Majumdar, R. et al. (Eds.) (2025). Proceedings of the 1st International Conference on Learning Evidence and Analytics. Asia-Pacific Society for Computers in Education

The Effect of Feedback in Chatbot-based Pre-class Learning Environment

Vimeanseth THORNG[a*], Fumiya OKUBO[b] & Atsushi SHIMADA[b]
[a]*Graduate School and Faculty of Information Science and Electrical Engineering, Kyushu University, Japan*
[b]*Faculty of Information Science and Electrical Engineering, Kyushu University, Japan*
*thorng.vimeanseth.000@s.kyushu-u.ac.jp

Abstract: With the rapid advancement of large language models (LLMs), chatbots powered by these models have recently been introduced to support pre-class learning. However, a potential concern is that students may develop misunderstandings during their interaction with the chatbot. To address this issue, incorporating exercises and feedback after chatbot-based learning is considered effective. There are two main types of feedback: Knowledge of Correct Response (KCR), and Explanation Feedback (ExF). Each type is believed to have different effects on learning outcomes. However, no research has compared these effects in the context of chatbot-based pre-class learning. Therefore, in this study we developed LLM-based Chatbot and Feedback-Tool to promote early awareness of misunderstandings and deepen understanding. Additionally, we evaluated the effectiveness of the developed learning support tools, as well as the effect of KCR and ExF on learning outcomes. The experimental results demonstrated the effectiveness of the learning support tools in the pre-class learning. Furthermore, it showed no significant difference in learning outcomes between the two types of feedback. However, students perceived ExF as more useful.

Keywords: Feedback types, Pre-class learning, Chatbot, Feedback-Tool, Learning outcomes, Large Language Model

1. Introduction

Prior knowledge at the beginning of a class has been shown to significantly influence students' academic achievement, supporting the view that learning builds on preexisting understanding (Thompson, et al., 2003). In recent years, with the rapid advancement of large language models (LLMs) and their text generation capabilities, there has been growing interest in using LLM-based chatbots to support pre-class learning. By utilizing chatbots in pre-class learning, students can gain a preliminary understanding of the lesson content. However, a major challenge is the possibility of students developing misconceptions during the knowledge acquisition process. One possible solution is to provide practice exercises after chatbot-based learning, accompanied by appropriate feedback on students' responses.

Shute (2008) classified feedback into simple feedback and elaborated feedback (EF) based on the complexity of the content. Simple feedback includes "Knowledge of Results (KR)", which only informs students of correctness, and "Knowledge of Correct Response (KCR)", which provides the correct answer in case of an incorrect response. On the other hand, EF consists of "Explanation Feedback (ExF)", which explains the reasons behind correctness or incorrectness, and "Prompt Feedback", which suggests the next course of action (Cai, et al., 2023).

Previous studies have suggested that KCR is more effective than KR. However, opinions on the effectiveness of elaborated feedback remain inconsistent. Van der Kleij et al. (2015) reported that elaborated feedback is superior to KCR. In contrast, Shute (2008) pointed out that excessive information may hinder cognitive processing. Yang (2017) has also shown that elaborated feedback does not offer a significant advantage over KCR. Considering these

previous studies, it is important to note that there is no consistent conclusion regarding the relationship between the amount of information in feedback and its effectiveness. Moreover, to the best of the author's knowledge, no study has examined the impact of KCR and one form of elaborated feedback, ExF on learning outcomes within a pre-class learning environment powered by LLM-based chatbot and feedback tool.

Therefore, the aim of this study is to promote early awareness of misunderstandings and deepen comprehension during pre-class learning by developing LLM-based chatbot and feedback tool. Furthermore, we aim to investigate the following research questions:

- **RQ1**: How effective are the developed learning support tools (LLM-based Chatbot and Feedback-Tool) in enhancing students' understanding or learning outcomes?
- **RQ2**: Between KCR and ExF, which type of feedback is more effective in enhancing students' learning outcomes in chatbot-based pre-class learning environment?

2. Chatbot-based Pre-class Learning Design

First, the overall flow of the lesson incorporating the chatbot-based pre-class learning phase is described. The lesson consists of two phases. In the first phase as shown in Figure 1, students engage in learning using tools such as Chatbot and Feedback-Tool, allowing them to become familiar with key concepts and identify questions before class. In the second phase, students participate in regular classroom instruction, focusing on those questions and deepening their understanding through the instructor's explanations. Figure 1 illustrates the two pre-class activities: "learning using the Chatbot" and "practicing using the Feedback-Tool". Note: The Chatbot refers to a chatbot developed using the Dify platform and utilizing GPT-4o for content generation. The Feedback-Tool is used for practicing exercises and generates feedback using GPT-4o. It was developed with Flask and MongoDB frameworks.

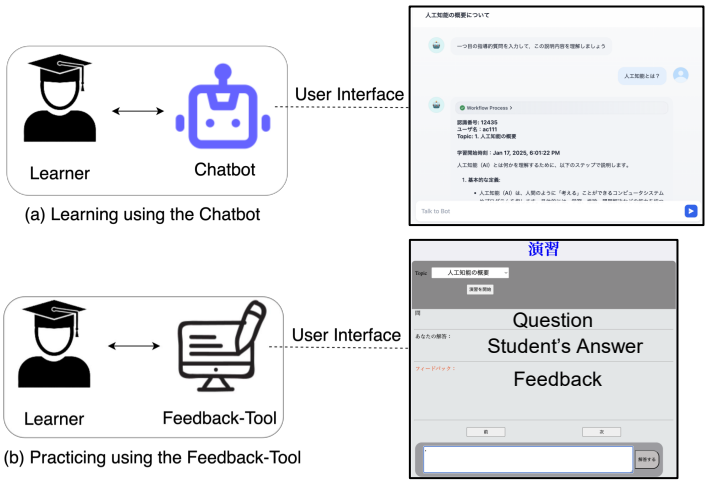

Figure 1. Chatbot-based pre-class learning activities

In the phase of learning using the Chatbot, each student engages in individual study by interacting with the chatbot based on a list of guiding questions (Huesca, et al., 2024). Guiding questions are questions designed by the instructor to stimulate students thinking and self-directed learning. They are designed to get to the point of the content, to aid understanding, and to support the acquisition of specific knowledge and skills. For example, in the case of a class on "Overview of Artificial Intelligence", the guiding questions are shown in Figure 2. In this process, students select guiding questions to ask the chatbot and can pose follow-up questions as needed. This allows for immediate, individualized support, helping students resolve doubts early and deepen their understandings.

In the phase of practicing using the Feedback-Tool, students engage with short-answer questions and write their responses. This tool will utilize GPT-4o to analyze the responses and provide feedback that includes the reasons for correctness or incorrectness, thereby deepening students' understanding and facilitating early awareness of misunderstandings.

1. What is Artificial Intelligence (AI)? What are Strong AI and Weak AI? What are the applications of AI?
2. What is the history of AI research (from the era of Inference-and-Search to the era of Generative-AI)?
3. What is the Frame Problem? What is the Adversarial Example? What is Explainable AI?
4. What are the challenges of AI?

Figure 2. List of Guiding Questions

3. Experiment and Result

3.1 Experimental Setup

The 102 participants were recruited from among undergraduate and graduate students at Kyushu University. The participants were evenly divided into two groups of 51 each. The experiment was conducted over a two-week period, from December 17 to December 30, 2024. As part of the experimental procedure, students engaged in two online pre-class learning sessions using the Chatbot and Feedback-Tool: one on “Overview of Artificial Intelligence” and another on “Overview of Machine Learning”. Group 1 received KCR in the first session and ExF in the second, while Group 2 received the feedback in the reverse order. To examine the RQ1 and RQ2 mentioned in section 1, Pre- and Post-test were administered for each session. Additionally, a 5-point Likert scale questionnaire was conducted after each Post-test.

To evaluate learning effectiveness, normalized learning gain (g_i) is widely used. However, this metric may not adequately reflect growth in low-scoring students, as it yields the same value (1.0) when both low- and high-scoring students reach full marks. To address this, we modified the metric by squaring the numerator, emphasizing improvement among low scorers. We also adjusted it to capture negative gains when post-test scores declined. This modified learning gain (g'_i) was used to analyze the effect of feedback types on learning outcomes.

$$g_i = \frac{\text{Post\_score}_i - \text{Pre\_score}_i}{\text{Max\_score} - \text{Pre\_score}_i} \tag{1}$$

$$g'_i = \begin{cases} \dfrac{(\text{Post\_score}_i - \text{Pre\_score}_i)^2}{\text{Max\_score} - \text{Pre\_score}_i}, & \text{Post\_score}_i \geq \text{Pre\_score}_i \\ -\dfrac{(\text{Post\_score}_i - \text{Pre\_score}_i)^2}{\text{Max\_score} - \text{Pre\_score}_i}, & \text{Post\_score}_i < \text{Pre\_score}_i \end{cases} \tag{2}$$

3.2 Experimental Results

3.2.1 Pre-post Test

Students who scored full marks (10 points) on the pre-test were excluded, as the modified normalized gain (g'_i) could not be calculated for them. As a result, Group 1 and 2 consisted of 41 and 40 students in the first session, and 45 and 38 in the second session, respectively.

RQ1: This study analyzed pre- and post-test results using the Wilcoxon signed-rank test. As shown in Table 1, in both sessions, Groups 1 and 2 showed significantly improved post-test scores (p=0.000), suggesting the effectiveness of the Chatbot and Feedback-Tool as learning support tools in the chatbot-based pre-class learning environment.

Table 1. Statistical Values of Pre- and Post-test Scores

Value	First session				Second session			
	Group 1		Group 2		Group 1		Group 2	
	pre	post	pre	post	pre	post	pre	post
N	41	41	40	40	45	45	38	38
Mean (±SD)	6.37 (1.46)	8.88 (0.89)	6.78 (1.13)	9.08 (0.98)	5.8 (1.98)	8.67 (1.51)	6.16 (1.37)	9.8 (1.02)
p-value	**0.000****		**0.000****		**0.000****		**0.000****	

N: Number of students; SD: Standard Deviation; pre: pre-test scores; post: post-test scores; **p < 0.01

RQ2: To address this research question, the Mann–Whitney U test was conducted to compare the impact of KCR and ExF on students' learning outcomes. In the first session, Group 1 (KCR) had a mean modified learning gain of 1.97 (Standard Deviation [SD] = 1.47), while Group 2 (ExF) had a mean of 1.96 (SD = 1.54). The resulting p-value was 0.887, indicating no significant difference between the groups. In the second session, the mean for Group 1 was 2.14 (SD = 1.91), and for Group 2, it was 2.47 (SD = 1.56), with a p-value of 0.191, again showing no statistically significant difference. These results suggest that, under the conditions of this study, KCR and ExF feedback types do not lead to a significant difference in students' learning outcomes.

3.2.2 Questionnaire

RQ1: As shown in Table 2, the number of respondents was *N*=87. For item (a), over 90% of students reported understanding the chatbot's explanations. For item (b), about 70% asked additional questions to the Chatbot, indicating its effective use in individual learning. For item (c) and (d), more than 80% of students reported that feedback helped them notice misunderstandings and improve understanding. These results demonstrate the effectiveness of the developed Chatbot and Feedback-Tool in supporting individual pre-class learning.

Table 2. Effectiveness of Chatbot and Feedback-Tool Questionnaire (5-point Likert Scale Ranging from 5: Strongly Agree to 1: Strong Disagree)

Items	Session	Number of students				
		5	4	3	2	1
(a) I understood the explanation provided by the Chatbot	1st	48	36	3	0	0
	2nd	29	52	2	4	0
(b) I asked additional questions to the Chatbot (e.g., about doubts)	1st	39	28	15	3	2
	2nd	34	30	17	3	3
(c) Feedback made you aware of misunderstandings or lack of understanding	1st	41	34	8	2	2
	2nd	52	29	6	0	0
(d) Feedback contributed to improving your understanding	1st	46	29	8	3	1
	2nd	47	30	6	4	0

Table 3. Feedback Effect Questionnaire (5-point Likert Scale Ranging from 5: Strongly Agree to 1: Strongly Disagree)

Session	Items	Group	Percentage of students [%]				
			5	4	3	2	1
1st	Q1	1 (KCR)	39.1	41.3	13.0	4.3	2.2
		2 (ExF)	56.1	36.6	4.9	0	2.4
	Q2	1 (KCR)	39.1	41.3	13.0	4.3	2.2
		2 (ExF)	68.3	24.4	4.9	2.4	0
2nd	Q1	1 (ExF)	63.0	32.6	4.3	0	0
		2 (KCR)	56.1	34.1	9.8	0	0
	Q2	1 (ExF)	58.7	32.6	6.5	2.2	0
		2 (KCR)	48.8	36.6	7.3	7.3	0

KCR: Knowledge of Correct Response; ExF: Explanation Feedback; Q1: Feedback made you aware of misunderstandings or lack of understanding; Q2: Feedback contributed to improving your understanding

RQ2: This RQ was examined by investigating students' perceptions of the feedback effects. Using a 5-point Likert scale, responses were gathered for two items: "Q1: Feedback made you aware of misunderstandings or lack of understanding" and "Q2: Feedback contributed to improving your understanding". As shown in Table 3, in the first session, a higher percentage of students in Group 2 (who received ExF) responded "strongly agree" or "agree" to questions Q1 and Q2 compared to Group 1 (KCR), with a maximum difference of

approximately 29%. In the second session, even after switching the feedback types, Group 1 (who received ExF) again rated the feedback more positively than Group 2, with a maximum difference of approximately 6%. These results suggest that students find ExF as more useful.

4. Discussion and Conclusion

This study proposed a new pre-class learning method using LLM-based Chatbot and Feedback-Tool. It also evaluated the effectiveness of the learning support tools, as well as the effect of KCR and ExF on learning outcomes. Drawing from the results of the pre- and post-test as well as the questionnaire, two main findings emerged.

For RQ1, the findings indicated that integrating the Chatbot and Feedback-Tool into pre-class learning effectively supports individualized learning. The significant improvements in post-test scores suggest that students were able to deepen their understanding through the use of these learning support tools. High engagement—evident in students' comprehension and active follow-up questioning—highlights the Chatbot's potential for immediate, personalized support. This approach promotes autonomous learning and may foster metacognitive skills, as students reported feedback helped clarify misunderstandings.

For RQ2, although test scores revealed no significant difference in learning outcomes between KCR and ExF, the questionnaire results consistently indicated a student preference for ExF. This suggests that students value feedback that not only provides the correct answer but also explains the reasoning behind it. Notably, the positive perception of ExF persisted even after the feedback types were switched in the second session. These findings imply that while KCR may be sufficient for surface-level correction, ExF facilitates deeper understanding and reflection, which are especially important in pre-class learning environments. One possible explanation for this discrepancy is that ExF enhances learners' perceived understanding and confidence by providing more meaningful insights, which may not be immediately reflected in test scores. Moreover, its explanatory nature is likely to foster long-term retention and conceptual clarity that short-term assessments may fail to capture.

A key limitation of this study is the narrow scope of its dataset, which focused solely on university students and content related to Artificial Intelligence. As such, the generalizability of the findings to other educational levels (e.g., K–12) or subject areas (e.g., humanities) remains unclear. Additionally, the study did not examine the long-term impact of the intervention. Future research should examine its impact across diverse contexts and over extended periods.

Acknowledgements

This work was supported by JST CREST Grant Number JPMJCR22D1 and JSPS KAKENHI Grant Number JP22H00551, Japan.

References

Thompson, R., & Zamboanga, B. (2003). Prior Knowledge and Its Relevance to Student Achievement in Introduction to Psychology. Teaching of Psychology. 30. 96-101.

Shute, V. J. (2008). Focus on Formative Feedback. Review of Educational Research. 78. 153-189.

Cai, Z., Gui, Y., Mao, P., et al. (2023). The effect of feedback on academic achievement in technology-rich learning environments (TREs): A meta-analytic review. Educational Research Review, Vol 39.

Van der Kleij, F. M., Feskens, R. C. W., & Eggen, T. J. H. M. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. Review of Educational Research, 85(4), 475–511. https://doi.org/10.3102/0034654314564881

Yang, K. (2017). Learning behavior and achievement analysis of a digital game-based learning approach integrating mastery learning theory and different feedback models. Interactive Learning Environments, 25, 235 - 248.

Huesca, G., Martínez-Treviño, Y., et al. (2024). Effectiveness of Using ChatGPT as a Tool to Strengthen Benefits of the Flipped Learning Strategy. Education Sciences, 14(6), 660.