g., [14, 26]), results of this work demonstrate the positive impact personalized break timing has on learning outcomes. While children in all conditions seemed to enjoy the breaks, only those in the personalized (reward and refocus) conditions showed significant learning gains (Figure 5 (a)). Moreover, the children with breaks on a personalized schedule outperformed those with breaks on a fixed schedule in terms of learning gains (Figure 5 (b)). These results show the importance of break timing during a learning interaction. Although we did not observe significant learning differences emerge as a result of employing different personalized strategies, our implementation of the reward and refocus strategies provide insight into how these strategies might be realized. Additional work is needed to explore alternative personalized strategies, such as a combination of the reward and refocus strategies, as well as other plausible implementation.

Our analysis also revealed that certain break triggers led to immediate changes in efficiency and accuracy during the tutoring interaction, providing design implications for robot tutoring systems. Specifically, breaks triggered based on negative performance changes led to desired immediate effects during learning, showing that performance-based metrics are useful features for providing breaks for cognitive rest. For example, after taking a break initiated by trigger type 5 (efficiency drop possibly signifying a negative affective state, such as disengagement), students improved their efficiency in problem solving, indicating a potential restorative effect following the break. Similarly, providing a break after trigger type 7 (timing drop, potentially due to guessing) prompted students to spend more time on problems, suggesting the break's potential to refocus the students on the learning task. Finally, breaks provided after trigger type 8 (drop in accuracy, possibly signifying frustration or confusion) led to an increase in accuracy following the break, again showing the potential of these breaks to refocus young learners. Together, these findings showed that the refocus strategy providing "positive time-out" had a positive impact on immediate learning performance.

However, these personalized triggers must be carefully designed, as not all of them led to positive performance changes. Initiating trigger type 2 to provide a break when students were showing improved accuracy caused accuracy to drop after the break. Interestingly, students receiving breaks as success-based rewards still benefitted over the whole session as evidenced by their improved test scores. More research is necessary to obtain a more comprehensive understanding of the broader effects of this type of trigger.

6.3 Limitations and Future Work

While showing benefits of personalized timing strategies for providing breaks, this work has limitations that inform future research. In this work, we contextualized our study in a single 40-minute one-on-one tutoring interaction. Participating students only interacted with our robot system once to learn two specific math concepts. Future research needs to consider how such tutoring systems may help children learn over several sessions and on a variety of learning topics. Moreover, we employed performance metrics of accuracy and efficiency to drive personalized strategies. While these metrics directly indicate how well the student learned, they do not capture information about visual attention, cognitive load, and other related processes involved in learning. Future work should include a variety of channels of information (e.g., EEG [30] or head pose [13]) to more accurately model user state during learning. A richer representation of user state allows a robot tutor to provide a greater personalized learning experience. In addition, the robot's social character and role can affect a child's engagement during a learning task [35]. Future work should explore how the social dynamics and the role of the robot (peer or teacher) during tutoring may influence learning outcomes. Lastly, this study focused specifically on when to provide cognitive rest during tutoring. Additional work should explore different implementations of various types of breaks.

7. CONCLUSION

In this paper, we investigated personalized timing strategies for providing non-task breaks to children during a robot tutoring interaction. We built an autonomous robot tutoring system and conducted a field study to compare the effectiveness of different break timing strategies in promoting learning. Results from our study showed that students receiving personalized strategies were able to improve their test scores from pretest to posttest significantly more than those receiving breaks on a fixed schedule. Furthermore, we found that certain types of break triggers provided immediate benefits for students to solve problems more efficiently and accurately. Results of this work have positive implications for creating effective, personalized tutoring interactions.

8. ACKNOWLEDGMENTS

9. REFERENCES

[1] A. Ariga and A. Lleras. Brief and rare mental "breaks" keep you focused: Deactivation and reactivation of task goals preempt vigilance decrements. *Cognition*, 118(3):439–443, 2011.

[2] D. I. Cordova and M. R. Lepper. Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of educational psychology*, 88(4):715, 1996.

[3] F. N. Dempster. The rise and fall of the inhibitory mechanism: Toward a unified theory of cognitive development and aging. *Developmental review*, 12(1):45–75, 1992.

[4] G. Gordon, S. Spaulding, J. K. Westlund, J. J. Lee, L. Plummer, M. Martinez, M. Das, and C. Breazeal. Affective personalization of a social robot tutor for children's second language skills. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[5] J. Greczek, E. Short, C. E. Clabaugh, K. Swift-Spong, and M. Mataric. Socially assistive robotics for personalized education for children. In *AAAI Fall Symposium on Artificial Intelligence and Human-Robot Interaction (AI-HRI)*, 2014.

[6] O. A. B. Henkemans, B. P. Bierman, J. Janssen, M. A. Neerincx, R. Looije, H. van der Bosch, and J. A. van der Giessen. Using a robot to personalise health education for children with diabetes type 1: A pilot study. *Patient education and counseling*, 92(2):174–181, 2013.

[7] D. Hood, S. Lemaignan, and P. Dillenbourg. 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*, pages 83–90. ACM, 2015.

[8] C.-M. Huang and B. Mutlu. Modeling and evaluating narrative gestures for humanlike robots. In *Robotics: Science and Systems*, pages 57–64, 2013.

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

[10] J. Kennedy, P. Baxter, E. Senft, and T. Belpaeme. Higher nonverbal immediacy leads to greater learning gains in child-robot tutoring interactions. In *International Conference on Social Robotics*, pages 327–336. Springer, 2015.

[11] I. Leite, M. McCoy, D. Ullman, N. Salomons, and B. Scassellati. Comparing models of disengagement in individual and group interactions. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 99–105. ACM, 2015.

[12] I. Leite, A. Pereira, G. Castellano, S. Mascarenhas, C. Martinho, and A. Paiva. Modelling empathy in social robotic companions. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 135–147. Springer, 2011.

[13] S. Lemaignan, F. Garcia, A. Jacq, and P. Dillenbourg. From real-time attention assessment to with-me-ness in human-robot interaction. In *The Eleventh ACM/IEEE International Conference on Human Robot Interation*, pages 157–164. IEEE Press, 2016.

[14] 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*, pages 423–430. ACM, 2014.

[15] 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)*, 2012.

[16] J. Lim and K. Kwok. The effects of varying break length on attention and time on task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, page 0018720815617395, 2015.

[17] K. E. McKay, J. M. Halperin, S. T. Schwartz, and V. Sharma. Developmental analysis of three aspects of information processing: Sustained attention, selective attention, and response organization. *Developmental Neuropsychology*, 10(2):121–132, 1994.

[18] J. Nelsen. *Positive Time-out: And Over 50 Ways to Avoid Power Struggles in the Home and the Classroom*. Harmony Books, 1999.

[19] R. Pekrun, T. Götz, W. Titz, and R. P. Perry. Academic emotions in students' self-regulated learning and achievement: a program of qualitative and quantitative research. *Educational Psychologist*, 37(2):91–105, 2002.

[20] A. D. Pellegrini and C. M. Bohn. The role of recess in children's cognitive performance and school adjustment. *Educational Researcher*, 34(1):13–19, 2005.

[21] A. D. Pellegrini, P. D. Huberty, and I. Jones. The effects of recess timing on children's playground and classroom behaviors. *American Educational Research Journal*, 32(4):845–864, 1995.

[22] 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*, pages 1253–1256. International Foundation for Autonomous Agents and Multiagent Systems, 2008.

[23] G. Petty. Teaching today: A practical guide. 2009.

[24] W. D. Pierce and J. Cameron. A summary of the effects of reward contingencies on interest and performance. *The Behavior Analyst Today*, 3(2):221, 2002.

[25] A. Powers, S. Kiesler, S. Fussell, and C. Torrey. Comparing a computer agent with a humanoid robot. In *Human-Robot Interaction (HRI), 2007 2nd ACM/IEEE International Conference on*, pages 145–152. IEEE, 2007.

[26] A. Ramachandran, A. Litoiu, and B. Scassellati. Shaping productive help-seeking behavior during robot-child tutoring interactions. In *The Eleventh ACM/IEEE International Conference on Human Robot Interation*, pages 247–254. IEEE Press, 2016.

[27] M. Saerbeck, T. Schut, C. Bartneck, and M. D. Janse. 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, pages 1613–1622. ACM, 2010.

[28] B. Schadenberg, M. Neerincx, F. Cnossen, and R. Looije. Personalising game difficulty to keep children motivated to play with a social robot: a bayesian approach. *Cognitive Systems Research*, 2016.

[29] E. Short, K. Swift-Spong, J. Greczek, A. Ramachandran, A. Litoiu, E. C. Grigore, D. Feil-Seifer, S. Shuster, J. J. Lee, S. Huang, et al. 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*, pages 924–929. IEEE, 2014.

[30] D. Szafir and B. Mutlu. Pay attention!: designing adaptive agents that monitor and improve user engagement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 11–20. ACM, 2012.

[31] F. Tanaka and S. Matsuzoe. Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning. *Journal of Human-Robot Interaction*, 1(1), 2012.

[32] T. C. Toppino, J. E. Kasserman, and W. A. Mracek. The effect of spacing repetitions on the recognition memory of young children and adults. *Journal of experimental child psychology*, 51(1):123–138, 1991.

[33] J. Wainer, D. J. Feil-Seifer, D. A. Shell, and M. J. Mataric. Embodiment and human-robot interaction: A task-based perspective. In *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication*, pages 872–877. IEEE, 2007.

[34] J. K. Westlund and C. Breazeal. The interplay of robot language level with children's language learning during storytelling. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, pages 65–66. ACM, 2015.

[35] C. Zaga, M. Lohse, K. P. Truong, and V. Evers. The effect of a robot's social character on children's task engagement: Peer versus tutor. In *International Conference on Social Robotics*, pages 704–713. Springer, 2015.