Towards AI-Enhanced Classroom Response System Eliciting Self-Explanations in Computer Science Courses

Ozogul, G., Zheng, X., Min, W., Lee, S., Carpenter, D., & Lester, J.

Abstract: This study investigates the implementation of a classroom response system in STEM education in a higher education context. The study used ExplainIt, a web-based classroom response system designed to support students' self-explanations and provide instant feedback. Data were collected from 32 undergraduate students using four instruments including demographic information, self-efficacy, engagement, and system evaluation. The results showed that students reported positive learning experiences, demonstrated increased self-efficacy in STEM content, and indicated high levels of engagement following their use of ExplainIt.

Introduction

The use of Artificial Intelligence (AI) has become very prominent in higher education contexts in the last five years (Chu et al., 2022), particularly in enabling faculty members to pose questions and provide just-in-time feedback to students in large courses (Dever et al., 2020). This has been a game-changing feature of AI in higher education, as higher education faculty often faced barriers to implementing feedback (Cragg, 2024) and received criticism for the inadequacies of feedback provided

to students in large classrooms (Boud & Molloy, 2013). Specifically, there has been a rise in publications related to automated feedback systems, particularly within STEM in higher education (Deeva et al., 2021). Automated intelligent tutoring systems can provide adaptive feedback to assist students during problem solving (Kochmar et al., 2022) and offer guidance on practice items (Yilmaz et al., 2022).

Feedback in higher education has shown positive effects on learners' self-efficacy (Ozogul et al., 2008), motivation (Dawson et al., 2019), learning outcomes (Ozogul & Sullivan, 2009; Schunk & Ertmer, 2000), and self-regulation (Zimmerman, 2000). In recent years, technology has become a support tool for instructors while providing feedback in large classes, such as clickers (Hunsu et al., 2016) and mobile app response systems (Teo & Chew, 2015). These solutions for feedback fall short in developing students' deep conceptual understanding and may not push students to achieve high-level cognitive outcomes in STEM courses (Shapiro et al., 2017). This limitation stems from these solutions' reliance on multiple-choice questions, which do not allow students to construct their own responses and engage in interactive learning (Chi & Wylie, 2014). Emerging AI technologies may present solutions to reduce the issue of reliance on multiple-choice questions for feedback and address the absence of instant feedback in STEM classes (Roll & Wylie, 2016).

One way to improve student learning and engagement is by developing technology that allows students to write their self-explanations and receive feedback. Self-explanation is a cognitive process that learners can actively engage into processing new information, acquiring skills by associating it with prior knowledge, and articulating responses in their own words to express conceptual knowledge (Bisra et al., 2018; Fonseca & Chi, 2011). Self-explanations coupled with personalized and specific feedback are crucial to maximize the effectiveness of this pair (Nakamoto et al., 2023). Self-explanations have been found to be effective for student learning in STEM domains, especially in Computer Science (Sudol-DeLyser, 2015).

Context of Study

ExplainIt is a web-based classroom response system designed to support students' self-explanations and provide feedback. Specifically, after creating and posing a question (e.g., "why" or "how" questions) using the Instructor App, students view the question and submit their self-explanation responses by using the Student App. With these functionalities in place, ExplainIt is undergoing further development to support automatic assessment of students' self-explanation responses and provide personalized feedback to students with large language model (LLM)-based natural language processing (NLP) techniques. Preliminary findings for automatically assessing student self-explanations show promising results by achieving high predictive accuracy, overcoming the limitations of manual evaluation in large classes (Carpenter et al., 2024). This highlights the potential of LLMs to efficiently and effectively assess student explanations, toward implementing an AI-enabled feedback system. In this initial implementation of ExplainIt, we did not utilize the LLM/NLP function but focused on the system and student experiences.

Research Questions

1. Are there any changes in Computer Science (CS) students' self-efficacy before and after using the ExplainIt classroom response system?
2. How do CS students perceive their engagement when using the ExplainIt system?
3. How do CS students perceive the usability of the ExplainIt system?

Method

The case study took place at a public university in the southeastern United States. The researchers investigated the effects of ExplainIt on students' self-efficacy, engagement, and usability perceptions.

Participants

Participants were 32 undergraduate students in a face-to-face Computer Science course. Participants were between 18 and 28 years old. There were 8 (25%) females, 23 (72%) males, and 1 (3%) did not report gender. In terms of ethnicity, 16 (50%) were white, 13 (41%) were Asian, and 3 (9%) were other.

Data Collection and Analysis

Four instruments were used to collect data: demographic, self-efficacy, user engagement, and system evaluation. The demographic and self-efficacy survey instruments were administered before the ExplainIt implementation. After the course, the self-efficacy, user engagement, and system evaluation instruments were given. Once all the instruments were completed, researchers downloaded and analyzed the data.

Results

The completion rates for the instruments varied. For self-efficacy, 16 students completed both pre and post surveys. For self-efficacy, the mean score was 3.69 in the pre-survey and 3.80 in the post-survey. The question with the highest gain on self-efficacy was: “I am knowledgeable about the concepts covered in the course,” which increased from 3.23 to 3.89. However, there were no significant differences in overall self-efficacy levels between pre- and post-surveys.

Regarding engagement, 18 students completed the post-survey. The average agreement rating for engagement was 3.48. Particularly, three statements received the highest ratings: “ExplainIt was not confusing to use” (4.28); “ExplainIt was not stressful” (4.28); and “Using ExplainIt to learn was an interesting experience” (4.11).

For system evaluation, 17 students completed it. Students rated the system’s ease of use at 3.94 and the ease of using the system to complete tasks at 3.47, where “1” was coded as “very hard” and “5” as “very easy”. The ExplainIt system evaluation yielded an average agreement of 3.85. The highest-rated statements were “time provided to write answers through ExplainIt was generally enough” (4.35) and “ExplainIt system’s complexity was appropriate” (4.29). Students also rated “the system is easy to learn quickly” very high (4.18).

The thematic analysis of the open-ended questions on usability uncovered several themes. Most students found ExplainIt helpful for learning and understanding the CS course concepts. Students also highlighted the simplicity and usability of the system. In terms of disliked aspects, a few students mentioned software bugs, such as a timeout issue. Students suggested additional features for ExplainIt, such as the ability to display a history log of submitted answers, results with graphs, differentiate between graded and non-graded questions, and an overall summary of answered questions.

Discussion

In terms of students’ self-efficacy, there was a positive change between pre- and post-instruments. This may be due to the self-explanation prompts and responses that helped students learn CS concepts and improve their self-efficacy perceptions. However, since there were no statistically significant differences in the gain scores, these positive gains were minimal. The observed gains may be attributed to the use of the ExplainIt self-explanation system or simply to students learning the course content and thus building confidence over time. Further research is needed to isolate the effects of the ExplainIt system through an experimental study to determine its specific impact on self-efficacy.

Students evaluated the experience with their system positively and found it easy to use and engaging. As higher education faculty often faced barriers in implementing formative feedback (Cragg, 2024), students may have appreciated the newer instructional approach and self-explanation support in their CS courses. This finding is promising, even though students in this study also reported usability issues, such as system timeouts and software bugs. Despite these issues, students remained engaged with the system and rated the system as easy to use. This may be due to the open nature of interaction in CS courses or the fact that CS majors tend to be more lenient toward software bugs and technical issues. Future studies should investigate a more robust version of the system, free of technical issues, and include students from other STEM disciplines.

To effectively integrate ExplainIt into diverse educational settings, instructors should receive training on best practices for designing and implementing self-explanation prompts. Workshops and professional development sessions can help educators craft questions that elicit meaningful self-explanations and support students in interpreting AI-generated feedback effectively. Additionally, ensuring that AI-generated feedback remains equitable and free from biases will be crucial. Finally, large-scale deployment may require integration with various learning management systems and robust server capacity. Addressing these challenges will be essential to enhancing ExplainIt's impact across diverse learning environments.

References

Bisra, K., Liu, Q., Nesbit, J. C., Salimi, F., & Winne, P. H. (2018). Inducing self-explanation: A meta-analysis. *Educational Psychology Review, 30*, 703-725. https://doi.org/10.1007/s10648-018-9434-x

Boud, D., & Molloy, E. (2012). Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in Higher Education, 38*(6), 698–712. https://doi.org/10.1080/02602938.2012.691462

Carpenter, D., Min, W., Lee, S., Ozogul, G., Zheng, X., & Lester, J. (2024). Assessing student explanations with large language models using fine-tuning and few-shot learning. In *Proceedings of the 19th Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 403-413). https://aclanthology.org/2024.bea-1.33/

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219-243.https://doi.org/10.1080/00461520.2014.965823

Chu, H., Tu, Y., & Yang, K. (2022). Roles and research trends of artificial intelligence in higher education: A systematic review of the top 50 most-cited articles. *Australasian Journal of Educational Technology, 38*(3), 22–42. https://doi.org/10.14742/ajet.7526

Cragg, J. (2021). *Digital formative assessments in higher education: Faculty recommendations for overcoming barriers to effective implementation* (Publication No. 28645307) [Doctoral dissertation, Northeastern University]. ProQuest Dissertations Publishing.

Dawson, P., Henderson, M., Mahoney, P., Phillips, M., Ryan, T., Boud, D., & Molloy, E. (2019). What makes for effective feedback: Staff and student perspectives. *Assessment & Evaluation in Higher Education, 44*(1), 25-36. https://doi.org/10.1080/02602938.2018.1467877

Deeva, G., Bogdanova, D., Serral, E., Snoeck, M., & De Weerdt, J. (2021). A review of automated feedback systems for learners: Classification framework, challenges and opportunities. *Computers & Education, 162*, 104094. https://doi.org/10.1016/j.compedu.2020.104094

Dever, D. A., Azevedo, R., Cloude, E. B., & Wiedbusch, M. (2020). The impact of autonomy and types of informational text presentations in game-based environments on learning: Converging multi-channel processes data and learning outcomes. *International Journal of Artificial Intelligence in Education, 30*(4), 581–615. https://doi.org/10.1007/s40593-020-00215-1

Fonseca, B. A., & Chi, M. T. (2011). Instruction based on self-explanation. In *Handbook of research on learning and instruction* (pp. 310-335). Routledge.

Hunsu, N. J., Adesope, O., & Bayly, D. J. (2016). A meta-analysis of the effects of audience response systems (clicker-based technologies) on cognition and affect. *Computers & Education, 94*, 102-119. https://doi.org/10.1016/j.compedu.2015.11.013

Kochmar, E., Vu, D. D., Belfer, R., Gupta, V., Serban, I. V., & Pineau, J. (2022). Automated data-driven generation of personalized pedagogical interventions in intelligent tutoring systems. *International Journal of Artificial Intelligence in Education, 32*(2), 323-349.https://doi.org/10.1007/s40593-021-00267-x

Nakamoto, R., Flanagan, B., Dai, Y., Yamauchi, T., Takami, K., & Ogata, H. (2023). Enhancing Self-Explanation Learning through a Real-Time Feedback System: An Empirical Evaluation Study. *Sustainability, 15*(21), 15577. https://doi.org/10.3390/su152115577

Ozogul, G., Olina, Z., & Sullivan, H. (2008). Teacher, self and peer evaluation of lesson plans written by preservice teachers. *Educational Technology Research and Development, 56*, 181-201. https://doi.org/10.1007/s11423-006-9012-7

Ozogul, G., & Sullivan, H. (2009). Student performance and attitudes under formative evaluation by teacher, self and peer evaluators. *Educational Technology Research and Development, 57*, 393-410. https://doi.org/10.1007/s11423-007-9052-7

Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education, 26*, 582-599.https://doi.org/10.1007/s40593-016-0110-3

Schunk, D. H., & Ertmer, P. A. (2000). Self-regulation and academic learning: Self-efficacy enhancing interventions. In *Handbook of self-regulation* (pp. 631-649). Academic Press.

Shapiro, A. M., Sims-Knight, J., O'Rielly, G. V., Capaldo, P., Pedlow, T., Gordon, L., & Monteiro, K. (2017). Clickers can promote fact retention but impede conceptual understanding: The effect of the interaction between clicker use and pedagogy on learning. *Computers & Education, 111*, 44-59. https://doi.org/10.1016/j.compedu.2017.03.017

Sudol-DeLyser, L. A. (2015). Expression of abstraction: Self explanation in code production. In *Proceedings of the 46th ACM technical symposium on computer science education* (pp. 272-277). https://doi.org/10.1145/2676723.2677222

Teo, Y. Z. & Chew, E. (2015) A Mobile Personal Response for Assessment and Feedback in Computing Education. In Teaching and Learning with Technology: Proceedings of the *2015 Global Conference on Teaching and Learning with Technology*(pp. 47-57). https://doi.org/10.1142/9789814733595_0004

Yilmaz, R., Yurdugül, H., Yilmaz, F. G. K., Şahin, M., Sulak, S., Aydin, F., ... & Ömer, O. R. A. L. (2022). Smart MOOC integrated with intelligent tutoring: A system architecture and framework model proposal. *Computers and Education: Artificial Intelligence, 3*, 100092.https://doi.org/10.1016/j.caeai.2022.100092

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In *Handbook of self-regulation* (pp. 13-39). Academic press. https://doi.org/10.1016/B978-012109890-2/50031-7

Acknowledgments

This research was supported by funding from the National Science Foundation under Grants DUE-2111473 and DUE-2111216. Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.