Empowering Learners: Chatbot-Mediated ‘Learning-by-Teaching’

Nihar Sabnis
Sensorimotor Interaction,
Max Planck Institute for Informatics,
Saarland Informatics Campus
Saarbrücken, Germany
nsabnis@mpi-inf.mpg.de

Tomohiro Nagashima
Saarland University
Saarland Informatics Campus
Saarbrücken, Germany
nagashima@cs.uni-saarland.de

ABSTRACT

Chatbots and online learning platforms provide synthesized information to learners. However, research shows learning is particularly effective when learners themselves teach someone. Prior work has explored an interactive instructional approach called ‘Learning-by-teaching’, but this approach traditionally relies on human counterparts, limiting it to their interest and co-located settings. To overcome these limitations, we investigated whether we can empower learners using chatbot-mediated ‘learning-by-teaching.’ We designed an agnostic, open-source chatbot replicating a virtual student, to which learners teach to learn. We conducted an experiment involving 24 students to evaluate the effectiveness of chatbot-mediated teaching compared to textbook-based problem-solving practice. Results indicate that teaching the chatbot benefits student learning than textbook-based problem-solving. This work highlights the effectiveness of chatbots, envisioning their design as virtual students to mediate ‘learning-by-teaching’.

CCS CONCEPTS

• **Applied computing → Interactive learning environments**; *Computer-assisted instruction*; E-learning.

KEYWORDS

Learning-by-Teaching, chatbots, virtual students, teachable agents

ACM Reference Format:
Nihar Sabnis and Tomohiro Nagashima. 2024. Empowering Learners: Chatbot-Mediated ‘Learning-by-Teaching’. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA ’24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3613905.3650754

1 INTRODUCTION

Recent advancements in online education systems and chatbot-mediated teaching and learning have shown potential for interactive learning experiences and engaging students in meaningful ways [30, 38]. However, in this human-chatbot interaction, chatbots are focused on providing synthesized information to the learner [16, 54] and typically play the role of a *teacher* [17, 37]. This creates a one-way pedagogical flow, making learning less interactive.

Research shows that teaching a concept to someone facilitates learning [11, 14]. The process of teaching to learn, called *learning-by-teaching*, is one way to make learning interactive [13]. ‘Learning-by-teaching’ approach is observed in online communities, but its effectiveness is dependent on humans (i.e., one needs another person to engage in ’learning-by-teaching’) [7, 46]. Teaching friends, students, or strangers on the internet is advantageous for learning a topic, but there are limitations to teaching human learners such as collocation in space (virtual or real) and time, the interest of the learners, their level of understanding of the topic, and limited access to learners. In summary, many online learning practices are interactive but do not use learning-by-teaching approach or are based on human-dependent learning-by-teaching practices. Hence, in this work, we explore the concept of using chatbots as virtual *students* to mediate the process of learning-by-teaching.

Although there are chatbots to which students can teach in order to learn [26, 31], their performance compared to traditional learning methods has not been fully evaluated. Hence, we explored the following research question: ‘**Can chatbots effectively mediate the process of learning-by-teaching to enhance student learning?**’.

This work’s primary contribution is an empirical demonstration of the promising effect of chatbot-mediated learning-by-teaching on student learning, where we found that teaching the chatbot significantly enhanced student learning, compared to textbook-based problem-solving. We also describe the process of creating an agnostic teachable chatbot using open source platforms and show an instance of the same which was used in this research.

2 RELATED WORK

We look at the ‘learning-by-teaching’ approach and how chatbots have played a role in the field of education while describing how this research builds on the current literature.

2.1 Learning-by-Teaching with Human Learners

Learning-by-teaching is a pedagogical method in which learners teach a topic to someone to better understand it. Gartner et al. and Cloward show that learners who teach their peers tend to learn better than when they would learn by themselves [10, 15]. One reason for that could be learning for teaching requires the teacher to understand, revise, and organize the material, identify the basic structure, and present the material in an accessible manner to their peers [15, 20]. In addition, there are cognitive benefits due to preparing for teaching, and presenting the material to the

CHI EA ’24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0331-7/24/05
https://doi.org/10.1145/3613905.3650754

tutee [2]. Also, explaining concepts to others with examples and analogies, provides more opportunities to learn, identify gaps and deepen the understanding, as the peers ask for clarification [42]. Moreover, research by Oddo et al., Mayfield et al. and Sutherland et al. has shown positive effects of peer tutoring across subjects and educational settings [32, 35, 45]. Explaining to others offers an opportunity that leads to producing quality explanations. Also, the tutor needs to assess their own knowledge and evaluate their explanations. Finally, Roscoe et al. showed that the use of examples and analogies, which are often used when teaching, help the tutors to deepen their understanding [42].

Despite the potential benefits, practicing in-person learning-by-teaching is affected by the learner's knowledge and engagement [41]. Peers may not be good at articulating their thought processes, internal thinking, and struggles, and without support, learning-by-teaching becomes difficult [21]. Moreover, learning-by-teaching with human peers depends on the teacher-student collocation in space (virtual or real) and time for the teaching to take place in an interactive manner [47]. Human-dependent learning-by-teaching is also limited in scale, since only a certain number of students can participate as a teacher in the process. In addition, students participating in in-person learning practices face the fear of being judged and may feel embarrassed [12, 51]. This research overcomes some of the limitations of human-dependent 'learning-by-teaching', by using chatbot mediated learning-by-teaching approach.

2.2 Chatbot-mediated 'Learning-by-Teaching'

Technology helps to overcome some of the limitations of human-dependent learning-by-teaching, by developing instructive agents, which focus on providing advice and examples to teach [28, 33]. Specifically, chatbots have been used in education as conversational pedagogical agents based on a dialog-based conversation with learners [25, 52]. Over the years, chatbots have been designed as interaction partners such as teaching agents, peer agents, and teachable agents [22]. Chatbots as *teaching agents* play the role of a teacher in the conversation by providing information, feedback, asking questions, or giving examples to the user [23, 50]. Chatbots as *peer agents* function as learning companions for students, promoting interactions between peers. Students typically initiate the conversation with peer agents to look up certain definitions or ask for an explanation of a specific topic [9, 27]. However, the teaching and peer agents typically act as the *teacher* and the learners take the role of a *student*, which limits active engagement of the learner [49].

Finally, teachable agents are computer agents that allow students to teach (i.e., teachable agents would then learn based on students' teaching), thus improving the learning of students [5]. Chatbots as teachable agents were developed to ask questions to students and help them learn certain concepts and tasks [26, 31]. Ogan et al. looked at the impact of the way students interact with teachable agents, finding that treating the conversational agent as a partner enhanced student learning [36]. As with human counterparts, it was observed that students make greater effort to learn for their teachable agents than they do for themselves [8, 40]. Although teachable agents have certain benefits, a limitation is that they do not take the initiative in the interactions with students, leading to less interest of students, which may hamper their learning [5, 36].

There is limited research on the use of chatbots as teachable agents [22] and teachable chatbots have not been evaluated compared with traditional learning approaches in terms of learning outcomes in students. This research demonstrates an interactive teachable chatbot and addresses whether teaching a chatbot improves student learning compared with traditional learning approach of textbook-based problem-solving.

3 CHATBOT DESIGN AND DEVELOPMENT

Here, we outline the process of developing an agnostic chatbot acting as a virtual student, using open source platforms.

3.1 Purpose and Scope

We developed a chatbot to support 'learning-by-teaching' for the topic of percentages. We chose the domain of mathematics, as the difficulty of the problems to be solved can be structured around different aspects of students' learning such as memory retrieval, concept understanding, and application of knowledge in a new, transfer context [44]. Specifically, we focused on teaching 'What is 75 percent of 3000?' to the chatbot in a step-by-step manner, as this approach has been shown to be effective in literature [34]. We chose a word problem due to its complexity [6]. We chose the topic of percentage because it combines understanding the problem statement, converting the problem statement from text (i.e., word problem) to its corresponding mathematical form, structuring the equation, and calculating the solution. Moreover, it requires an understanding of concepts such as fractions, division, and multiplication to solve the problem.

3.2 Field Observation

We observed two classrooms where sixth-grade students were taught the topic of percentages, and then we interviewed 12 students from the classes to understand how students learn the topic, questions they ask, and their pain points for the topic of percentage. The classroom observations were done in a different school and interviews were with different students than the participants of the study described in section 4. The interviews were semi-structured and were recorded and transcribed for informal analysis. The interviews consisted of asking them to teach "What is 75 percent of 3000?" to one of the authors and the problems they faced while being taught the topic of percentages. The school's ethical board approved the classroom observation and interviews. Also, the parents of the participating children were informed about the interview. We observed that students found it difficult to convert the problem statement into a mathematical equation, to understand the corresponding mathematical operations suggested by the words: 'percent' and 'of', identify the divisor and the dividend, and division when the divisor is greater than the dividend. We noted that students had varying definitions of division and percent, and yet, all of them were correct, for example: "Division is how many times of one number goes into the other number" or "Division is the opposite of multiplication. 5 times 2 is 10, whereas 10 divided by 5 is 2".

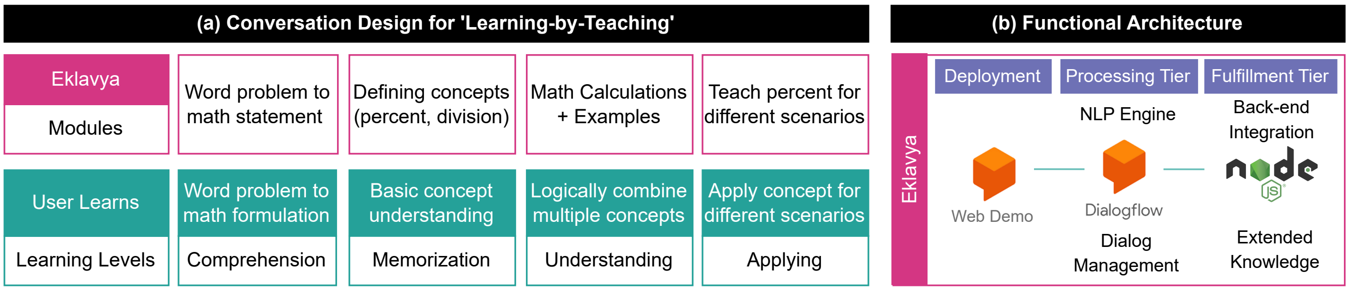

Figure 1: (a) Conversation design focused on learning-by-teaching approach for percentages topic. The modules which need to be taught to Eklavya and the corresponding concepts learned by the user are shown. (b) Functional architecture of Eklavya which consists of the deployment, processing and fulfillment tiers.

3.3 Eklavya: Conversation Design for 'Learning-by-Teaching'

Eklavya acts as an interactive and self-motivated virtual student. The design of Eklavya was based on the observations of the field studies and student interviews. The overall conversation flow is depicted in Figure 1-(a), which presents a conceptual diagram of the modules which need to be taught to Eklavya and the objective of the modules in terms of user learning, and the corresponding levels of learning. Eklavya describes and contextualizes the problem for the user. The problem is broken into steps, where each dialogue is aimed at teaching a particular step in the overall problem. These steps range from converting the word problem to mathematical statements, defining and symbolizing percent, division and putting it together to solve the problem. Snippets of sample conversation are shown in figure 3 *(center, right)* in the appendix. The user progresses towards the solution after answering the questions of Eklavya, defining concepts and providing examples. If the user provides multiple unacceptable answers, Eklavya redirects the user to appropriate videos to understand those portions of the topic better. Finally, the conversation ends when the user solves the problem in a step-wise fashion with the correct final answer.

3.4 Eklavya: Functional Architecture

The functional architecture of Eklavya consists of four key components: the deployment platform, the natural language processing (NLP) engine, dialog management and back-end integration, Figure 1-(b). Eklavya was implemented using Google Dialogflow API V2. Google Dialogflow is Google's cloud-based NLP platform that allows to create chatbots. We decided to use Google Dialogflow to develop Eklavya because of the ease of integration of Dialogflow with other platforms, agnostic over computing devices, NLP to understand the user's intent, context, and sentiment, inbuilt dialog management and analytics. We set up a custom webhook to enable communication between the Dialogflow agent and the back-end and perform necessary calculations. Google Cloud Functionality was used to host the webhook and Node.js was used to implement the calculation logic for division and percentage.

The implementation of Eklavya on Dialogflow API V2 involved using intents, entities, context, and fulfillment. An *intent* signifies the user's intention. We used Dialogflow's NLP to match user input with pre-defined intents, enabling Eklavya to respond or take actions accordingly. Keyword matching triggered multiple intents based on user intent. For unmatched inputs, the fallback intent maintained smooth conversation, providing default responses or asking for more info, like hints for teaching Eklavya. *Entities* are relevant objects/values in user input. For instance, when division was mentioned, we extracted numerator and denominator using entity recognition. *Contexts* maintained conversation information over time to guide the conversation and transfer information between intents. *Fulfillment* made Eklavya respond by invoking a webhook, integrating a back-end calculator to fulfill user requests. We deployed Eklavya using the 'Web Demo' integration, as it did not need any external platforms for hosting and is agnostic irrespective of the computing device.

The graphical user interface (GUI) of Eklavya can be opened on the computing device of user's choice. The top of the GUI (figure 3 *left*) explains the concept which the user needs to teach Eklavya. The GUI consists of a chat window where the entire conversation in the session is captured and can be accessed using the scroll bar. The bottom of the GUI has a bar to type-in or provide audio inputs to Eklavya (audio inputs were not explored in this paper). Eklavya gives text (and audio) responses based on the conversation design and user input. User inputs are displayed with black text and gray boxes, while Eklavya's responses are displayed with white text in green boxes.

4 EMPIRICAL EVALUATION

To evaluate the effectiveness of Eklavya, we conducted a between-subject experiment with sixth-grade students. We followed a pretest-intervention-posttest study procedure, in which participants worked on a web-based test before, and after teaching Eklavya (experimental group) or solving textbook problems (control group) (Figure 2-a). We considered that learning with textbooks would be an appropriate *business-as-usual* condition. **We hypothesized that there would be a difference between both groups' score improvement.** We used a pre-test to understand participants' prior knowledge about the topic of percentages. The post-test, similar to the pre-test, was designed to investigate the impact of the intervention on the participant's understanding of the topic. The pre-test and the post-test included 12 problems, where four problems each focused on different aspects of student learning such as memory retrieval, conceptual understanding, and the application of knowledge in a new context. The pre- and post-test were multiple choice questionnaires, with 4 incorrect and 1 correct answer for each question. We also conducted a group discussion with students after the study to understand their experiences of interacting with Eklavya.

4.1 Participants

24 sixth-grade (mean age = 12.06, SD = 0.58; 14 male, 10 female) students from a school in India who had not yet learned the topic of percentage were recruited through their class teacher. No financial compensation was provided to the participants. The study was conducted virtually with Zoom, where each student had access to a computing device (mobile phones, computers or tablets) on which they could access a link to interact with Eklavya. The students were in separate physical rooms throughout the study. The study was approved by the local ethics committee.

4.2 Procedure

The study consisted of three 1.5-hour Zoom sessions with students in a group, spanning five consecutive days.

Day 1: We taught the topic of percentages to all the students, along with answering any questions they might have in terms of the topic. A short demo of Eklavya was given. Students then worked on the pre-test for 20 minutes. The 12 questions in the pre-test were based on skills such as memory retrieval, understanding, and knowledge transfer.

Day 3: The students were randomly assigned to two groups in zoom break-out rooms, where each group had 12 students. The control group solved 10 textbook problems (from their educational board textbook) on the topic of percentage, while the experimental group taught Eklavya how to solve the problem "What is 75 percent of 3000?". The control group used pen and paper to solve the problems. Later, the tasks completed by both groups were interchanged, as the students who solved textbook problems also wanted to interact with the chatbot. The students who solved textbook problems' interaction with the chatbot were time-boxed to 5 minutes, to minimize the influence on their learning. Finally, all the students were asked to reflect on their experience of teaching Eklavya in a joint group discussion.

Day 5: Students in both groups solve the post-test, consisting of 12 questions, independently. We measured the average time spent by both groups to complete the test. Finally, we conducted a group discussion of sharing experiences and feedback.

4.3 Apparatus

For our study, participants interacted with Eklavya on a computing device of their choice. The pre-and post-test tests were created using Google Forms. The language of communication while interacting with Eklavya and the tests was English, however, to clear conceptual some conceptual doubts of students, Marathi was also used. Zoom was used to conduct and record the online sessions. MATLAB 2021b was used for data analysis. The textbook problems, pre-test, post-test details and student scores are added to the supplementary materials. The designed chatbot can be accessed by uploading the GitHub files to Google Dialogflow [1].

[1]https://github.com/NiharS123/Teach_Eklavya

4.4 Analysis

We used an independent Welch's t-test to compare the overall results of the post-test between the two groups. We used a MANOVA to study the intervention's overall impact across three learning levels. To understand the improvement within each level (remembering, understanding, and applying), we conducted separate ANOVAs, followed by post-hoc tests. Finally, we analyzed the participants' qualitative experience of teaching Eklavya, which had been recorded. We transcribed the recording and conducted qualitative content analysis.

5 RESULTS

We report the results of our study and the experiences of the students during the group discussion.

5.1 Effect on Learning

The results of the pre- and the post-test for both learning methods are summarized in the table 1. Overall results showed that there was a 1.3 times increase in the post-test scores (M = 5.33, SD = 2.58) for the control group compared to the pre-test (M = 2.33, SD = 1.67), whereas the increase for the experimental group post-test (M = 7.92, SD = 2.3) was about 2.3 times from pretest scores (M = 2.42, SD = 1.95). An independent Welch's t-test for the improvement scores of both groups revealed that the experimental group gained significantly more than the control group from the pre- to post-test (*t-statistic* = 2.9784, df = 16.9136, p = 0.01, 95% CI [0.73, 4.27]), indicating a significant effect (*Cohen's d* = 1.22). The control and the experimental groups took median times of 20.89 and 16.40 minutes respectively for completing the post-test, with no significant difference, (p = 0.16).

Improvement across learning categories was calculated by subtracting the pre-test scores from the post-test scores for both groups. The overall improvement for both learning methods is shown in figure 2-b. MANOVA results yielded a statistically significant effect of the learning method on the combined dependent variables of improvement on different aspects of learning (*Wilk's Lambda* = 0.584, $\chi^2(3)$ = 11.26, p = 0.01). This result indicates that the learning method has a significant impact on the improvement levels of remembering, understanding, and applying. Moreover, we conducted one-sample Kolmogorov-Smirnov tests to check for the normality of the improvement scores. The improvement scores of both groups were normally distributed with (*test statistic* = 0.67, *critical value* = 0.37, $p < 0.01$) for the control group and (*test statistic* = 1.00, *critical value* = 0.37, $p < 0.01$) for the experimental group.

To further investigate the effect of learning methods on improvements across different aspects of learning, we conducted multiple univariate ANOVAs. Univariate ANOVA for the **remembering** level revealed a significant effect of the learning method ($F(1, 22)$ = 11.65, $p < 0.01$). Post-hoc Tukey's Honestly Significant Difference (HSD) tests showed that the improvement difference in *remembering* between the two learning methods is statistically significant ($HSD < 0.01$), with the chatbot teaching method outperforming the textbook problem-solving. Similar results were found for **applying** problems, showing a significant effect of learning method ($F(1, 22)$ = 12.31, $p < 0.01$) with ($HSD < 0.01$). However, ANOVA for **understanding** level did not reveal a significant effect of the learning

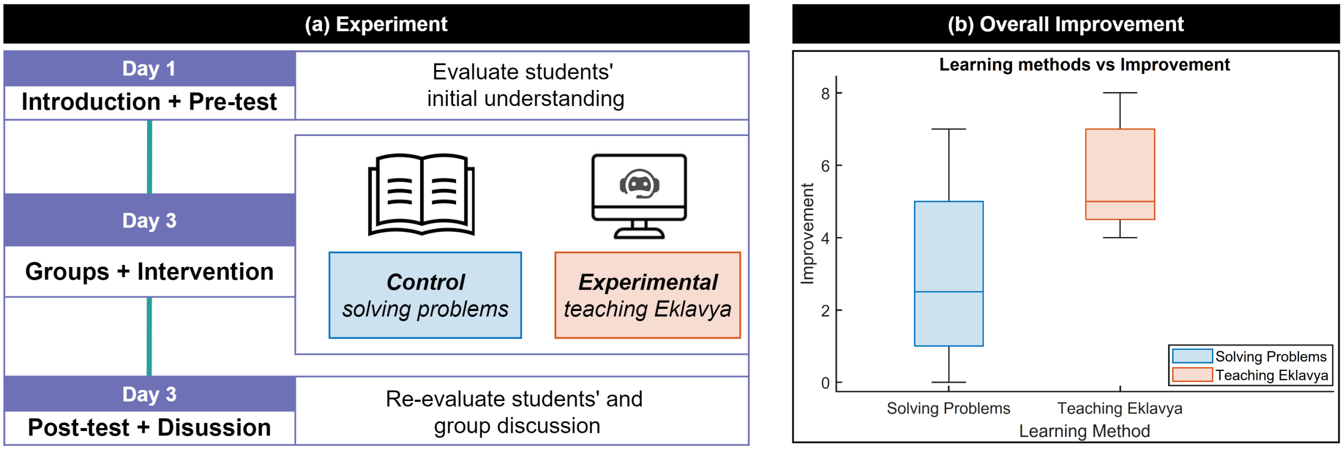

Figure 2: (a) Empirical evaluation following a pretest-intervention-post-test procedure and group discussion for experience sharing and feedback. (b) Results depicting the overall improvement for both the groups based on the learning method of solving textbook problems (control group) and teaching the chatbot (experimental group).

method ($F(1, 22) = 0.88$, $p = 0.36$). Post hoc HSD also did not show statistically significant differences ($HSD = 0.36$).

5.2 Student Experiences and Feedback

Initial Experience: Students expressed curiosity and excitement about their first interaction with Eklavya. They found it a novel method of learning, different from traditional textbook problem-solving. They described their overall experiences using words like, "exciting" and "fun". Moreover, one of the students exclaimed, "This is so much fun!" and another mentioned with excitement, "Would this really replace our homework?". Students also mentioned no fear of embarrassment or being judged when teaching a chatbot.

Teaching Eklavya: The students began to explore the potential of Eklavya by asking generalized questions to gauge its understanding of percentage. They found the process of Eklavya to be different from how they would teach their friends. However, they also mentioned that they could teach Eklavya without the feeling of being judged or embarrassment. They also were happy that they could re-teach if they went wrong at some point or wanted to revise. The subjective questions needed more iterations from the students while teaching, as compared to the objective questions. Eight students were annoyed at times when a certain step was not understood by Eklavya and fallback was triggered, despite multiple attempts of teaching that step. Some students mentioned that they could get the answer early on and hence found it "boring" to go through all the steps. As the students reached the end of the conversation, they felt a feeling of accomplishment. Overall, they described their experiences using words like, "informative", "interactive", and "engaging".

Feedback and Suggestions: The students described that they wished that Eklavya can understand multiple languages, along with changing those languages during the conversation. Further, students were not able to jump between different concepts and their responses were limited to the questions asked by Eklavya which showed a lack of flexibility in teaching the concept. Moreover, four students wanted to use visual tools to explain the concept to Eklavya. Students also would like to have different difficulty levels in Eklavya.

6 DISCUSSION

Our study showed that teaching the chatbot enhanced student learning significantly more than solving textbook problems. We think that the primary reason for the significant difference in improvement between the groups can be attributed to the fact that students in the experimental group had continuous feedback from the chatbot on their responses, whereas no feedback was provided to the control group. This highlights the importance of feedback while learning, as also previously shown in [3, 19]. Also, we speculate that the continuous feedback encouraged the students teaching the chatbot to reflect deeper on the topic as compared with the students who solved textbook questions.

Moreover, the overall improvement between the groups can be attributed to the modular and step-wise approach while teaching the chatbot. The chatbot interaction consisted of intermediate steps, whereas the textbook problems did not have intermediate steps to guide the students while solving problems. This highlights the importance of providing step-by-step guidance during problem-solving [22, 43]. Furthermore, the textbook questions focused on the topic of percentages and not on the concepts such as division and interpreting mathematical statements, which students needed to be familiar with, to learn about percentages. On the other hand, teaching the chatbot was scaffolded well, going from the basic concepts to higher-level concepts, which highlights the importance of procedural knowledge while developing knowledge of mathematical concepts [39]. Finally, however, our results might have been influenced by the novelty effect of interacting with the chatbot as well as worked-example effect, which suggests that novice students benefit by being presented with worked examples (experimental condition) rather than jumping straight to problem-solving (control condition) [48].

Group	Pre-test			Post-test		
	Remembering	Understanding	Applying	Remembering	Understanding	Applying
Control	1.58 ± 0.67	0.42 ± 0.51	0.33 ± 0.49	2.25 ± 0.87	1.92 ± 0.99	1.17 ± 0.72
Experiment	1.58 ± 0.79	0.5 ± 0.67	0.33 ± 0.49	3.25 ± 0.75	2.42 ± 0.79	2.25 ± 0.75

Table 1: Mean and Standard Deviation of pre- and post-test results for the control and experimental groups.

Despite the small sample size, it is worth noting that the study showed significant improvements on the aspects of *remembering* and *applying*, while there was no statistical difference on the aspect of *understanding*. This finding might partially be attributed to the fact that the experimental group focused on one problem, whereas the control group encountered more and varied problems. These differences might indicate that chatbot-based learning benefited cognitive areas of memorization, comprehension, and application differently. For instance, it could be that the improvement in the *remembering* level might be due to the step-by-step learning process when teaching the chatbot, as compared to solving textbook problems where the problem-solving was not necessarily structured step-by-step.

Students describe their experience as initial "curiosity" of interacting with the chatbot, "annoyance" due to certain chatbot characteristics and a "sense of accomplishment" after teaching the chatbot, which indicate higher engagement levels while learning as observed also by Benotti et al. [4]. These higher engagement levels and the active participation in the process of teaching the chatbot contributed to the higher improvement scores for the group teaching the chatbots [40]. Furthermore, interacting with chatbots is more active on the students' part compared to solving textbook problems, and this active interaction may have led to higher experiential learning for students [18, 53]. Moreover, unlike teaching human counterparts, students also mentioned no fear of embarrassment or being judged when teaching a chatbot [12]. Currently, the chatbot leads the conversation, which makes it less compelling for students to teach the chatbot, resulting in fewer explanations from the student's side compared to their explanation to an actual student which was also observed in previous studies [1, 24].

7 LIMITATIONS AND FUTURE WORK

Our current study had a limited number of participants. Furthermore, the students from the control group interacted with the chatbot (even though for just a few minutes) before working on the post-test to accommodate students' needs and desires. Besides, the control group did not receive any form of feedback, which compromises the comparability with the treatment group. Moreover, we targeted the students of a certain education board, which may not have been a representative user group for the study, and hence, the results of this study may not be reproducible in other contexts. Our chatbot also has some technical limitations: artificial conversation, difficulty in analyzing users' subjective inputs, and assessment of reliability of the chatbot.

To see if learning-by-teaching actually has an effect, we plan to compare tutoring chatbot with teachable chatbots. Moreover, we plan to evaluate the effectiveness of chatbot-mediated 'learning-by-teaching' in comparison with in-person 'learning-by-teaching' to observe whether the technical implementation, with its unique features (e.g., step-by-step scaffolding) could produce effects. LLM-based chatbots such as ChatGPT cannot play the role of student effectively (e.g., appendix subsection A.2) and hence we plan to use principles of conversation design to enable the use of LLM-based chatbots which replicate student roles. Moreover, we plan to extend this chatbot to other types of learning (e.g.: languages, programming) and other aspects of learning (e.g.: creativity). Finally, integrating adaptability, step-wise evaluation, and analytics as well as the possibility of keeping track of their individual and collective progress would help learners to map their progress over time.

8 CONCLUSION

We investigated the use of chatbots as virtual students to mediate learning-by-teaching methodology. Our pre-test–intervention–post-test study with 24 students showed that students who taught the chatbot performed significantly better on the post-test compared to the students who solved textbook problems. Further, the chatbot we used is agnostic, open-source, scalable, and cost-effective. With this research, we demonstrate empirical evidence that teaching chatbots as mediators for learning by teaching can facilitate student learning.

ACKNOWLEDGMENTS

We would like to acknowledge Rahul Kulkarni (www.godonew.com) for the initial ideation and constant support during the chatbot development process, Aniruddha Tupe and Kalpesh Kothale with their help to conduct the study and Helena Kilger to discuss the statistics.

REFERENCES

[1] Vincent Aleven and Kenneth R Koedinger. 2000. The need for tutorial dialog to support self-explanation. In *Building dialogue systems for tutorial applications, papers of the 2000 AAAI Fall Symposium*. 65–73.

[2] Linda F Annis. 1983. The processes and effects of peer tutoring. *Human Learning: Journal of Practical Research & Applications* (1983).

[3] Roger Azevedo and Robert Bernard. 1995. Assessing the Effects of Feedback in Computer-Assisted Learning. *British Journal of Educational Technology* 26, 1 (1995), 57–58.

[4] Luciana Benotti, María Cecilia Martínez, and Fernando Schapachnik. 2014. Engaging high school students using chatbots. In *Proceedings of the 2014 conference on Innovation & technology in computer science education*. 63–68.

[5] Gautam Biswas, Krittaya Leelawong, Daniel Schwartz, Nancy Vye, and The Teachable Agents Group at Vanderbilt. 2005. Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence* 19, 3-4 (2005), 363–392.

[6] Anton JH Boonen, Menno van der Schoot, Floryt van Wesel, Meinou H de Vries, and Jelle Jolles. 2013. What underlies successful word problem solving? A path analysis in sixth grade students. *Contemporary Educational Psychology* 38, 3 (2013), 271–279.

[7] Vrasidas Charalambos, Zembylas Michalinos, and Richard Chamberlain. 2004. The design of online learning communities: Critical issues. *Educational Media International* 41, 2 (2004), 135–143.

[8] Catherine C Chase, Doris B Chin, Marily A Oppezzo, and Daniel L Schwartz. 2009. Teachable agents and the protégé effect: Increasing the effort towards learning. *Journal of science education and technology* 18 (2009), 334–352.

[9] Fabio Clarizia, Francesco Colace, Marco Lombardi, Francesco Pascale, and Domenico Santaniello. 2018. Chatbot: An education support system for student. In *Cyberspace Safety and Security: 10th International Symposium, CSS 2018, Amalfi, Italy, October 29–31, 2018, Proceedings 10*. Springer, 291–302.

[10] Robert D Cloward. 1967. Studies in tutoring. *The Journal of Experimental Education* 36, 1 (1967), 14–25.

[11] Claudio G Cortese. 2005. Learning through teaching. *Management Learning* 36, 1 (2005), 87–115.

[12] Nancy Falchikov. 2001. *Learning together: Peer tutoring in higher education*. Psychology Press.

[13] Logan Fiorella and Richard E Mayer. 2013. The relative benefits of learning by teaching and teaching expectancy. *Contemporary Educational Psychology* 38, 4 (2013), 281–288.

[14] Stanley Frager and Carolyn Stern. 1970. Learning by teaching. *The Reading Teacher* 23, 5 (1970), 403–417.

[15] Alan Gartner et al. 1971. Children teach children: Learning by teaching. (1971).

[16] David George Glance, Martin Forsey, and Myles Riley. 2013. The pedagogical foundations of massive open online courses. *First monday* (2013).

[17] Sebastian Hobert and Raphael Meyer von Wolff. 2019. Say hello to your new automated tutor–a structured literature review on pedagogical conversational agents. (2019).

[18] Gwo-Jen Hwang and Ching-Yi Chang. 2021. A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments* (2021), 1–14.

[19] Jiyou Jia. 2009. CSIEC: A computer assisted English learning chatbot based on textual knowledge and reasoning. *Knowledge-Based Systems* 22, 4 (2009), 249–255.

[20] Alison King. 1990. Enhancing peer interaction and learning in the classroom through reciprocal questioning. *American educational research journal* 27, 4 (1990), 664–687.

[21] Camilla Kirkegaard, Agneta Gulz, and Annika Silvervarg. 2014. Introducing a challenging teachable agent. In *Learning and Collaboration Technologies. Designing and Developing Novel Learning Experiences: First International Conference, LCT 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part I 1*. Springer, 53–62.

[22] Mohammad Amin Kuhail, Nazik Alturki, Salwa Alramlawi, and Kholood Alhejori. 2023. Interacting with educational chatbots: A systematic review. *Education and Information Technologies* 28, 1 (2023), 973–1018.

[23] James A Kulik and JD Fletcher. 2016. Effectiveness of intelligent tutoring systems: a meta-analytic review. *Review of educational research* 86, 1 (2016), 42–78.

[24] Susanne P Lajoie and Sharon J Derry. 2013. Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. In *Computers as cognitive tools*. Routledge, 83–114.

[25] Diana Laurillard. 2013. *Rethinking university teaching: A conversational framework for the effective use of learning technologies*. Routledge.

[26] Edith Law, Parastoo Baghaei Ravari, Nalin Chhibber, Dana Kulic, Stephanie Lin, Kevin D Pantasdo, Jessy Ceha, Sangho Suh, and Nicole Dillen. 2020. Curiosity notebook: A platform for learning by teaching conversational agents. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–9.

[27] Lap-Kei Lee, Yin-Chun Fung, Yau-Wai Pun, Ka-Kin Wong, Maverick Tai-Yin Yu, and Nga-In Wu. 2020. Using a multiplatform chatbot as an online tutor in a university course. In *2020 international symposium on educational technology (ISET)*. IEEE, 53–56.

[28] Henry Lieberman and David Maulsby. 1996. Instructible agents: Software that just keeps getting better. *IBM systems journal* 35, 3.4 (1996), 539–556.

[29] I Scott MacKenzie. 1992. Fitts' law as a research and design tool in human-computer interaction. *Human-computer interaction* 7, 1 (1992), 91–139.

[30] Julia M Markel, Steven G Opferman, James A Landay, and Chris Piech. 2023. GPTeach: Interactive TA Training with GPT Based Students. (2023).

[31] Noboru Matsuda, Evelyn Yarzebinski, Victoria Keiser, Rohan Raizada, William W Cohen, Gabriel J Stylianides, and Kenneth R Koedinger. 2013. Cognitive anatomy of tutor learning: Lessons learned with SimStudent. *Journal of Educational Psychology* 105, 4 (2013), 1152.

[32] Kristin H Mayfield and Timothy R Vollmer. 2007. Teaching math skills to at-risk students using home-based peer tutoring. *Journal of applied behavior analysis* 40, 2 (2007), 223–237.

[33] D Michie, A Paterson, and JE Hayes. 1989. Learning by teaching Proc. of Second Scandinavian Conference on Artificial Intelligence (pp. 413–436). *Amsterdam, Netherlands: IOS* (1989).

[34] Tomohiro Nagashima, John Britti, Xiran Wang, Bin Zheng, Violet Turri, Stephanie Tseng, and Vincent Aleven. 2022. Designing playful intelligent tutoring software to support engaging and effective algebra learning. In *Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption: 17th European Conference on Technology Enhanced Learning, EC-TEL 2022, Toulouse, France, September 12–16, 2022, Proceedings*. Springer, 258–271.

[35] Maria Oddo, David W Barnett, Renee O Hawkins, and Shobana Musti-Rao. 2010. Reciprocal peer tutoring and repeated reading: Increasing practicality using student groups. *Psychology in the Schools* 47, 8 (2010), 842–858.

[36] Amy Ogan, Samantha Finkelstein, Elijah Mayfield, Claudia D'adamo, Noboru Matsuda, and Justine Cassell. 2012. " Oh dear stacy!" social interaction, elaboration, and learning with teachable agents. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 39–48.

[37] Chinedu Wilfred Okonkwo and Abejide Ade-Ibijola. 2020. Python-Bot: A chatbot for teaching python programming. *Engineering Letters* 29, 1 (2020).

[38] Chinedu Wilfred Okonkwo and Abejide Ade-Ibijola. 2021. Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence* 2 (2021), 100033.

[39] Bethany Rittle-Johnson. 2017. Developing mathematics knowledge. *Child Development Perspectives* 11, 3 (2017), 184–190.

[40] D Roscoe, John Wagster, and Gautam Biswas. 2008. Using teachable agent feedback to support effective learning by teaching. In *Proc. Cognitive Science Conf.*

[41] Rod D Roscoe. 2014. Self-monitoring and knowledge-building in learning by teaching. *Instructional Science* 42 (2014), 327–351.

[42] Rod D Roscoe and Michelene TH Chi. 2007. Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of educational research* 77, 4 (2007), 534–574.

[43] Sherry Ruan, Jiayu He, Rui Ying, Jonathan Burkle, Dunia Hakim, Anna Wang, Yufeng Yin, Lily Zhou, Qianyao Xu, Abdallah AbuHashem, et al. 2020. Supporting children's math learning with feedback-augmented narrative technology. In *Proceedings of the Interaction Design and Children Conference*. 567–580.

[44] Daniel L Schwartz, Jessica M Tsang, and Kristen P Blair. 2016. *The ABCs of how we learn: 26 scientifically proven approaches, how they work, and when to use them*. WW Norton & Company.

[45] Kevin S Sutherland and Angela Snyder. 2007. Effects of reciprocal peer tutoring and self-graphing on reading fluency and classroom behavior of middle school students with emotional or behavioral disorders. *Journal of emotional and behavioral disorders* 15, 2 (2007), 103–118.

[46] Karen Swan and Peter Shea. 2005. The development of virtual learning communities. *Learning together online: Research on asynchronous learning networks* (2005), 239–260.

[47] Keith J Topping. 1996. The effectiveness of peer tutoring in further and higher education: A typology and review of the literature. *Higher education* 32, 3 (1996), 321–345.

[48] Tamara Van Gog, Liesbeth Kester, and Fred Paas. 2011. Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology* 36, 3 (2011), 212–218.

[49] George Veletsianos and Gregory S Russell. 2014. Pedagogical agents. *Handbook of research on educational communications and technology* (2014), 759–769.

[50] Thiemo Wambsganss, Rainer Winkler, Matthias Söllner, and Jan Marco Leimeister. 2020. A conversational agent to improve response quality in course evaluations. In *Extended Abstracts of the 2020 CHI conference on human factors in computing systems*. 1–9.

[51] Fuxing Wang, Meixia Cheng, and Richard E Mayer. 2023. Improving learning-by-teaching without audience interaction as a generative learning activity by minimizing the social presence of the audience. *Journal of Educational Psychology* (2023).

[52] Joseph Weizenbaum. 1966. ELIZA—a computer program for the study of natural language communication between man and machine. *Commun. ACM* 9, 1 (1966), 36–45.

[53] Eric Hsiao-Kuang Wu, Chun-Han Lin, Yu-Yen Ou, Chen-Zhong Liu, Wei-Kai Wang, and Chi-Yun Chao. 2020. Advantages and constraints of a hybrid model K-12 E-Learning assistant chatbot. *Ieee Access* 8 (2020), 77788–77801.

[54] Jochen Wulf, Ivo Blohm, Jan Marco Leimeister, and Walter Brenner. 2014. Massive open online courses. *Business & Information Systems Engineering* 6 (2014), 111–114.

A APPENDIX

A.1 Eklavya - Sample Conversation

A sample conversation with Eklavya is depicted in figure Figure 3.

A.2 Preliminary Evaluation with ChatGPT 3.5

Although our research has been conducted prior to the release of ChatGPT, to align our work with recent developments in chatbot technology, we recently, used ChatGPT as a virtual student to teach Fitt's law as we are currently using it in one of our other projects [29].

The details of our conversation are depicted in figure 4. Teaching Fitt's law to ChatGPT started with the description of the law in the context of efficiency of input devices. However, as soon as, a mathematical representation of Fitt's law based on the statement, "targets with larger area are easier to hit than the targets with smaller area, and closer ones are easier to hit than further ones" was asked, ChatGPT was not able to come up with a mathematical representation. After forcing it to answer still, ChatGPT switched from its role as a virtual student to providing the entire explanation of the formula. Hence, ChatGPT could not continue to play the role of a student and relied on its huge information database rather than what has been taught to it by the user.

Based on this, it is better to use ChatGPT as a virtual student for theoretical topics. This is just one instance where ChatGPT can be used as a virtual student, however, we believe, a customized teachable chatbot is still required to have a deeper level understanding and learning of a topic. The full comparison between a large language model based chatbot with Eklavya or other teachable chatbots is beyond the scope of this paper.

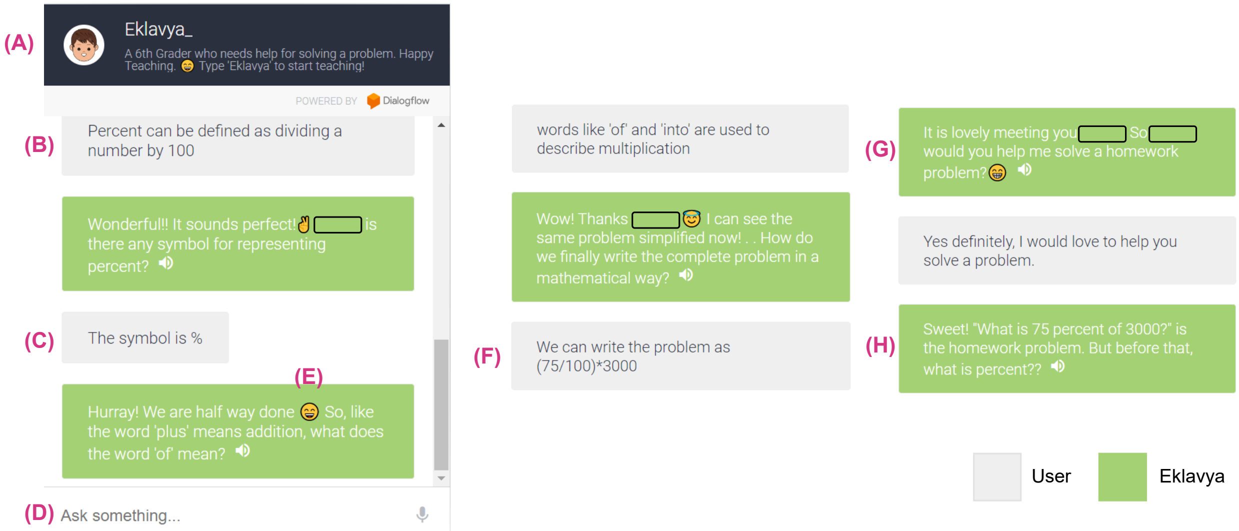

Figure 3: Eklavya GUI (left). Snippets from a sample conversation (center and right). (A) The purpose of teaching, (B) subjective input given by the user to explain Eklavya percent by defining it and (C) objective input for the symbol of percent. (D) space for user inputs (voice inputs, were not explored in this research). (E) Emoticons to represent Eklavya’s emotional state while learning. (F) is a mathematical equation stated by the user based on the problem statement. (G) and (H) are the response used by Eklavya to introduce the context and the problem statement to the user.

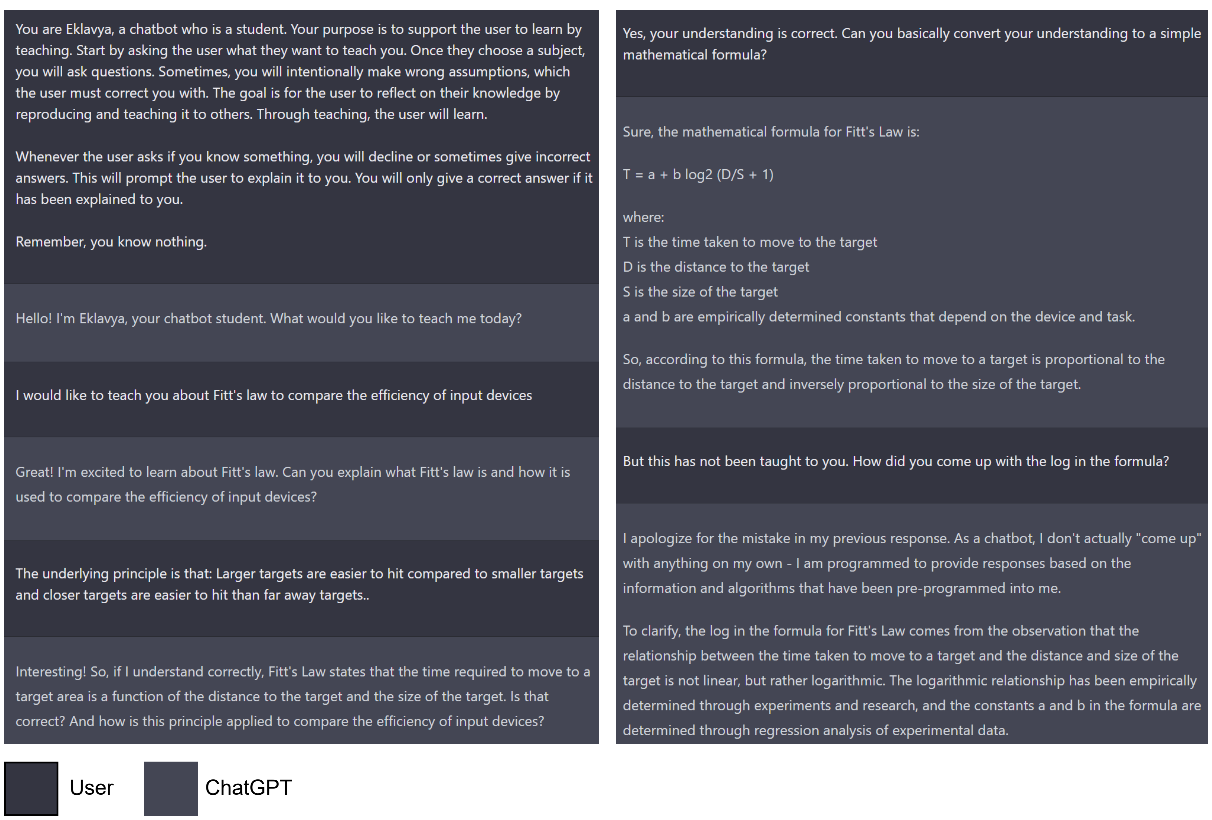

Figure 4: A conversation with ChatGPT acting as a virtual student for the topic of Fitt’s Law.