EUROPEAN JOURNAL OF STEM EDUCATION

European Journal of STEM Education,
2026, 11(1), 40
eISSN: 2468-4368

lectito

Research paper

Student Experiences with Static versus Adaptive AI Feedback on Self-Explanations in Computer Science Courses

Gamze Ozogul [1*], Xiaoying Zheng [1], Wookhee Min [2], Seung Lee [2], Jordan Esiason [2], Dan Carpenter [3], James C. Lester [2]

[1] *Indiana University*, UNITED STATES
[2] *North Carolina State University,* UNITED STATES
[3] *Amplify AI*, UNITED STATES

*Corresponding Author: gozogul@iu.edu
This paper is part of the special issue: AI and STEM: Exploring the Impact of Artificial Intelligence on Science, Technology, Engineering, and Mathematics Education

Citation: Ozogul, G., Zheng, X., Min, W., Lee, S., Esiason, J., Carpenter, D., & Lester, J. (2026). Student experiences with static versus adaptive AI feedback on self-explanations in computer science courses. *European Journal of STEM Education, 11*(1), Article 40. https://doi.org/10.20897/ejsteme/18919

Published: July 7, 2026

ABSTRACT

Feedback is a critical component of formative assessment yet remains underutilized in higher education STEM courses due to large class sizes and instructor workload. Emerging artificial intelligence (AI) technologies are reshaping STEM education by enabling scalable, personalized feedback and promoting active learning. This study explores the instructional potential of AI-supported feedback in addressing these challenges, with a particular focus on self-explanations. It reports on two iterations of a classroom response system in an undergraduate computer science (CS) course across two semesters. The first iteration provided static feedback based on instructor-prepared sample responses, while the second incorporated AI-generated adaptive feedback powered by natural language processing. Using a comparative case study design, the authors investigated differences in student self-efficacy, engagement, and perceptions of system usability. Both implementations sustained student engagement and participation in CS courses, demonstrating the system's effectiveness in fostering active learning. Findings highlight the potential of AI-driven feedback to enhance scalability and interactivity, deepen conceptual understanding, and reduce instructor workload. These results underscore the promise of AI-driven feedback systems for enhancing interactivity and learning in STEM courses, particularly in CS contexts with growing enrollments. Future work should explore question frequency, question formats, instructor's perspectives and pedagogy with feedback tools, and improve AI accuracy to reduce over-reliance.

Keywords: student engagement, computer science, artificial intelligence, feedback, STEM in higher education

Science, Technology, Engineering and Math (STEM) fields in higher education are being transformed by artificial intelligence (AI) technologies. These technologies are often being used to support instructors and students in curriculum design, monitoring engagement, providing and receiving customized feedback, and conducting just-in-time assessment of learning outcomes (Joseph & Uzondu, 2024; Seco et al., 2025). As AI technologies advance in their capacity to provide automated and customized feedback, higher education institutions are increasingly leveraging AI-supported feedback mechanisms within their learning environments (Acar et al., 2025; Chen, 2025; Slimi et al., 2025).

INTERACTIVITY AND FEEDBACK IN STEM COURSES

In higher education settings, student engagement and interaction within courses play a crucial role in promoting learning and engagement (Beauchamp & Kennewell, 2010; González-Cacho & Abbas, 2022). One way to increase interactivity and productive engagement in courses is embedding formative feedback mechanisms which offer timely insights on student progress. Feedback provides students with information about their performance, helps them interpret their performance results, and enables them to identify areas of improvement, thereby supporting their learning (Henderson et al., 2019a, 2019b). By making students aware of how their current performance compares to the expected performance, feedback creates an opportunity for students to address the "gaps" in their learning, which is an essential premise of formative assessment in education (Hattie & Timperley, 2007). Despite its recognized effectiveness in supporting instruction, student learning processes, and performance in higher education (Baartman & Quinlan, 2024; Pitt & Quinlan, 2022; Shute, 2008), feedback remains underutilized. This is largely due to limited instructor time, large class sizes, and the heavy workload it imposes on instructors (Burner et al., 2025; Henderson et al., 2019a, 2019b; Mirhosseini et al., 2023).

This underutilization is particularly common in higher education STEM courses, which are often dominated by lecture-based teaching methods that lack meaningful interactivity and formative feedback (Gonsar et al., 2021; Ortiz Moreno et al., 2025). The lack of interactivity and the dominance of lectures in STEM courses pose serious issues in undergraduate education. Prior studies have shown that when STEM faculty implement active learning strategies in college courses, student achievement improves across all STEM disciplines (Freeman et al., 2014; Villegas-Ch et al., 2025), and these active learning strategies contribute to narrowing achievement gaps for underrepresented STEM students, helping to create a more equitable learning environment (Ortiz Moreno et al., 2025; Theobald et al., 2020). Similarly, when feedback is used in STEM courses, results show positive outcomes such as increased student motivation and engagement in course content (Park et al., 2023), improved attendance and comprehension (Zhao, 2023), improved interest in course topic and higher course grades (Park et al., 2024), and stronger intentions to persist in the course content area (Marwan et al., 2020).

In line with these findings, a growing number of technological solutions have been developed to enhance interactivity and facilitate timely, personalized feedback that addresses the long-standing challenges in instructional scalability, instructor time, learner engagement, and learner support. For example, classroom response systems which allow instructors to pose multiple-choice or short answer questions in real time, ranging from clickers (Blasco-Arcas et al., 2013; Jammeh et al., 2025; Mayer et al., 2009), to audience response systems (Diaz et al., 2024; Hunsu et al., 2016). Research has demonstrated that such systems can foster active participation and support the acquisition of factual knowledge (Campbell & Mayer, 2009; Hunsu et al., 2016) but also may limit the development of a deep conceptual understanding of underlying concepts and principles, as these technologies can momentarily disrupt the cognitive processes (Shapiro et al., 2017). This suggests that educational technologies do not inherently foster learning outcomes, rather their success relies on intentional pedagogical scaffolding.

Rationale for self-explanations in CS

Similarly, AI tools by themselves are inadequate for promoting deep learning and reflection. Their effectiveness depends on well-designed well-structured pedagogical practices and deliberate choices of instructors (Adamakis & Rachiotis, 2025). To address this limitation, we introduced an explanation-based classroom response system that fosters students' metacognitive processes, informs instructor about students' understanding in the course concepts, and provides time and space for instructor for reflection and instant instructional pivot. The system, ExplainIt, is an AI-based classroom response system that supports students' self-explanation-based responses to promote deeper learning while leveraging AI for real-time customized feedback generation (Carpenter et al., 2024; Ozogul et. al., 2025). Self-explanation is a cognitive process through which learners develop inferences regarding conceptual relationships or causal connections within the learning material they have studied by articulating concepts in their own words (Bisra et al., 2018; Elme et al., 2022). As self-explanation promotes meaning making and fosters building cohesive mental models, it is promising for STEM higher education and for preparing students for today's job market (Jelks & Crain, 2020). We intentionally incorporated self-explanations into the design of this tool for CS higher education courses to provide students with opportunities to interact with the digital tool while helping them develop critical and deep-thinking skills. The system immediately summarizes students' self-explanation responses, enabling instructors to adjust the CS curriculum and use this technology-rich tool to provide just-in-time feedback to students, consistent with the 2019 UNESCO Beijing Consensus on Artificial Intelligence and Education on leveraging AI to support just-in-time feedback in education. By integrating AI-supported tools that provide feedback on students' written self-explanations, these systems may relieve instructors of this timeconsuming task and allow them to focus on higher-order discussions during class time, based on AI-summarized student responses (Fırat, 2023).

Self-explanation has been recognized as a powerful strategy to foster active participation in learning, help construct cohesive mental models (Roy & Chi, 2005), and enhance problem-solving abilities (Chi et al., 1989). Encouraging students to generate self-explanations can provide significant learning benefits in undergraduate STEM classrooms. Research consistently shows that students who engage in self-explanation learn far more effectively than those who do not (Chen et al., 2025; Fonseca & Chi, 2011). Studies investigated the self-explanation effect in computer science (CS) higher education courses consistently reported positive impact on student learning and engagement (Sandoval-Medina et al., 2024; Tamang et al., 2020; Vihavainen et al., 2015). Effectively translating these benefits of self-explanation in CS courses requires tools that facilitate real-time interaction, promptly assess students' responses, and generate feedback to guide cognitive processes. Without such tools, instructors would struggle to evaluate student self-explanation responses instantly and efficiently during class time (Nakamoto et al., 2023, 2024).

Given the cognitive benefits and pedagogical potential of self-explanations coupled with feedback in STEM courses, particularly in CS, it is essential to explore how AI can support their implementation at scale. In this regard, AI-supported feedback offers a compelling solution for enhancing undergraduate STEM education, especially in courses often criticized for their lack of interactivity, a factor closely linked to lower student performance (Freeman et al., 2014). By enabling scalable and customized solutions, AI technologies may help address this challenge, especially in disciplines that involve complex problem solving, such as CS (Adekola, 2025; Muangprathub et al., 2020; Roll & Wylie, 2016). As long as instructor oversight and transparency are preserved when using these tools, they can help instructors deliver customized formative feedback in a timely manner and support instructors administratively (Zacharis & Papadakis 2025), which are central to constructive learning activities (Chiu & Chi, 2014). In a study by López-Pernas et al. (2025), authors found that when CS students used AI in course work, they relied on instant solutions provided by AI and skipped higher metacognitive processes such as self-explanations or reflections when solving CS problems. Similar findings were reported by Prasad & Sane (2024), indicated that integration of AI technologies without self-explanation or reflection processes embedded, may not sufficiently support the development of problem-solving skills among novice CS students.

These studies underscore the need for intentionally designed AI-supported learning tools that scaffold metacognitive engagement rather than to replace it, particularly in CS undergraduate education, where rising enrollments make it increasingly difficult for instructors to provide timely, personalized feedback to every student (Tang et al., 2024). In a study by Marwan et al. (2020), novice CS students who received adaptive feedback during programming courses reported higher engagement, improved learning outcomes, and greater intentions to persist in the field. Similar results were reported by Villegas-Ch et al. (2025) for CS students in logic and sequential problem solving. Students who received adaptive and targeted feedback in an active course setting significantly outperformed the ones who were taught with a traditional approach. These findings emphasize the potential of AI-driven feedback systems enhance academic performance and persistence in CS.

Building on these findings, it is important to consider how students perceive and access AI-generated feedback. Recent studies reveal diverse student perceptions toward AI-generated feedback (Chan & Hu, 2023) and issues related to its accessibility (Khairuddin et al., 2024). Student perceptions become crucial given the widespread integration of AI technologies in academic activities (Hooda et al., 2022). For example, a comparative study between instructor-generated and AI-generated feedback showed that students consistently rated instructor feedback as significantly more useful (Er et al., 2025). Therefore, understanding these perceptions is essential for designing feedback systems support student learning.

Additionally, as AI integration in higher education is moving rapidly, it is important to consider instructors understanding of pedagogical implications of integrating AI tools to their courses (Uğraş et. al. 2024) and bring in the contextualized nuanced understanding that instructors provide in a hybrid approach while integrating AI (Kaliisa et al., 2026). In the case of AI-generated feedback it is important that instructors be in the loop and engaged as students interact with these technology-rich environments.

This study keeps the instructor in the loop and does not impose any requirements regarding how often or in what ways the feedback is used on students' self-explanations in the context of the instructor's CS courses. Instead, it examines how these two feedback versions are used in a natural setting, contribute to CS students' self-efficacy, perceptions of the feedback delivery system's usability, and engagement with the feedback. As emphasized by Uğraş et al. (2024) and Shishakly (2025) students should move beyond passively receiving information or answers from AI and instead engage in critical reflection and evaluation. Accordingly, the self-explanations incorporated into this study's design support critical and deep learning by encouraging students to actively interact with digital technology while receiving feedback.

The following four research questions were posed:

1. How did the instructor and students use static feedback and adaptive AI feedback in the two CS courses?

2. How does student self-efficacy in CS courses differ between static feedback and adaptive AI-generated feedback?
3. How do students' perceptions of classroom response system usability compare between the static feedback and adaptive AI-generated feedback versions in CS courses?
4. How do students' perceptions of engagement differ between static and adaptive AI-generated feedback in CS courses?

CONTEXT

The CS course

The two feedback type iterations via a classroom response system were implemented across two semesters in a 3-credit undergraduate CS course titled "Automated Learning and Data Analysis." This course introduces key concepts and computational methods for extracting insights from data. The course had ten objectives, including "List and explain the problems arising in preparing data for analysis, and the methods for addressing these problems" and "Explain and contrast methods for evaluating the performance of automated learning algorithms." The course was offered face-to-face and taught by the same instructor.

The explanation-based classroom response system

ExplainIt is a web-based classroom response system that elicits students' self-explanations in response to instructor-generated questions and delivers feedback through the Instructor and Student Apps. The system uses self-explanations, active and constructive learning, and advanced natural language processing to enhance the CS course and its instruction. In the first semester implementation, the system featured static feedback based on instructor-provided representative responses, allowing students to compare their self-explanations with sample answers. In the second iteration, the system incorporated AI-generated adaptive feedback powered by the GPT-4o mini large language models (LLMs), which automatically evaluated student responses and provided customized and formative feedback on self-explanations. This enhancement represented a transition from instructor-dependent static instant feedback to AI-generated adaptive instant feedback, enabling more customized learning support.

The ExplainIt apps

This explanation-based classroom response system includes three Apps: Instructor Authoring Tool, Instructor Dashboard, and Student Response App. In the Instructor Authoring Tool, instructors can add the course and topics to this App. Under the Specific Course and Topic section, instructors can create questions along with corresponding sample answers (**Figure 1**). In the Instructor Dashboard, instructors can filter courses, topics, and questions to review the number of responses and the dates when questions were posed (**Figure 2**). To view a student's response to a specific question in detail, instructors can double-click the corresponding row in the chart to open a pop-up window (**Figure 3**). Regarding the student's response app, students can submit their response to the question posed by the instructor. Then they can review all the response logs as well as the feedback they received to each response in the same App section. The main difference between the two versions of the system lies in the feedback content students receive based on their submitted responses. In the first iteration, students could view only their own responses and the corresponding sample answer from the instructor (**Figure 4**). In the second iteration, students can view customized AI-feedback in addition to a correctness rating (Figure).

Figure 1

Instructor authoring tool interface for creating questions and sample answers

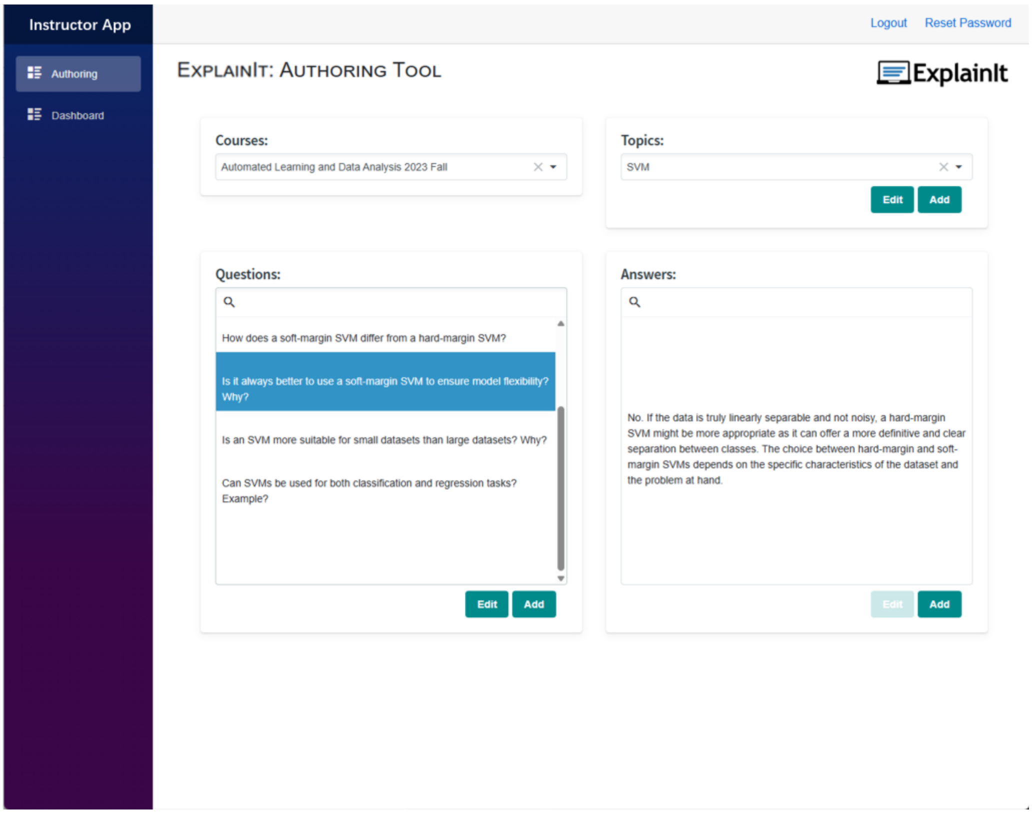


Figure 2

Instructor dashboard interface for viewing response correctness charts

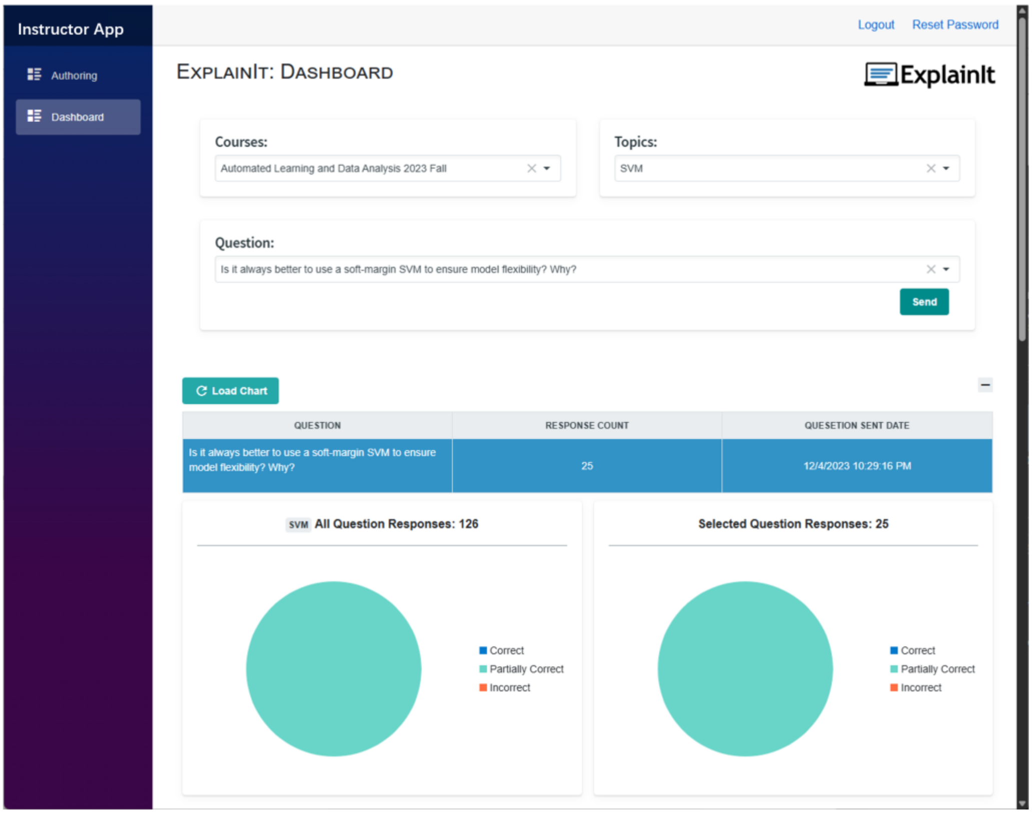


Figure 3

Instructor dashboard interface for viewing student responses to specific questions

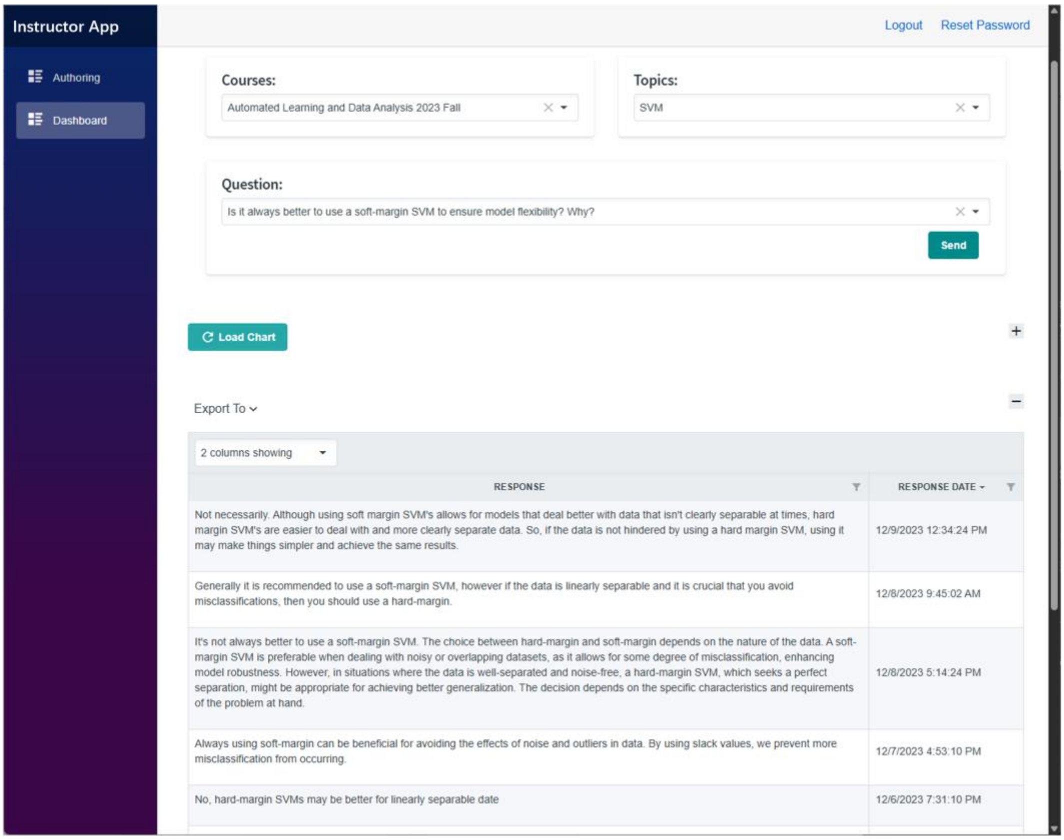


Figure 4

Student Response App Interface for the Static Feedback Version of ExplainIt

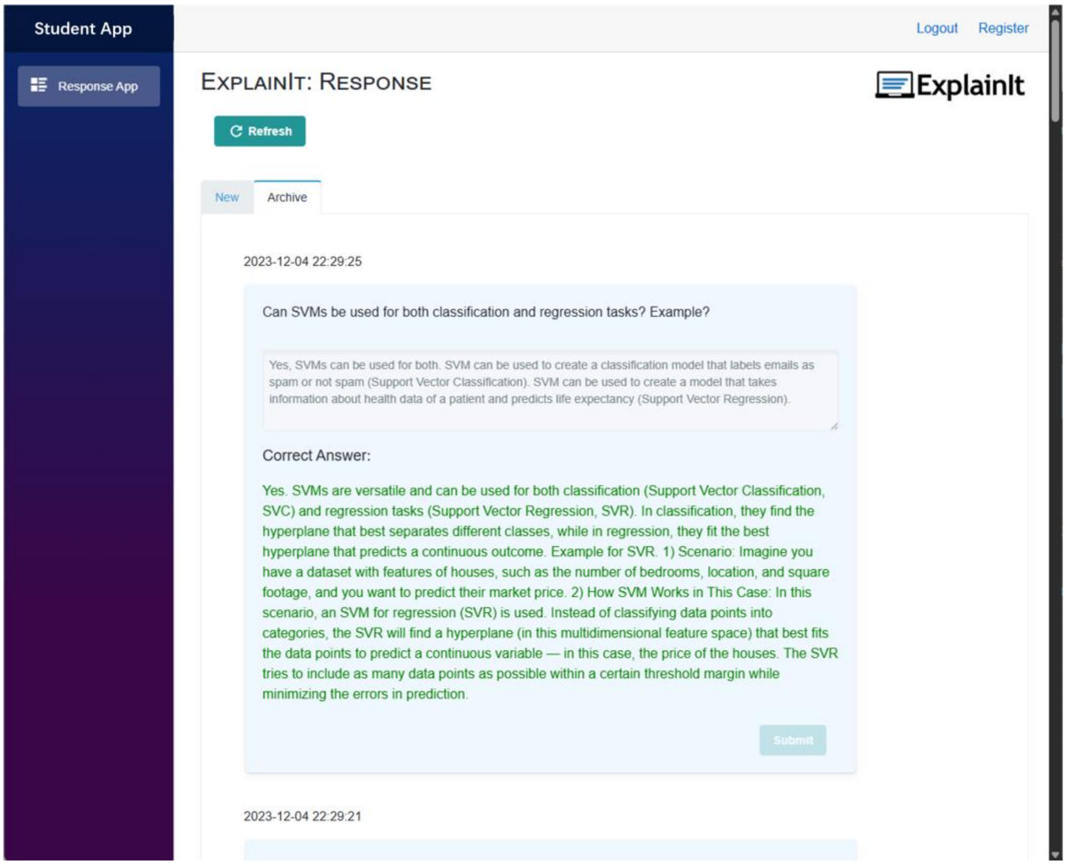
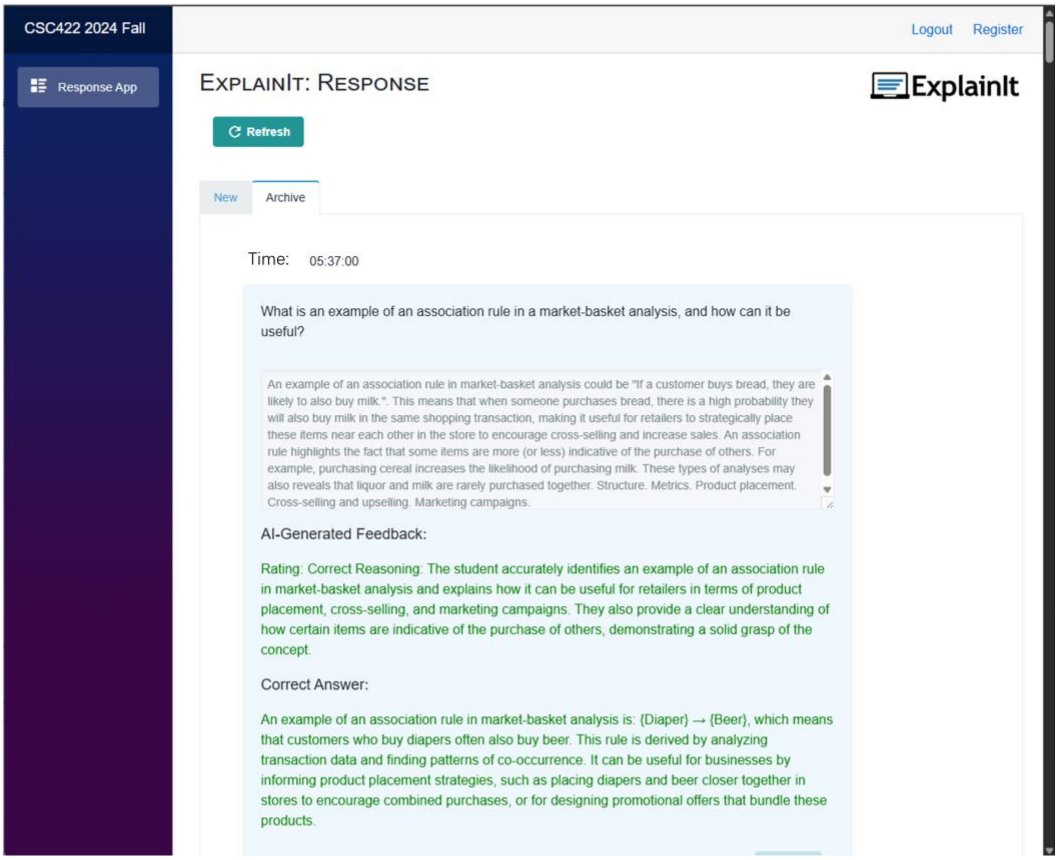


Figure 5

Student Response App Interface for the AI Feedback Version of ExplainIt



METHOD

This study adopts a comparative case study design (Yin, 2017) to explore student experiences with the two distinct feedback iterations by the system and focusing on instructional implementation in two CS courses. This approach enables an in-depth comparison of how modifications in feedback (static versus AI-generated) influence students' learning experiences. The design allows for a rich, multi-faceted comparison of two software iterations that employ different types of feedback within the same CS course context and under the same instructor.

The first-semester implementation focused on instructor posting questions and students writing self-explanation responses, followed by receiving static feedback based on instructor-prepared sample responses entered in advance. In contrast, the second semester implementation incorporated AI-based answer classification and introduced AI-generated adaptive feedback within the same course during the subsequent semester.

Participants

The participants were undergraduate CS students and one CS instructor at a public university in the Southeast U.S. The number of students enrolled in the two courses was 36 and 50, respectively. Of the enrolled students, 32 (89%) students in the first implementation and 28 (56%) students in the second implementation completed the demographic survey.

In the first implementation, participating students reported ages ranging from 18 to 28 years. Twelve students (37%) were juniors, and 20 (63%) were seniors. The gender distribution was 8 female (25%), 23 male (72%), and 1 did not report gender (3%). Regarding ethnicity, 16 students (50%) identified as White, 13 (41%) Asian, and 3 (9%) preferred not to answer. In the second implementation student ages ranged between 20 to 23 years. Twenty-one students (75%) were seniors, 6 (21%) were juniors, and 1 (4%) was sophomore. The participants included 23 male (82%), 4 female (14%), and 1 student (4%) who did not report gender. Regarding ethnicity, 10 students (36%) identified as White, 10 (36%) Asian, 3 (11%) Black or African American, 1 (3%) Hispanic or Latino, 1 (3%) Native American, 2 (8%) Other, and 1 (3%) did not respond.

Data collection procedures

The Institutional Review Board (IRB) approval was obtained before the study began, and all research procedures adhered to IRB requirements. The instructor was recruited via email outlining the study and its procedures. After expressing interest, the research team scheduled an online introduction and a training session to demonstrate the system's features and address questions to the instructor. During this session, the instructor received login credentials, was shown system features, and received guideline document for writing self-explanation questions. Before the start of the first semester, the research team met with the instructor again to address specific questions regarding setup and other issues. The team remained available for consultation via email or Zoom at any time during the study. At the beginning of the first CS course, the instructor introduced the system to his students, provided students with the study information sheet, consent form, and links to pre-surveys. Students were asked to review the study information sheet and electronically sign the IRB approved consent form to allow their demographic information and study response log data to be used for research purposes. Then all students were given a generic ID to interact with the system, thus when interacting with third-party AI tools no personally identifiable information was included or collected. De-identified student response data for surveys were stored on secure university approved cloud-based storage systems. All instructor questions and student responses were captured within ExplainIt. At the end of the course, participating students completed post-surveys. The research team downloaded all stored data and survey responses for analysis.

Data sources

Surveys

Data were collected electronically from the students using identical instruments across two semesters. Three survey instruments were used: one for student demographics (pre), one for self-efficacy (pre and post), and one for engagement and system evaluation (post). The eight-item demographic survey collected student background information such as grade level, gender and ethnicity. The self-efficacy surveys captured students' confidence in the knowledge and skills related to course objectives. The survey instruments were adapted from Unfried et al. (2015), O'Brien et al. (2018), and Brooke (1996), with minor modifications to align with the CS context of this study. The self-efficacy and user engagement surveys had five-point Likert scale. The system usability survey included both Likert-scale and open-ended questions. The open-ended questions asked students what features they "liked best" and "liked least," and which additional features would support their learning.

Student and instructor interaction system data from the system

Another data source was the instructor-posed questions, student self-explanation responses, and feedback data stored in ExplainIt.

DATA ANALYSIS

Once all survey responses were collected, the data were downloaded and prepared for analysis. Quantitative data were analyzed with descriptive statistics and non-parametric tests. These analyses provided an overview of students' perceptions of self-efficacy, engagement, and system usability across implementations. For the qualitative data, a thematic analysis (Braun & Clarke, 2012, Chapter 4) was conducted on open-ended responses. This analysis aimed to identify key patterns and themes about the quality of system feedback, usability, and perceived learning benefits.

RESULTS

Research question 1. How did the instructor and students use static feedback and adaptive AI feedback in the two CS courses?

The instructor posed varying numbers of questions across the two semesters while teaching the same course. In the first semester implementation, the instructor created questions on three CS topics—Deep Learning, Clustering, and Support Vector Machine (SVM). A total of 18 questions and sample answers were prepared by the instructor before the semester began, of which 13 were posed during course.

In the second semester, the instructor developed questions for 13 CS topics, including topics such as Data Types and Decision Trees. For these topics, he prepared 27 questions prior to semester start, and posed all questions to students throughout the semester, indicating a more comprehensive integration of feedback intervals into the instructional process. A comparison of the topics and questions across the two implementations revealed some overlap. The common topics were Deep Learning and SVM, and there was one question that appeared in both semesters under the SVM topic. Table 1 shows examples of instructor posed questions to illustrate the nature of the self-explanation questions. None of the questions posed over two semesters required a mathematical formula response, as parsing mathematical expressions can pose a significant challenge for LLMs; all required conceptual and textual self-explanations.

Table 1

Examples of instructor-created questions

Number	Topic	Question
Semester 1		
1	Deep Learning	What does the term "deep" in deep learning refer to?
3	Deep Learning	In a neural network, what function is responsible for introducing non-linearity to the model?
10	SVM	How does soft-margin SVM differ from hard-margin SVM?
Semester 2		
1	SVM	How do soft-margin SVM differ from hard-margin SVM?
11	Model Evaluation	What are the steps of k-fold cross-validation?
17	Data Types	What is the main difference between these four data types (Nominal, Ordinal, Interval, Ratio)? Please describe them in turn and compare them to the data types that preceded each category.

In the ExplainIt Student App, feedback students received differed across two semesters. In the first semester, students received instructor-prepared static feedback delivered through the system. In the second implementation, students received both AI-based correctness ratings (correct, partially correct, or incorrect) and AI-generated adaptive feedback. Table 2 and Table 3 (in APPENDIX) shows examples of feedback provided to students on their self-explanation responses.

Table 2

Examples of static instructor feedback students received in the first CS course

Question posed by the instructor	Topic	Student self-explanation response	Human Rating	Static instructor feedback
What is the main difference between K-means and hierarchical clustering?	Clustering	K-means is a partitional clustering algorithm that divides data into a fixed number of clusters (k), while hierarchical clustering creates a tree-like structure of nested clusters without requiring a predefined number of clusters.	Correct	The main difference is that K-means clustering partitions the data into K non-overlapping clusters, without any hierarchical structure, while hierarchical clustering creates a tree of clusters called a dendrogram. Hierarchical clustering does not require the number of clusters to be specified in advance and can produce a set of nested clusters at different levels of granularity.
In a neural network, what function is responsible for introducing non-linearity to the model?	Deep Learning	Nonlinear transformations in its layers	Partially Correct	Activation functions are responsible for introducing non-linearity to the model. Examples include the ReLU (Rectified Linear Unit), sigmoid, and tanh functions.
Is it always better to use a soft-margin SVM to ensure model flexibility? Why?	SVM	Yes because a soft-margin SVM is often more robust with noise and outliers.	Incorrect	If the data is truly linearly separable and not noisy, a hard-margin SVM might be more appropriate as it can offer a more definitive and clear separation between classes. The choice between hard-margin and soft-margin SVMs depends on the specific characteristics of the dataset and the problem at hand.

Table 3

Examples of adaptive AI-feedback students received in the second CS course

Question posed by the instructor	Topic	Student self-explanation response	AI Rating	AI feedback
What are the Common Properties of Similarity?	Exploratory Data Analysis	Common properties of similarity include: Reflexivity: An object is similar to itself. Symmetry: If object A is similar to object B, then B is similar to A. Transitivity: If object A is similar to B, and B is similar to C, then A is similar to C.	Correct	The student response is correct. They accurately identified the common properties of similarity, including reflexivity, symmetry, and transitivity. Each property is explained clearly and accurately. Therefore, the response is rated as correct.
What is the primary purpose of using convolutional layers in deep learning?	ANNs and Deep Learning	Convolutional layers help to reduce the complexity of a piece of data in order to store the information. For images, it will look at a grid of pixels, give it a value, and pass that on in order to reduce the amount of information present in the image.	Partially Correct	Partially correct. The primary purpose of using convolutional layers in deep learning is to extract features from the input data (such as images) by applying filters to detect patterns and structures. While it is true that convolutional layers help to reduce the complexity of the data by extracting important features, the main goal is not to simply store information but rather to learn and identify patterns that are crucial for making predictions or classifications.
What is the primary difference between partitional and hierarchical clustering?	Clustering	Hierarchical clustering allows subclusters to form within clusters, while partitional clustering does not allow for any cluster overlap.	Incorrect	The student response is incorrect. The primary difference between partitional and hierarchical clustering is that partitional clustering requires the user to specify the number of clusters beforehand, while hierarchical clustering does not require the user to specify the number of clusters. Hierarchical clustering can be agglomerative (bottom-up) or divisive (top-down) and does allow for subclusters to form within clusters. Therefore, the student's response is incorrect.

In terms of student performance, across 314 student responses in the static-feedback the percentage rated as correct for each question ranged from 21% to 100%. The question with the highest correctness rate (100%) was the final question: *"Can SVMs be used for both classification and regression tasks? Example?"* The question with the lowest correctness rate (21%) was the first question: *"What does the term 'deep' in deep learning refer to?"* Overall, 11 out of 13 questions received responses with correctness rates above 50%, indicating strong performance through self-explanation responses. Additionally, 18 out of 314 responses were rated as incorrect, with incorrect rates ranging from 2% to 27% across all questions.

In the adaptive AI-generated feedback implementation, across 1,322 student self-explanation responses, the correctness rate ranged from 36% to 98%. The question receiving the highest correct rating at 98% was "*Why Dimensionality Reduction?"* while the lowest correct rating as 36% "*How to calculate the misclassification error at node t?*" A total of three responses marked as incorrect by the AI.

Research question 2. How does student self-efficacy in CS courses differ between static feedback and adaptive AI-generated feedback?

The number of completed student self-efficacy surveys varied both across different instruments within each implementation and across the two implementations for the same instrument. For the static-feedback semester implementation 16 (44%) students completed both pre- and post- self-efficacy surveys. Regarding the AI-generated feedback implementation, 13 (26%) students completed both the pre- and post-self-efficacy surveys.

Given the student responses resulted in a small sample in both semesters and the ordinal nature of the data, the Wilcoxon signed-rank test was conducted for each pair of pre- and post self-efficacy data for the two iterations of the system respectively. The Wilcoxon signed-rank test indicated no significant differences in self-efficacy during the static-feedback implementation between pre- and post-surveys (Z= -.63, p= .53, r= .16). For the AI-generated feedback, the Wilcoxon signed-rank test indicated no significant differences in self-efficacy between pre- and post-surveys (Z= -.36, p= .72, r= .11). These results suggested that students' self-efficacy remained mainly stable through the pre- and post-surveys in either first or second implementation.

Research question 3. How do students' perceptions of classroom response system usability compare between the static feedback and adaptive AI-generated feedback versions in CS courses?

For the static-feedback implementation, system usability data were collected from 17 (47%) students. The thematic analysis of the open-ended question responses uncovered several key themes. Most students indicated that the classroom response system was helpful for their learning and deepening their understanding of course concepts. In terms of disliked aspects, a few students mentioned timeout issues that required reloading the ExplainIt student application. Additionally, students suggested features such as viewing answer history, visualizing instructor questions with graphs, distinguishing graded from non-graded questions (e.g., participation), and summarizing responses.

For the AI-generated feedback implementation, open-ended responses were obtained from 17 (34%) students. Students highlighted several favorable aspects, appreciation for the instant and insightful feedback that fostered deeper thinking (7) and the AI-generated evaluative feedback (5). Three additional positive responses included the system's ease of use, its simple interface design, and the engaging nature of the instructor questions. Additional suggestions included making the feedback area more prominently displayed with figures and graphs, adding a hint function, and incorporating a discussion board.

Research question 4. How do students' perceptions of engagement differ between static and adaptive AI-generated feedback in CS courses?

In the first implementation with instructor static-feedback, 18 (50%) students completed the user engagement survey after using ExplainIt for a semester. The mean engagement ratings for each question across all participants was 3.48. This showed that most students perceived themselves as engaged while system being used in their CS course. The following three items in the survey received the highest ratings: "Feedback system was not confusing to use" (4.28); "Feedback system was not stressful (4.28)"; and "Using the feedback system to learn was an interesting experience" (4.11).

As for the AI-based feedback implementation, 19 (38%) students completed the engagement survey. The overall mean engagement score across all items was 3.25. The three statements that received the highest ratings were: "Using the feedback system was rewarding." (4.17); "Using the feedback system was worthwhile." (4.11); "Using the feedback system to learn was an interesting experience." (4.06).

DISCUSSION

This comparative case study examined how static feedback, compared to adaptive AI-generated feedback, influenced instructors' and students' use of digital feedback system, students' self-efficacy, and engagement in CS courses. The results showed that the instructor posed more self-explanation questions in the adaptive AI-generated feedback implementation compared to the static-feedback implementation, resulting in higher student engagement (27 questions, generating 1,322 total responses from 48 to 50 students per question). The higher number of questions posed by the instructor may reflect multiple rationales. One might be the distinct features of the two systems and their feedback design. In the AI-generated feedback condition, the instructor did not need to create sample responses, allowing more time to generate self-explanation questions and fully leverage the tool's AI capabilities. This increase in number of instructor-generated questions corresponded with greater student participation in the AI-supported feedback iteration versus the static-feedback iteration, suggesting that adaptive, personalized feedback fostered consistent participation and deeper conceptual understanding.

Another possible explanation for differences in student engagement and performance may relate to the self-explanation question topics covered across the two semesters. Although the same course was taught by the same instructor, the instructor posed questions on some common topics, while a greater number of different topics of self-explanation questions were posed in the second semester. As a result, student engagement and participation may have been influenced by issues like topic difficulty, student interest, or relevance of the topics to CS students.

A second explanation is that the instructor may have developed greater confidence in using the system during the first semester and subsequently implemented it more frequently in the second semester, which coincided with the use of AI-generated feedback. Lastly, the instructor may recognize the value of formative feedback and time efficiencies associated with using such feedback in his courses, leading to more frequent use of self-explanation questions. These possible interpretations align with prior research suggesting that AI may have potential for low-stakes formative feedback while reducing instructor workload (Burner et al. 2025; & Zacharis & Papadakis, 2025). Future research may investigate the same topic coverage under different feedback conditions by using a controlled experimental design and further explore instructors' experiences to better understand instructional decision-making related to formative feedback cycles in STEM contexts.

Another interesting observation is that although participation was higher in AI-generated feedback, student engagement ratings were slightly lower compared to the static feedback implementation (3.25 versus 3.48). One possible explanation might be that topics covered in second semester was more difficult or less interesting to students. Another possibility is that the instructor posed more questions and did so more frequently in the AI-supported setting. While this likely increased students' course participation frequency, it may have made the activity feel routine, thereby reducing its perceived novelty and overall student engagement. Similar findings were reported by Zhao (2023) in a CS course, indicating that increased frequency of posing questions diminished students' motivation. A further explanation for the lower engagement could be also the nature of cognitive involvement required of students while processing the feedback they received. In the static-feedback condition, students actively analyzed the instructor's sample answer and compared it with their own, fostering deeper cognitive processing. In contrast, the AI-supported version provided detailed, customized explanations, which may have scaffolded learning but reduced the need for students to engage in active analysis. Future research could explore the optimal number and frequency of instructor questions per session to maintain engagement without causing fatigue. Another avenue might involve diversifying question formats (e.g., figures, videos, cases), or varying question types (e.g., multiple choice, definition), to sustain student engagement and interest.

Another challenge noted in the literature is that STEM instructors' often have limited time to provide immediate formative feedback in courses (Henderson et al., 2019a, 2019b; Nakamoto et al., 2023, 2024). In this study, the implementation of AI-generated adaptive feedback in CS courses effectively addressed this barrier: a single instructor managed 1,322 open-ended self-explanation responses by leveraging AI technologies. As a result, students not only received assessments of the correctness of their responses, but received adaptive feedback tailored to their self-explanations. This streamlined feedback approach enabled the instructor to post several self-explanation questions, known to promote deeper cognitive engagement with course content (Shapiro et al., 2017), and to focus on broader range of the CS content. This approach was beneficial for student understanding and engagement in the course, prioritizing pedagogy over concerns about workload and time commitment.

Building on this streamlined approach, ExplainIt further enhanced classroom interactivity by addressing the issue of providing formative feedback in large STEM courses, a persistent concern mentioned by Freeman et al. (2014), Mirhosseini et al. (2023), and Villegas-Ch et al. (2025). This system enhanced classroom interactivity by addressing the challenge of providing formative feedback in large STEM courses. The number of responses submitted by each student in both iterations indicates active engagement with course material, contrasting with the passive reliance on lectures often observed in CS courses (Gonsar et al., 2021). This finding is in line with prior research and demonstrates that AI-supported feedback systems may fulfill their promise by enabling instructors to provide customized feedback and perform just-in-time assessments of students within courses (Joseph & Uzondu, 2024; Seco et al., 2025).

Lastly, student performance assessed by the system, along with the self-efficacy and perceptions of engagement survey results, suggests that both the static and adaptive versions of the feedback system helped maintain moderate to high levels of engagement among undergraduate CS students. This indicates the system's viability as an instructional support tool. The observed increase in student engagement in the AI-enhanced iteration further suggests that adaptive, just-in-time feedback fosters sustained engagement and possibly deeper conceptual understanding, critical for STEM persistence and outcomes (Freeman et al., 2014; Marwan et al., 2020, Mirhosseini et al., 2023).

The AI-supported feedback implementation covered more topics (13 vs. 3) and more questions (27 vs. 13) than static-feedback implementation, which may also indicate greater integration of the system into instruction in the second semester. In the static-feedback implementation, correctness improved from early to later questions, suggesting gradual learning. In the adaptive feedback implementation, correctness was consistently high across most questions, indicating that adaptive feedback may accelerate learning and reduce variability. The adaptive feedback likely contributed to improved accuracy and reduced errors, as students received tailored guidance rather than static responses. This finding is similar with Villegas-Ch et al. (2025) study that dynamic, AI-generated feedback may better support conceptual understanding and error correction in CS courses. Future studies may focus on AI-supported and allowing students to review the feedback they received and re-submit responses.

The integration of static feedback and adaptive AI-generated feedback combined with assessments of the correctness of student responses in a CS course, demonstrated its capacity to foster increased student participation and improved student performance. The static feedback implementation showed a clear progression in correctness, starting with the lowest rate (21%) on the first question and reaching 100% on the final question. This pattern indicates gradual improvement in conceptual understanding, as course topics become more complex over the semester, likely driven by repeated exposure to static feedback and students' active participation. In contrast, the adaptive feedback implementation maintained consistently high correctness ratings across most questions, suggesting that dynamic, adaptive feedback may accelerate learning and may reduce variability in student performance.

Despite the promise of AI-supported feedback and active learning in CS courses, a few limitations should be noted. First, there is the potential risk of instructors over-relying on AI-based systems to assess student work without verifying the accuracy of the feedback students receive. In this project, the team attempted to ensure accuracy by evaluating the system's ability to assess student self-explanations (Carpenter et al., 2024; Ozogul et al., 2025), and the instructor was supported by a research team continuously to allow for monitoring of output, and by doing sample ratings. In addition to these "human in the loop" precautions, reliability was further bolstered using a low temperature (0.2) for the LLM, which encourages the LLM to behave much more predictably and produce very similar outputs for the same input (Fastowski et al., 2025). Even so, to make sure that there will be no issues with LLMs the research team needs to iteratively refine the LLM prompts and conduct repeated testing to evaluate the model's assessment and feedback generation performance over time. In other contexts, the responsibility for accuracy checks may also fall on the instructor. Until such checks are completed for any comparable system, instructors should approach reliance on AI-supported feedback with caution. Second, because this study focused on self-explanations in CS, due to the nature of the field there was limited variations in student responses. In other STEM contexts, greater variability may affect the AI's ability to accurately evaluate and rate responses. Third, we did not capture instructor perspective in this study as it was focused on student experiences. As such, our interpretations of instructor rationales and use of the feedback versions is limited. Lastly, even though both implementations were done in the same course, the set of topics covered in the second semester was not identical to those in the first. Due to this we must further examine our attributions of student engagement that may result from exposure to the feedback version and consider the potential differences in topic difficulty or student interest in a topic.

CONCLUSION

Overall, the findings highlight the potential of AI-generated adaptive CS by offering more personalized guidance than static feedback. Broader integration of ExplainIt may have important implications for STEM education, as instructors' repeated use across semesters suggests increased confidence in the system's instructional value and its potential to reduce workload while maintaining pedagogical quality. For students, sustained exposure to adaptive AI-supported feedback may foster deeper conceptual understanding, improved self-efficacy, and persistence in STEM fields. Future research should examine whether these benefits generalize across STEM disciplines and instructional contexts, and investigate long-term effects on motivation, retention, engagement, and learning outcomes, including how such systems complement active learning strategies.

Acknowledgement

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.

Funding

This research was supported by funding from the National Science Foundation under Grants DUE-2111473 and DUE-2111216.

Ethical statement

IRB approval was obtained prior to start of the study and all study procedures adhered to the IRB that is approved. IRB number 13268

Competing interests

The authors declare that they have no relevant financial or non-financial interests to disclose.

Author contributions

The authors jointly contributed to the conception, design and implementation, data analysis of the study. The preparation of this manuscript and discussion sections was led by Ozogul and Zheng. All authors provided final approval and review of our manuscript.

Data availability

All data collected or analyzed during this study are included in this research article. The data are made available in a way that ensures their integrity. In accordance with applicable data protection and privacy rules, we followed the best practices in data handling data and sharing findings with readers.

AI disclosure

Artificial intelligence tools were not used.

Biographical sketch

Gamze Ozogul, is a Professor of Instructional Systems Technology in the Department of Learning, Design, and Adult Education in the School of Education at Indiana University Bloomington. She earned her Ph.D. in Educational Technology from Arizona State University in 2006 and subsequently completed a postdoc in Electrical Engineering at the same institution. Her research focuses on instructional design and evaluation, feedback design and delivery, and investigating instructional strategies to support student outcomes and engagement in emerging learning environments.

Xiaoying Zheng, is a Ph.D. student in Instructional Systems Technology in the Department of Learning, Design, and Adult Education in the School of Education at Indiana University Bloomington. She holds an M.S. in Computer Technology. *Her* research focuses on technology integration in higher education, the design and improvement of instructional feedback, and the promotion of self-directed learning. She is currently investigating integrating an AI-based classroom response and feedback system to support undergraduate STEM learning.

Wookhee Min, is a Senior Research Scientist in the Department of Computer Science at North Carolina State University (USA). He also serves as Managing Director of the NSF AI Institute for Engaged Learning (EngageAI) and is affiliated with the Center for Educational Informatics. He earned his Ph.D. in Computer Science from North Carolina State University in 2016. His research lies at the intersection of artificial intelligence and education, with expertise in adaptive learning technologies. His key research interests include game-based learning environments, multimodal learning analytics, user modeling, and natural language processing.

Seung Lee, is a Research Scientist in the Center for Educational Informatics at North Carolina State University. His research centers on applying artificial intelligence and machine learning to creating adaptive learning environments for applications in education and training with a focus on developing intelligent tutoring systems and multimodal natural language dialogue systems for intelligent narrative-centered learning environments.

Jordan Esiason, is a PhD student in North Carolina State University's Computer Science department in the U.S.A. He has previously *completed* a B.S. in statistics and a B.M. in music composition. His primary field of research is computer science education with a focus on intelligent classroom support systems, broadening access to computer science education, and postsecondary student pathways.

Dan Carpenter is an AI Engineer at Amplify Education (USA). He earned his Ph.D. in Computer Science from North Carolina State University in 2023. His research interests center around the design and implementation of AI-supported learning experiences, specifically exploring how natural language processing, narrative-centered learning, and self-regulated learning can be synthesized to improve student outcomes.

James C. Lester, is the Goodnight Distinguished University Professor in Artificial Intelligence and Machine Learning in the Department of Computer Science at North Carolina State University, where he is the Director of Center for Educational Informatics. His current work ranges from AI-driven narrative-centered learning environments and conversational agents for learning to multimodal learning analytics and sketch-based learning environments. The AI-driven learning environments he and his team create have been used by thousands of K-12 students. He is a Fellow of the Association for the Advancement of Artificial Intelligence (AAAI).

Disclaimer/Publisher's Note

REFERENCES

Acar, E., Yigit, F., & Deiri, Y. (2025). A focused review of artificial intelligence in education: Evolution and challenges. *Journal of Interdisciplinary Research in Artificial Intelligence and Society*, *1*(1), Article 3. https://doi.org/10.20897/jirais/17640

Adamakis, M., & Rachiotis, T. (2025). Artificial intelligence in higher education: A state-of-the-art overview of pedagogical integrity, artificial intelligence literacy, and policy integration. *Encyclopedia*, *5*(4), Article 180. https://doi.org/10.3390/encyclopedia5040180

Adekola, S. (2025). Integrating AI into STEM education: Challenges and opportunities. In P. A. Okebukola (Ed.), *Handbook of artificial intelligence and quality higher education: AI and curriculum development for the future* (Vol. 2, pp. 139–144). Sterling Publishers.

Baartman, L. K., & Quinlan, K. M. (2024). Assessment and feedback in higher education reimagined: Using programmatic assessment to transform higher education. *Perspectives: Policy and Practice in Higher Education*, *28*(2), 57–67. https://doi.org/10.1080/13603108.2023.2283118

Beauchamp, G., & Kennewell, S. (2010). Interactivity in the classroom and its impact on learning. *Computers & Education*, *54*(3), 759–766. https://doi.org/10.1016/j.compedu.2009.09.033

Bisra, K., Liu, Q., Nesbit, J. C., Salimi, F., & Winne, P. H. (2018). Inducing self-explanation: A meta-analysis. *Educational Psychology Review*, *30*(3), 703–725. https://doi.org/10.1007/s10648-018-9434-x

Blasco-Arcas, L., Buil, I., Hernández-Ortega, B., & Sese, F. J. (2013). Using clickers in class: The role of interactivity, active collaborative learning and engagement in learning performance. *Computers & Education*, *62*, 102–110. https://doi.org/10.1016/j.compedu.2012.10.019

Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology* (Vol. 2, pp. 57–71). American Psychological Association. https://doi.org/10.1037/13620-004

Brooke, J. (1996). SUS: A 'quick and dirty' usability scale. In P. W. Jordan, B. Thomas, I. L. McClelland, & B. Weerdmeester (Eds.), *Usability evaluation in industry* (pp. 189–194). Taylor & Francis.

Burner, T., Lindvig, Y., & Wærness, J. I. (2025). "We should not be like a dinosaur"—Using AI technologies to provide formative feedback to students. *Education Sciences*, *15*(1), Article 58. https://doi.org/10.3390/educsci15010058

Campbell, J., & Mayer, R. E. (2009). Questioning as an instructional method: Does it affect learning from lectures? *Applied Cognitive Psychology*, *23*(6), 747–759. https://doi.org/10.1002/acp.1513

Carpenter, D., Min, W., Lee, S., Ozogul, G., Zheng, X., & Lester, J. (2024, June). 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 (BEA 2024)* (pp. 403–413). https://aclanthology.org/2024.bea-1.33.pdf

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, *20*(1), Article 43. https://doi.org/10.1186/s41239-023-00411-8

Chen, P., Yang, D., Zhao, J., Yang, S., & Lavonen, J. (2025). The effects of self-explanation on secondary school students' computational thinking and programming behaviour. *Journal of Computer Assisted Learning*, *41*(5), Article e70116. https://doi.org/10.1111/jcal.70116

Chen, Q. (2025). Students' perceptions of AI-powered feedback in English writing: Benefits and challenges in higher education. *International Journal of Changes in Education*. https://doi.org/10.47852/bonviewIJCE52025580

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*(2), 145–182. https://doi.org/10.1207/s15516709cog1302_1

Chiu, J. L., & Chi, M. T. H. (2014). Supporting self-explanation in the classroom. In V. A. Benassi, C. E. Overson, & C. M. Hakala (Eds.), *Applying science of learning in education: Infusing psychological science into the curriculum* (pp. 91–103). American Psychological Association.

Diaz, P., Hrastinski, S., & Norström, P. (2024). How using a response system in blended synchronous seminars encourages online and onsite student participation. *Education and Information Technologies*, *29*, 19889–19911. https://doi.org/10.1007/s10639-024-12665-4

Elme, L., Jørgensen, M. L., Dandanell, G., Mottelson, A., & Makransky, G. (2022). Immersive virtual reality in STEM: Is IVR an effective learning medium and does adding self-explanation after a lesson improve learning outcomes? *Educational Technology Research and Development*, *70*, 1601–1626. https://doi.org/10.1007/s11423-022-10139-3

Er, E., Akçapınar, G., Bayazıt, A., Noroozi, O., & Banihashem, S. K. (2025). Assessing student perceptions and use of instructor versus AI-generated feedback. *British Journal of Educational Technology*, *56*(3), 1074–1091. https://doi.org/10.1111/bjet.13558

Fastowski, A., Prenkaj, B., & Kasneci, G. (2025). From confidence to collapse in LLM factual robustness. *Conference on Empirical Methods in Natural Language Processing*. https://www.semanticscholar.org/reader/5e51fe92bb1ebd2e89b06afb9f776c65b8c18771

Firat, M. (2023). What ChatGPT means for universities: Perceptions of scholars and students. *Journal of Applied Learning and Teaching*, *6*(1), 57–63. https://doi.org/10.37074/jalt.2023.6.1.22

Fonseca, B. A., & Chi, M. T. H. (2011). Instruction based on self-explanation. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 310–335). Routledge.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Gonsar, N., Patrick, L., & Cotner, S. (2021). Graduate- and undergraduate-student perceptions of and preferences for teaching practices in STEM classrooms. *Disciplinary and Interdisciplinary Science Education Research*, *3*, Article 6. https://doi.org/10.1186/s43031-021-00035-w

González-Cacho, T., & Abbas, A. (2022). Impact of interactivity and active collaborative learning on students' critical thinking in higher education. *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, *17*(3), 254–261. https://doi.org/10.1109/RITA.2022.3191286

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Henderson, M., Ajjawi, R., Boud, D., & Molloy, E. (2019a). Identifying feedback that has impact. In M. Henderson, R. Ajjawi, D. Boud, & E. Molloy (Eds.), *The impact of feedback in higher education* (pp. 15–34). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-25112-3_2

Henderson, M., Ryan, T., & Phillips, M. (2019b). The challenges of feedback in higher education. *Assessment & Evaluation in Higher Education*, *44*(8), 1237–1252. https://doi.org/10.1080/02602938.2019.1599815

Hooda, M., Rana, C., Dahiya, O., Rizwan, A., & Hossain, M. S. (2022). Artificial intelligence for assessment and feedback to enhance student success in higher education. *Mathematical Problems in Engineering*, *2022*, Article 5215722. https://doi.org/10.1155/2022/5215722

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

Jammeh, A. L. J., Karegeya, C., & Ladage, S. (2025). Clicker-integrated instruction and conventional instruction: The comparative evaluations of students' performances in chemistry. *Education and Information Technologies*, *30*, 5331–5351. https://doi.org/10.1007/s10639-024-12992-6

Jelks, S. M., & Crain, A. M. (2020). Sticking with STEM: Understanding STEM career persistence among STEM bachelor's degree holders. *The Journal of Higher Education*, *91*(5), 805–831. https://doi.org/10.1080/00221546.2019.1700477

Joseph, O. B., & Uzondu, N. C. (2024). Integrating AI and machine learning in STEM education: Challenges and opportunities. *Computer Science & IT Research Journal*, *5*(8), 1732–1750.

Kaliisa, R., Misiejuk, K., López-Pernas, S., & Saqr, M. (2026). How does artificial intelligence compare to human feedback? A meta-analysis of performance, feedback perception, and learning dispositions. *Educational Psychology*, *46*(1), 80–111. https://doi.org/10.1080/01443410.2025.2553639

Khairuddin, Z., Shahabani, N. S., Ahmad, S. N., Ahmad, A. R., & Zamri, N. A. (2024). Students' perceptions on the artificial intelligence (AI) tools as academic support. *Malaysian Journal of Social Sciences and Humanities*, *9*(11), Article e003087. https://doi.org/10.47405/mjssh.v9i11.3087

López-Pernas, S., Misiejuk, K., Oliveira, E., & Saqr, M. (2025). The dynamics of the self-regulation process in student-AI interactions. In *Proceedings of the 25th Koli Calling International Conference on Computing Education Research* (pp. 1–12). Association for Computing Machinery. https://doi.org/10.1145/3769994.3770043

Marwan, S., Gao, G., Fisk, S., Price, T. W., & Barnes, T. (2020). Adaptive immediate feedback can improve novice programming engagement and intention to persist in computer science. In *Proceedings of the 2020 ACM Conference on International Computing Education Research* (pp. 194–203). Association for Computing Machinery. https://doi.org/10.1145/3372782.3406264

Mayer, R. E., Stull, A., DeLeeuw, K., Almeroth, K., Bimber, B., Chun, D., Bulger, M., Campbell, J., Knight, A., & Zhang, H. (2009). Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary Educational Psychology*, *34*(1), 51–57. https://doi.org/10.1016/j.cedpsych. 2008.04.002

Mirhosseini, S., Henley, A. Z., & Parnin, C. (2023, March). What is your biggest pain point? An investigation of CS instructor obstacles, workarounds, and desires. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education* (Vol. 1, pp. 291–297). Association for Computing Machinery. https://dl.acm.org/doi/10.1145/3545945.3569816

Muangprathub, J., Boonjing, V., & Chamnongthai, K. (2020). Learning recommendation with formal concept analysis for intelligent tutoring system. *Heliyon*, *6*(10), Article e05227. https://doi.org/10.1016/j.heliyon.2020.e05227

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), Article 15577. https://doi.org/10.3390/su152115577

Nakamoto, R., Flanagan, B., Dai, Y., Yamauchi, T., Takami, K., & Ogata, H. (2024). Integrating self-explanation and operational data for impasse detection in mathematical learning. *Research and Practice in Technology Enhanced Learning*, *20*, Article 19. https://doi.org/10.58459/rptel.2025.20019

O'Brien, H. L., Cairns, P., & Hall, M. (2018). A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies*, *112*, 28–39. https://doi.org/10.1016/j.ijhcs.2018.01.004

Ortiz Moreno, H., Godoy-Rangel, C., & Ahumada-Lazo, R. (2025). From lectures to active learning: A case study on problem-based learning in the material balances course. *International Journal of Mechanical Engineering Education*, 1–22. https://doi.org/10.1177/03064190251356618

Ozogul, G., Zheng, X., Min, W., Lee, S., Carpenter, D., & Lester, J. (2025). Towards AI-enhanced classroom response system eliciting self-explanations in computer science courses. *The Journal of Applied Instructional Design*, *14*(2), 335–341. https://doi.org/10.59668/2222.20820

Park, L. E., O'Brien, C., Italiano, A., Ward, D. E., & Panlilio, Z. (2023). "That's a great question!" Instructors' positive responses to students' questions improve STEM-related outcomes. *Self and Identity*, *22*(6), 849–895. https://doi.org/10.1080/15298868.2023.2207836

Park, L. E., Ward, D. E., Moore-Russo, D., Rickard, B., Vessels, V., & Hundley, J. (2024). Positive feedback as a lever to boost students' STEM outcomes. *Personality and Social Psychology Bulletin*, *52*(1), 176–197. https://doi.org/10.1177/01461672241265954

Pitt, E., & Quinlan, K. M. (2022). *Impacts of higher education assessment and feedback policy and practice on students: A review of the literature 2016–2021*. Advance HE. https://www.advance-he.ac.uk/knowledge-hub/impacts-higher-education-assessment-and-feedback-policy-and-practice-students-review

Prasad, P., & Sane, A. (2024). A self-regulated learning framework using generative AI and its application in CS educational intervention design. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education* (Vol. 1, pp. 1070–1076). Association for Computing Machinery. https://doi.org/10.1145/3626252.3630828

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

Roy, M., & Chi, M. T. H. (2005). The self-explanation principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 271–286). Cambridge University Press.

Sandoval-Medina, C., Arévalo-Mercado, C. A., Muñoz-Andrade, E. L., & Muñoz-Arteaga, J. (2024). Self-explanation effect of cognitive load theory in teaching basic programming. *Journal of Information Systems Education*, *35*(3), 303–312. https://doi.org/10.62273/GMIV1698

Seco, D., Grösser, S., & Pedrosa, A. M. (2025). Use of generative artificial intelligence tools in higher education environments. *Multidisciplinary Journal for Education, Social and Technological Sciences*, *12*(1), 156–175. https://doi.org/10.4995/muse.2025.23623

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

Shishakly, R. (2025). Understanding AI in higher education: Gendered and intersectional students' experience with ChatGPT use. *European Journal of STEM Education*, *10*(1), Article 36. https://doi.org/10.20897/ ejsteme/17646

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Slimi, Z., Benayoune, A., & Alemu, A. E. (2025). Students' perceptions of artificial intelligence integration in higher education. *European Journal of Educational Research*, *14*(2), 471–484. https://doi.org/10.12973/eu-jer.14.2.471

Tamang, L. J., Alshaikh, Z., Ait Khayi, N., & Rus, V. (2020). The effects of open self-explanation prompting during source code comprehension. In *Proceedings of the Thirty-Third International Florida Artificial Intelligence Research Society Conference* (pp. 451–456). AAAI Press.

Tang, X., Wong, S., Huynh, M., He, Z., Yang, Y., & Chen, Y. (2024). SPHERE: Scaling personalized feedback in programming classrooms with structured review of LLM outputs. *arXiv*. https://doi.org/10.48550/arXiv.2410.16513

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., Chambwe, N., Cintrón, D. L., Cooper, J. D., Dunster, G., Grummer, J. A., Hennessey, K., Hsiao, J., Iranon, N., Jones, L., Jordt, H., Keller, M., Lacey, M. E., Littlefield, C. E., ... Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences*, *117*(12), 6476–6483. https://doi.org/10.1073/pnas.1916903117

Uğraş, H., Uğraş, M., Papadakis, S., & Kalogiannakis, M. (2024). ChatGPT-supported education in primary schools: The potential of ChatGPT for sustainable practices. *Sustainability*, *16*(22), Article 9855. https://doi.org/10.3390/su16229855

UNESCO. (2019). *Beijing consensus on artificial intelligence and education*. https://unesdoc.unesco.org/ark:/48223/pf0000368303

Unfried, A., Faber, M., Stanhope, D. S., & Wiebe, E. (2015). The development and validation of a measure of student attitudes toward science, technology, engineering, and math (S-STEM). *Journal of Psychoeducational Assessment*, *33*(7), 622–639. https://doi.org/10.1177/0734282915571160

Vihavainen, A., Miller, C. S., & Settle, A. (2015). Benefits of self-explanation in introductory programming. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education* (pp. 284–289). Association for Computing Machinery. https://doi.org/10.1145/2676723.2677260

Villegas-Ch, W., Buenano-Fernandez, D., Navarro, A. M., & Mera-Navarrete, A. (2025). Adaptive intelligent tutoring systems for STEM education: Analysis of the learning impact and effectiveness of personalized feedback. *Smart Learning Environments*, *12*, Article 41. https://doi.org/10.1186/s40561-025-00389-y

Yin, R. K. (2009). *Case study research and applications: Design and methods* (4th ed.). SAGE Publications.

Zacharis, G., & Papadakis, S. (2025). Can AI grade like a human? Validity, reliability, and fairness in university coursework assessment. *Educational Process: International Journal*, *19*, Article e2025591. https://doi.org/10.22521/edupij.2025.19.591

Zhao, X. (2023). *The impact of live polling quizzes on student engagement and performance in computer science lectures*. arXiv. https://doi.org/10.48550/arXiv.2309.12335