electronics

MDPI

Article

VClass Engager: A Generative AI-Based System for Enhancing Student Engagement in Synchronous Online Classes

Ali Alammary

College of Computing and Informatics, Saudi Electronic University, Riyadh 11673, Saudi Arabia; a.alammary@seu.edu.sa

Abstract

Student disengagement remains a major barrier to effective learning in synchronous online classrooms, where lack of interaction, limited feedback, and screen fatigue often contribute to passive participation. Despite the growing use of generative AI, there is a notable lack of empirical research investigating the application of generative AI in addressing engagement challenges in synchronous online sessions. This study introduces VClass Engager, a novel experimental system that utilizes generative AI to foster student participation and deliver instant, personalized feedback during live virtual sessions. The system integrates several features, including instant analysis of student answers to chat questions, real-time AI-generated feedback, and a leaderboard that displays students' cumulative scores to promote sustained engagement. To assess its effectiveness, the system was evaluated across multiple courses. Engagement was measured by tracking participation in three in-class formative questions, and response quality was analyzed using a cumulative link mixed model (CLMM). In addition, a post-session survey captured students' perceptions regarding usability, motivational impact, and feedback quality. Results demonstrated statistically significant increases in student participation and response quality in sessions using VClass Engager compared to baseline sessions. Survey responses revealed high levels of satisfaction with the system's ease of use and motivational aspects. By combining AI and gamification, this study provides early empirical evidence for a promising approach to enhancing engagement in synchronous online learning.

Keywords: artificial intelligence; generative AI; large language models; educational technology; formative assessment; gamification; online learning

check for updates

Academic Editor: George A. Tsihrintzis

Received: 9 September 2025
Revised: 22 October 2025
Accepted: 22 October 2025
Published: 23 October 2025

Citation: Alammary, A. VClass Engager: A Generative AI-Based System for Enhancing Student Engagement in Synchronous Online Classes. *Electronics* **2025**, *14*, 4154. https://doi.org/10.3390/electronics14214154

Copyright: © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With the increasing shift towards online learning, student engagement in synchronous online sessions has become a critical challenge. Unlike traditional classrooms, synchronous online sessions often suffer from reduced participation, limited interaction, and difficulties in maintaining student motivation [1]. Effective engagement is crucial for ensuring active learning and knowledge retention; however, many existing virtual learning strategies fail to sustain meaningful participation [2]. Formative assessment plays a crucial role in monitoring student engagement and understanding, providing immediate feedback, and guiding learning progress [3]. However, implementing effective formative assessments in synchronous online sessions remains a challenge due to difficulties in gathering real-time information on students' needs, checking students' understanding during live instruction, providing individualized feedback, and building meaningful relationships with students in the absence of face-to-face cues and interactions [4].

Generative AI, such as ChatGPT and Gemini, has shown promise in enhancing educational experiences by enabling real-time feedback, automating formative assessments, and improving student interaction [5]. AI-driven feedback has been shown to improve learning outcomes by delivering personalized insights, fostering greater learner engagement, and adapting instruction to individual needs [6]. Meanwhile, gamification techniques, such as leaderboards, badges, and points, have demonstrated effectiveness in increasing student motivation, fostering competitive learning environments, and enhancing participation [7]. Although the benefits of generative AI and gamification have been well documented, there remains a notable lack of empirical research investigating their application to address the ongoing challenges of student engagement in synchronous online classes. The integration of generative AI and gamification may offer a promising pathway toward enhancing interactivity and student engagement in synchronous online learning contexts.

This study proposed and empirically evaluated VClass Engager, a novel experimental system that integrates generative AI and gamification to address the persistent challenges of student disengagement and passive participation in synchronous online sessions. The system evaluates students' responses to instructor-posted questions in the chat using generative AI, assigns scores based on predefined criteria, and displays a gamified leaderboard to foster competition and collaboration among peers. The study presents the system along with a detailed explanation of its development and a comprehensive evaluation of its effectiveness through a comparative analysis of virtual sessions with and without the system. By merging these innovative technologies, the system aims to enhance the frequency and quality of student participation, thereby creating a more dynamic and engaging learning experience.

The remainder of this paper is structured as follows: Section 2 provides a literature review, followed by Section 3, which discusses related work. Section 4 presents the system architecture and functionalities of VClass Engager, while Section 5 outlines its implementation. Section 6 details the evaluation methodology, and Section 7 presents the results. Section 8 offers a discussion of the findings, and Section 9 concludes the paper with key findings and directions for future research.

2. Literature Review

2.1. Synchronous Online Learning

Synchronous online learning refers to a learning experience that occurs in real-time through video conferencing platforms, enabling direct interaction between students and instructors [8]. This mode of delivery gained increasing adoption due to the rapid transition to online learning caused by the COVID-19 pandemic [9,10]. Synchronous online sessions attempt to replicate the immediacy and interactions of traditional face-to-face classrooms, enabling real-time communication, active participation, and immediate feedback [1].

Several platforms support the delivery of synchronous online learning, including Zoom, Microsoft Teams, and Blackboard Collaborate. These platforms provide a range of interactive features, including real-time chats to facilitate immediate communication, live polling for rapid assessments, group discussions to promote collaborative work, and screen sharing to support effective content delivery. These functionalities aim to provide an engaging learning environment, promote active student participation, and allow instructors to deliver content more effectively [4].

2.2. Student Engagement in Online Learning

Student engagement is a central component of successful learning. Kuh [11] defines engagement as the time and effort students devote to educationally purposeful activities and how institutions support such involvement. Astin's theory of Student Involvement emphasizes that student learning and development are directly related to the quality and

quantity of physical and psychological energy students invest in academic and social activities [12]. In other words, the more actively students are involved, the greater their learning and personal growth. Fredricks et al. [13] categorized engagement into three dimensions: behavioral, emotional, and cognitive. Behavioral engagement refers to students' participation in academic and social activities, such as attending classes and completing assignments. Emotional engagement involves students' feelings of interest, belonging, and connection to the learning environment. Cognitive engagement reflects the investment in learning and the willingness to invest mental effort to understand complex ideas. In the context of online education, engagement not only reflects students' interaction with content but also their motivation, sense of connection, and active participation in learning activities. Several learning theories provide a foundation for understanding and fostering engagement in virtual settings. Constructivist learning theory, for instance, stresses the importance of active, learner-centered experiences in constructing knowledge [14,15]. According to this theory, students engage more deeply when they can collaborate, receive feedback, and apply knowledge in meaningful contexts. Self-Determination Theory further explains that students are more likely to be engaged when their basic psychological needs for autonomy, competence, and relatedness are supported [16].

Despite the variety of features offered by synchronous online learning platforms, maintaining consistent student engagement remains a critical challenge. Extended screen time during online sessions often leads to screen fatigue, reducing students' attention span, cognitive function, and active participation over time [17]. Furthermore, the remote nature of synchronous online learning limits access to non-verbal cues, such as body language and subtle gestures, which are essential for fostering meaningful connections and rich social interaction [18]. Privacy concerns regarding camera use and session recordings further restrain active engagement, as some students may feel uncomfortable showing their home environments [19]. Additionally, external distractions are common at home and often affect concentration and disrupt the flow of discussions or group activities. Managing large numbers of students in synchronous online sessions presents additional challenges, including difficulties in providing individualized feedback, maintaining interactive discussions, and ensuring adequate participation, all of which can negatively impact engagement and learning outcomes [20].

Effective engagement in synchronous online sessions is critical for achieving desired learning outcomes. Strategies that promote real-time interaction, provide personalized feedback, and foster a sense of belonging are particularly important in sustaining student participation and motivation [21]. Therefore, it is vital to adopt innovative approaches and integrate appropriate technologies to address engagement-related challenges and improve the quality and effectiveness of synchronous online learning experiences.

2.3. Formative Assessment in Virtual Classrooms

Formative assessment plays a vital role in supporting students' learning by providing continuous insights into their progress and understanding. Unlike summative assessments, which evaluate performance at the end of a learning period, formative assessments are used throughout the instructional process to inform both teaching and learning in real time [22]. They help instructors to identify misconceptions, track students' progress toward achieving learning outcomes, and adjust instructional strategies to better meet students' needs [23]. For students, formative assessment fosters self-awareness of their learning process, encourages reflection, and promotes the development of metacognitive skills. It also supports motivation by highlighting progress and providing timely, constructive feedback [24]. According to Taylor et al. [25], when implemented effectively, formative assessment can create a feedback-rich environment where learning becomes more per-

sonalized, adaptive, and learner-centered, which are key attributes for success in both face-to-face and online learning contexts.

In synchronous online platforms, a variety of tools can be used to implement formative assessments. These include live polling tools (e.g., Blackboard Collaborate Polls, Mentimeter), chat-based questions, interactive whiteboards, and short quizzes embedded within the virtual session. Such tools are designed to allow instructors to assess student comprehension, promote participation, and make timely pedagogical decisions. They are also intended to enable instructors to adapt instruction in real time based on students' responses, thereby enhancing instructional responsiveness and learning outcomes [26].

Despite their potential advantages, several challenges limit the effectiveness of existing formative assessment tools in virtual classrooms. Tools like polls and short quizzes often assess only surface-level understanding, making it difficult for instructors to gauge deeper conceptual learning or critical thinking [27]. Similarly, while interactive whiteboards support visual expression and collaboration, they typically allow only one or two students to contribute at a time, and require digital fluency and time, which can hinder real-time, equitable participation and quick feedback [28]. In contrast, text-based chat offers a more inclusive and scalable alternative. It allows all students to respond simultaneously, providing a broader and richer overview of student understanding in real time. However, a key limitation of using text chat for formative assessment is the challenge of delivering timely and individualized feedback, particularly in large classes, where instructors may struggle to evaluate multiple student responses efficiently during live instruction [18].

To address this challenge, the current study proposes a technology-driven solution that enhances the implementation of formative assessment within chat-based interactions in virtual classrooms. By leveraging generative AI for real-time analysis and feedback, the system aims to support scalable, personalized formative assessment in synchronous online learning environments.

2.4. Generative AI in Education

Machine learning is a core branch of artificial intelligence. It enables systems to learn from data and make predictions or decisions without explicit programming [29]. A subfield of machine learning is generative AI that refers specifically to models such as variational autoencoders, generative adversarial networks (GANs), and transformer-based language models that can produce new content (e.g., text, images, video) by learning patterns from existing data [30]. These models have advanced rapidly due to breakthroughs in deep learning and the availability of large-scale datasets, and they are now reshaping a wide range of industries from creative design to scientific research, with profound impacts on everyday life and work [31].

Generative AI includes various models and tools that are designed for specific content creation tasks. Transformer-based models like GPT-4 generate human-like text, while tools such as DALL·E, Midjourney, and Stable Diffusion focus on image generation [32]. More recent systems like Runway and Sora extend these capabilities to video creation, enabling text-to-video synthesis [33]. Built on top of these models, user-facing platforms such as ChatGPT, Gemini, Claude, and Perplexity provide intuitive interfaces that allow non-experts to engage in complex tasks, including writing assistance, summarization, code generation, and research support, significantly broadening the accessibility and usability of generative AI [34].

In education, generative AI has proven useful across a variety of tasks. It is used, for instance, to generate automated feedback on student assignments, assist in grading open-ended responses, support the development of instructional materials, and deliver personalized tutoring to students [35]. These applications promote personalization by

allowing learners to receive tailored explanations, examples, and learning paths that align with their individual needs. Additionally, AI tools enhance scalability by supporting large numbers of students simultaneously and provide immediacy by delivering real-time responses and guidance. For instructors, this translates into more efficient feedback cycles and greater opportunities for adaptive teaching [36].

Despite its potential, the use of generative AI in education presents several challenges. AI-generated content may contain inaccuracies or biases due to the nature of the data on which the models were trained [37]. There is also the risk of over-reliance, where students might substitute genuine learning efforts with AI-generated responses, potentially undermining critical thinking and academic integrity. Other concerns include the opacity of model decision-making, data privacy risks, and the environmental costs associated with training large-scale models [35]. These issues highlight the need for thoughtful integration of generative AI into educational practice to ensure ethical use, transparency, and the preservation of human-centered learning values.

2.5. Gamification in Online Learning

Gamification refers to the use of game design elements such as points, badges, leaderboards, and challenges in non-game contexts to increase motivation and engagement [38]. These elements have been integrated into educational platforms to encourage participation, reinforce desired behaviors, and provide immediate feedback. For example, students may earn points for completing tasks, receive badges for achieving milestones, or compete on leaderboards based on their performance [39].

A large body of literature suggests that gamification can enhance student motivation, engagement, and participation, particularly in virtual learning environments where traditional classroom dynamics are absent. By introducing elements of competition and achievement, gamified learning environments can foster greater interest in course content and promote consistent involvement [40,41]. In higher education, studies have shown that gamification can improve student performance, increase time-on-task, and encourage collaborative learning behaviors [42].

However, despite its benefits, gamification also has limitations. Some studies have found that the effectiveness of gamification may diminish over time if rewards become expected or lose novelty [43]. Poorly designed gamified systems can also lead to superficial engagement where students focus on earning rewards rather than meaningful learning. Overreliance on external rewards can also undermine inner motivation, a phenomenon known as the overjustification effect, leading to a decline in genuine engagement over time [44]. In addition, pointsification can occur when learners focus solely on accumulating points rather than achieving meaningful learning outcomes, thereby diminishing the intended educational purpose of gamification [45]. Additionally, not all students respond positively to competition, and leaderboards may create anxiety or discourage lower-performing students [46]. These concerns highlight the importance of thoughtful design and alignment with pedagogical goals to ensure that gamification supports, not replaces, sound instructional practices.

3. Related Work

As evidence of the educational benefits of generative AI and gamification continues to accumulate, interest in their integration to enhance student learning has grown substantially. Various tools and systems have been developed and evaluated in academic contexts, aiming to enhance engagement, feedback quality, and instructional personalization across different learning environments.

Recent developments have focused on using generative AI to support instructors in designing and delivering instructional materials. Dickey and Bejarano [47], for instance, developed a tool called GAIDE (Generative AI for Instructional Development) to support instructors in automatically generating quizzes, lesson plans, and assignments using ChatGPT. Evaluations of GAIDE indicated a notable reduction in lesson preparation time while maintaining instructional quality. Likewise, Pesovski et al. [48] developed a generative AI tool and integrated it within an existing learning management system. The tool automatically generates multiple versions of learning materials based on instructor-defined learning outcomes. Evaluation results showed that students found the diversity of materials engaging and motivating.

Moreover, generative AI has also been utilized to improve assessment and feedback. Becerra et al. [49] introduced an innovative generative AI-based tool designed to support MOOC learners by helping them monitor their progress and reduce dropout rates. The tool utilizes generative AI to analyze anonymized learner data such as course progression, assignment performance, time spent on content, and activity time logs to deliver personalized feedback and actionable guidance. Guo et al. [50] also developed a feedback system that leverages generative AI and a multi-agent architecture to automatically generate and refine feedback on student responses in a science education course. The tool was evaluated with middle and high school students and demonstrated improvements in learning outcomes and higher levels of satisfaction with the feedback received. Muniyandi et al. [51] proposed a personalized course recommendation system for e-learning using a generative AI model enhanced by an Improved Artificial Bee Colony Optimization (IABCO) algorithm. The system acts as an agent within a reinforcement-based learning framework and demonstrates high performance, achieving an average success rate of 86.5% and an accuracy of 95.6% in recommending suitable courses to learners with varying levels of maturity.

Researchers have also combined generative AI with Intelligent Tutoring Systems (ITS) to deliver smarter, more adaptable feedback. Li et al. [52] proposed Sim-GAIL, a generative imitation learning method for student modeling in ITS, which helps the system better predict student behavior and provide personalized support. Liu et al. [53] developed a GPT-4–based ITS that leverages natural language processing and deep learning to support Socratic-style, conversational learning, in which the system engages students through dialogue, generates dynamic learning content, and provides personalized feedback in real time. Balakrishnan et al. [54] proposed a Generative AI-powered ITS that integrates brain-signal analysis with ChatGPT and Bard to detect student confusion during lectures and provide instant feedback. Their proposed system combines EEG-based distress detection with real-time AI explanations to improve learners' understanding and identify all possible points of confusion.

Generative AI has also been applied to personalize learning pathways and provide informed recommendations tailored to individual learner needs. For example, Chen et al. [55] developed a scalable, personalized tutoring tool called GPTutor, which leverages large language models to offer students customized explanations and learning support based on their individual responses. Similarly, Wang et al. [56] proposed a generative recommendation system that utilizes transformer-based models to deliver next-generation personalized course suggestions by generating item descriptions aligned with learners' interests.

Recent research has also explored the use of generative AI to enhance student engagement in online learning. Lin et al. [57] conducted a case study on the integration of ChatGPT into asynchronous online discussions using the Canvas learning management system. They analyzed student log data and collected learner perceptions to assess ChatGPT's effect on interaction and participation. Their findings revealed a marked increase in discussion

activity and depth when ChatGPT use was encouraged. Dai et al. [58] developed an online project-based learning (PBL) platform that incorporates generative AI to enhance learning interactivity and engagement. They conducted a year-long experimental study analyzing learners' cognitive processes, learning strategies, and performance across multiple tasks. The findings revealed that generative AI significantly improved students' learning approaches, cognitive engagement, and overall effectiveness. Zhang et al. [59] investigated the effects of ChatGPT-based human–computer dialogic interaction on student engagement in a programming course. In a quasi-experimental study involving 109 students, participants completed programming tasks either through ChatGPT-facilitated dialogic interaction or traditional pair programming. The results showed that ChatGPT-driven activities substantially increased students' behavioral, cognitive, and emotional engagement, reducing off-task behaviors and enhancing higher-order cognitive skills. Kartika [60] examined the use of generative AI within gamified language learning applications, where AI-generated conversational content and quizzes were combined with game-like reward mechanisms such as points, badges, and progress tracking. The study found improvements in students' motivation, vocabulary acquisition, and engagement through these AI-driven, reward-based interactions.

While generative AI has shown promise in enhancing personalization and instructional efficiency, gamification has also gained widespread adoption as a strategy to increase student motivation, engagement, and participation in online learning environments. A large number of studies have explored gamified tools that incorporate game-like elements such as points, leaderboards, badges, and challenges into the learning process to foster a more engaging and motivating learning experience. Aibar-Almazán et al. [61] investigated the use of Kahoot, a game-based learning platform that enables instructors to create quizzes and interactive competitions, in physiotherapy courses. They found that 30 to 60-min weekly gameplay sessions significantly enhanced students' attention, creativity, and critical thinking compared to non-gamified instruction. Bucchiarone et al. [62] introduced PolyGloT, an adaptive e-tutoring system that leverages gamified elements to support neurodiverse learners. Their initial evaluations revealed increased engagement and learning effectiveness. Song et al. [63] presented LearningverseVR, an immersive game-based learning environment that integrates generative AI and virtual reality. The platform uses Unity to create a 3D virtual environment and employs generative AI to enable learners to have open-ended conversations with virtual characters. This allows learners to engage in interactive, personality-rich virtual scenarios that enhance engagement through realistic immersion. Similarly, Zhao and McClure [64] introduced Gather.Town, a 2D avatar-based virtual space, as a gamified video-conferencing tool that promoted natural interaction and online engagement among language learners.

While generative AI and gamification have each demonstrated benefits in education, their combined potential to address real-time engagement challenges in synchronous online classrooms remains underexplored. To date, no system has integrated these approaches to enhance live participation and motivation in synchronous virtual classrooms. This gap highlights the need for innovative solutions that unify generative AI and gamified strategies to improve engagement in synchronous learning environments.

4. Theoretical Foundation and System Architecture

4.1. Theoretical Framework

The design of VClass Engager is grounded in two theoretical frameworks: active learning and gamification. These two frameworks together explain how the system would foster engagement, motivation, and meaningful participation during synchronous online sessions. Active Learning theory, as has been explained earlier, emphasizes that students learn more

effectively when they actively construct knowledge through interaction, reflection, and feedback rather than passively receiving information [65]. Within synchronous online sessions, these principles are applicable through short, structured engagement pauses that encourage students to respond, reflect, and receive immediate feedback. By prompting learners to articulate understanding at regular intervals, VClass Engager transforms the lecture into a sequence of interactive learning moments consistent with constructivist principles [66].

Gamification, in contrast, complements active learning by introducing motivational design elements that sustain learners' interest and participation [41]. Within VClass Engager, gamification transforms traditional formative assessment into an engaging experience by integrating features such as points, leaderboards, and real-time feedback. These elements encourage students to participate actively, compare their performance with peers, and strive for improvement throughout the session. Integrating these two frameworks provides a coherent pedagogical and motivational basis for the system's architecture. Active learning shapes the formative questioning and reflection cycles, while gamification sustains continuous participation and motivation. Together, they ensure that VClass Engager's functionalities, including real-time evaluation, feedback generation, and competitive ranking, are theoretically grounded in evidence-based models of engagement and learning.

4.2. System Architecture

VClass Engager was designed as an experimental tool comprising three main components: the Formative Assessment Manager, the AI Evaluation Engine, and the Database (see Figure 1). This architecture provides a scalable foundation for future enhancements, including expanded AI capabilities and integration with multiple video conferencing platforms.

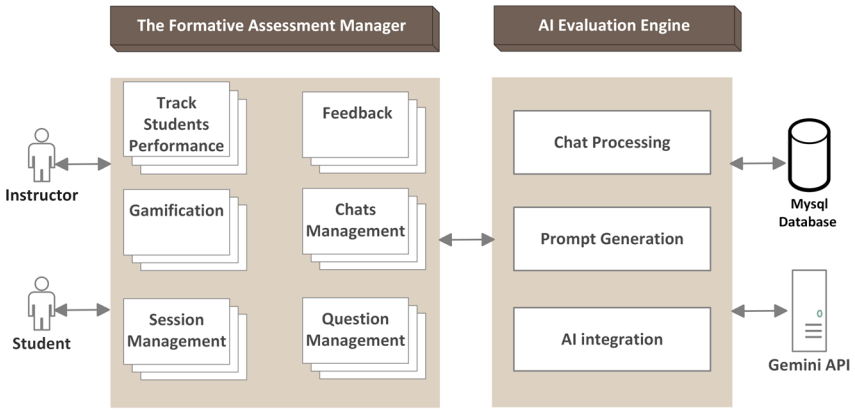


Figure 1. VClass Engager architecture.

The Formative Assessment Manager serves as the system's primary user interface for managing formative assessments and engagement activities. It allows instructors to create and organize synchronous sessions, add formative questions, define AI evaluation criteria, and submit chat transcripts for automated analysis. It also provides instructors with access to AI-generated feedback reports, while enabling students to review their responses, personalized feedback, and performance comparisons. The Formative Assessment Manager is responsible for handling all user interactions and coordinating the flow of information between end-users and the AI Evaluation Engine.

The AI Evaluation Engine is responsible for processing student responses and generating evaluation results. It receives instructor-defined questions, evaluation criteria, and extracted student responses from the Formative Assessment Manager. It constructs valid prompts compatible with the generative AI model, submits them for evaluation, and processes the AI-generated output. The Evaluation Engine computes student scores,

generates individualized feedback, detects copied answers, and produces summary reports to support instructor decision-making and gamified engagement.

The database securely stores all system data, including session records, assessment questions, evaluation criteria, student responses, AI-generated feedback, and gamification metrics. It ensures reliable data storage, integrity, and efficient retrieval for real-time performance tracking and system reporting.

4.3. Functional Requirements

The following functional requirements were identified to guide the development of VClass Engager. They are organized according to the main workflow stages: instructor setup, student participation, and AI-driven engagement analysis.

4.3.1. Instructor Setup and Session Management

1. The system should allow instructors to manage session details, including course information, date, and time of session.
2. Instructors should be able to create and manage a variety of formative questions specifying their type, text, and AI evaluation criteria. These supported question types and their instructional roles are described below.
3. Instructors should receive aggregated feedback reports summarizing student performance and suggested instructional actions.

4.3.2. Student Participation

1. The system must parse students' chat logs from synchronous sessions, automatically extracting questions, student names, and their responses.
2. Students should be able to review their answers, AI-generated feedback, and performance statistics.
3. Students should be able to view a dynamic leaderboard that displays their rankings based on their participation frequency and the quality of their responses.

4.3.3. AI-Driven Engagement Analysis

1. The system should generate valid, structured prompts for the generative AI model to support automated evaluation and feedback generation.
2. The system should evaluate student responses using generative AI according to instructor-defined criteria and generate personalized feedback.
3. A detection mechanism should flag copied responses and assign appropriate penalties.
4. The system should provide real-time analytics on the quality of answers and engagement levels during live sessions.

The supported formative question types in VClass Engager are based on evidence-based classroom assessment techniques that promote active learning at different stages of an online session [67–69]:

1. Concept Questions assess students' understanding of key ideas and definitions. They are typically used at the start or middle of a lecture to verify comprehension after a major explanation.
2. Background Knowledge Probe identifies students' prior knowledge and potential misconceptions before new content is introduced. It is best used at the beginning of a session or before a new unit.
3. Application Cards prompt students to apply newly learned concepts to practical examples or real-world scenarios. These are most effective during the middle or near the end of a lecture.

4. Misconception/Preconception Check uncovers incorrect assumptions or reasoning patterns that may hinder future learning. This type is typically used after initial explanations or before transitioning to advanced content.
5. Muddiest Point asks students to indicate which concept or explanation they found most confusing. It is most suitable near the end of a session or after covering a major topic.
6. Problem Identification encourages higher-order reasoning by asking students to identify an issue, limitation, or potential improvement in a presented concept or process. It is best positioned toward the end of a learning segment.

When creating any of these question types, the instructor is prompted to select how student responses should be evaluated. The system offers four AI-powered assessment modes: (1) similarity to a correct answer, wherein the instructor provides a reference solution; (2) coverage of key themes, where the instructor specifies the essential concepts that a good response should address; (3) relevance to lecture content, which requires a brief description of the lecture context or learning focus; and (4) AI-driven correctness judgement, in which the system autonomously determines the quality of responses without requiring predefined input. The instructor selects the most suitable evaluation strategy depending on the instructional purpose, e.g., Concept Questions and Background Knowledge Probes commonly benefit from similarity-based assessment, Application Cards and Problem Identification are best suited to theme-coverage evaluation, Misconception Checks work effectively with AI correctness assessment, and Muddiest Point activities align well with relevance-based analysis. This flexible design ensures pedagogical alignment while preserving instructor control and AI adaptability.

4.4. Non-Functional Requirements

In addition to the functional requirements, a set of non-functional requirements was identified to ensure the effectiveness, usability, and reliability of VClass Engager:

1. Usefulness: The functions and features of VClass Engager should meaningfully support instructors in monitoring student engagement and providing timely, AI-powered feedback during synchronous online sessions.
2. Ease of Use: VClass Engager should be intuitive and easy to use for both instructors and students, minimizing the learning curve and facilitating seamless integration into online teaching practices.
3. Information Clarity: All information and feedback presented by VClass Engager, including the quality of student responses, scores, and AI-generated insights, should be displayed in a clear and understandable manner.
4. Responsiveness: System operations, including AI evaluation, feedback generation, and leaderboard updates, should be executed in real-time or near-real-time to maintain engagement and provide immediate feedback during live sessions.

5. System Implementation

5.1. The Formative Assessment Manager

The Formative Assessment Manager was implemented as a web-based interface using PHP 8.3, HTML5/CSS3, JavaScript (ECMAScript 2023), and the Bootstrap 5.3 framework. These technologies were utilized to ensure responsive design, cross-platform compatibility, and seamless user interactions for both instructors and students. The interface was designed to support real-time teaching workflows in synchronous online environments and is structured around multiple interactive modules:

5.1.1. Formative Question Editor

The Formative Question Editor enables instructors to compose and manage formative assessment questions in preparation for synchronous online sessions. Questions must be created prior to the live session to ensure smooth integration into the chat-based engagement workflow. As shown in Figure 2, the interface includes a dropdown list for question type, text input fields for question content, and configurable options for selecting AI evaluation criteria. Supported question types align with evidence-based formative strategies, such as misconception checks, application cards, and muddiest point prompts. Depending on the selected evaluation criterion, instructors are prompted to enter correct answers, key themes, or lecture topics to guide AI assessment. To add a question, an associated session must first be created, ensuring that each question is properly linked to its instructional context. All questions can be edited or deleted through a centralized table view, offering structured management and revision capabilities.

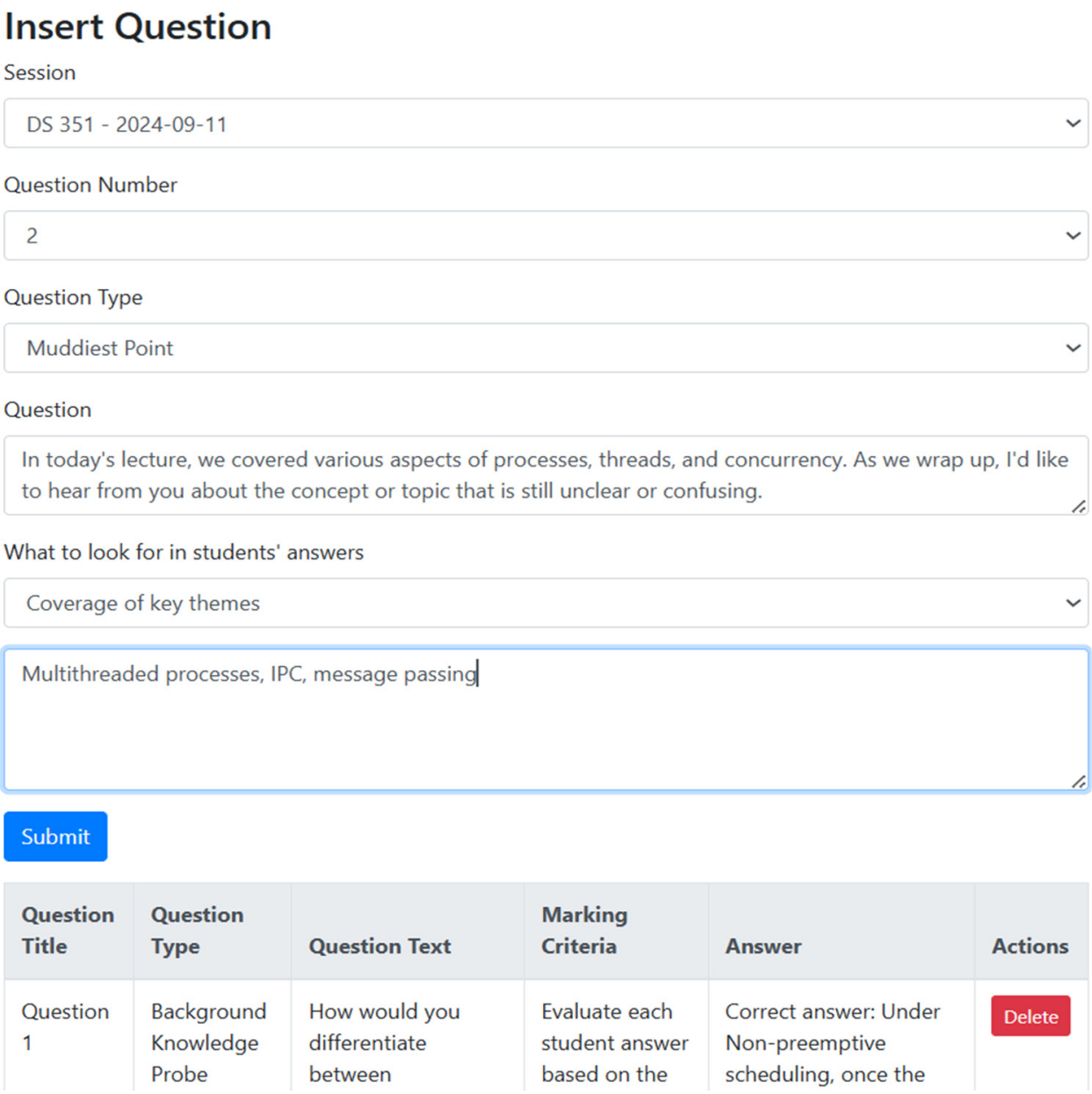


Figure 2. Formative question editor for creating, configuring, and managing formative questions.

5.1.2. Chat Analyzer

The Chat Analyzer enables instructors to process class interaction data by selecting a previously defined session and pasting the corresponding chat transcript copied from Blackboard Collaborate (see Figure 3). Due to the absence of a publicly available API for automated chat extraction, instructors are currently required to manually copy and paste the chat content, which is a straightforward process that typically takes only a few seconds.

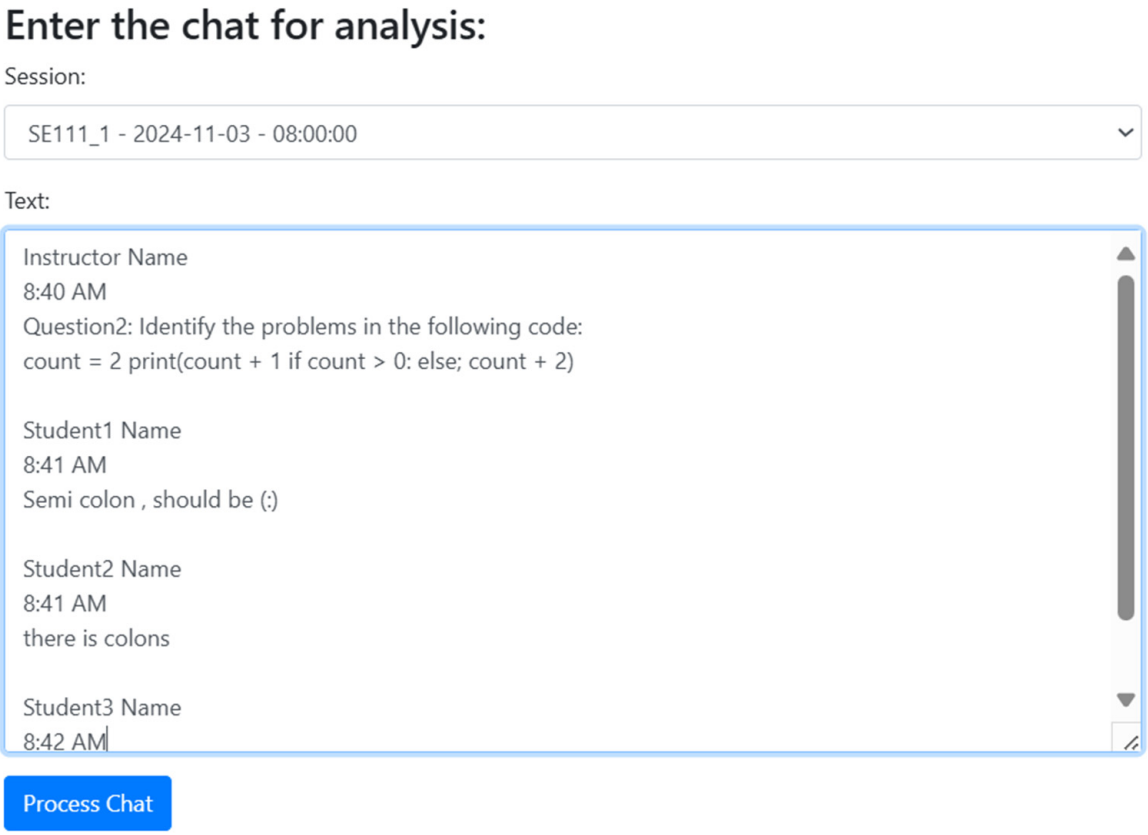


Figure 3. Chat analyzer used to extract and evaluate students' responses from the live session chat.

Once submitted, the transcript is passed to the AI Evaluation Engine, which parses the chat log to extract instructor-posted questions, student names, and associated responses. Although current support is limited to Blackboard Collaborate, the modular design of the parser allows for future adaptation to other video conferencing platforms such as Zoom or Microsoft Teams. The extracted data is then used by the AI Evaluation Engine to construct evaluation prompts and interact with the Gemini API (gemini-2.0-flash), returning both a score and personalized feedback for each student response. Instructors can review, modify, or approve the AI-generated evaluations.

5.1.3. Gamified Leaderboard

The leaderboard module presents a dynamically updating display of student rankings, calculated based on cumulative scores derived from AI-evaluated responses to formative questions throughout the session. This real-time visualization, as shown in Figure 4, integrates key gamification elements such as progress tracking and peer comparison to foster motivation and active participation. It is designed to be visible during synchronous sessions so it can serve as an instructional aid, enabling instructors to monitor individual engagement levels and identify students who may require additional support or encouragement.



Figure 4. Gamified leaderboard displaying students' rankings based on cumulative AI-evaluated responses.

5.1.4. Instructor Summary Feedback

This module enables instructors to see a synthesis of student responses to a selected formative question (see Figure 5). Upon selecting a session and question, the system retrieves the corresponding student answers and AI-generated scores. These are passed to the AI Evaluation Engine, which constructs a structured prompt and submits it to the Gemini API (gemini-2.0-flash) for analysis. The AI returns a summary report that identifies common strengths, misconceptions, and overall trends in student understanding. The resulting feedback is displayed to instructors in a concise format, along with instructional suggestions to support real-time pedagogical adjustments.

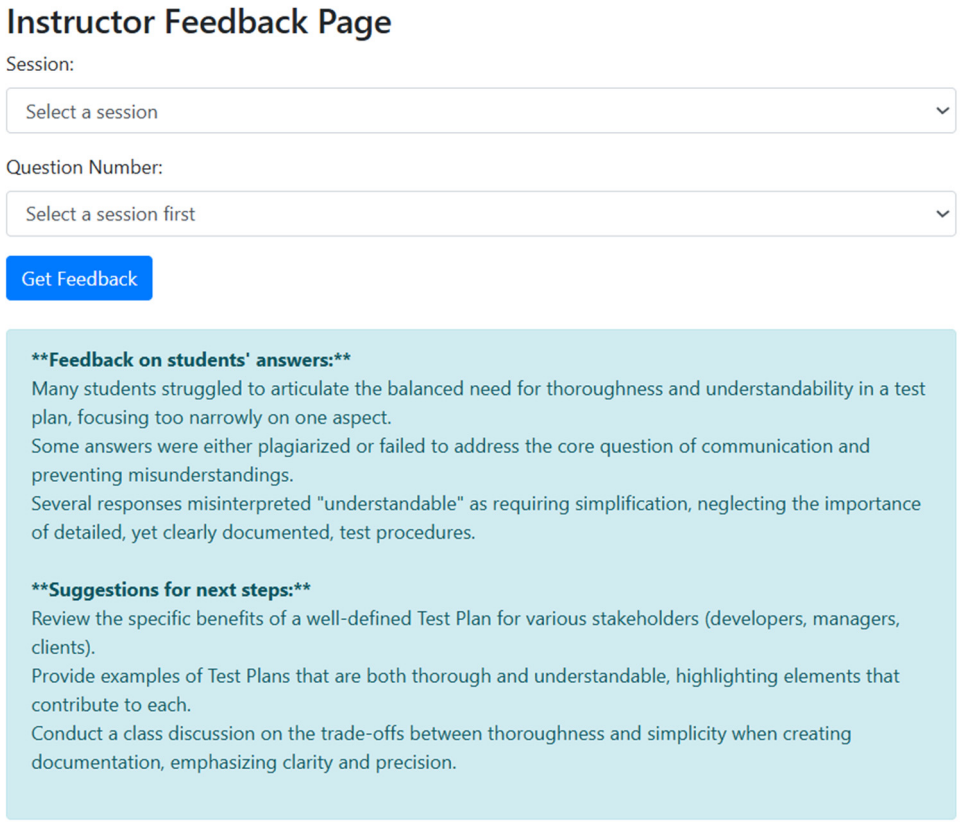


Figure 5. AI-generated summary providing feedback on students' answers and suggested instructional next steps.

5.1.5. Student Self-Review

To further enhance engagement and promote reflective learning, this module provides a dedicated self-review interface that students can access after each synchronous session. This interface allows students to revisit the formative questions they answered, view their submitted responses, and examine the AI-generated feedback tailored to their answers. For each question, the system presents the awarded score, along with comparative performance metrics such as the highest and median scores achieved by peers (see Figure 6). By offering transparency and context, the system enables students to better understand the quality of their responses, identify areas for improvement, and monitor their progress over time. This functionality encourages sustained attention during sessions and supports a growth-oriented learning experience.

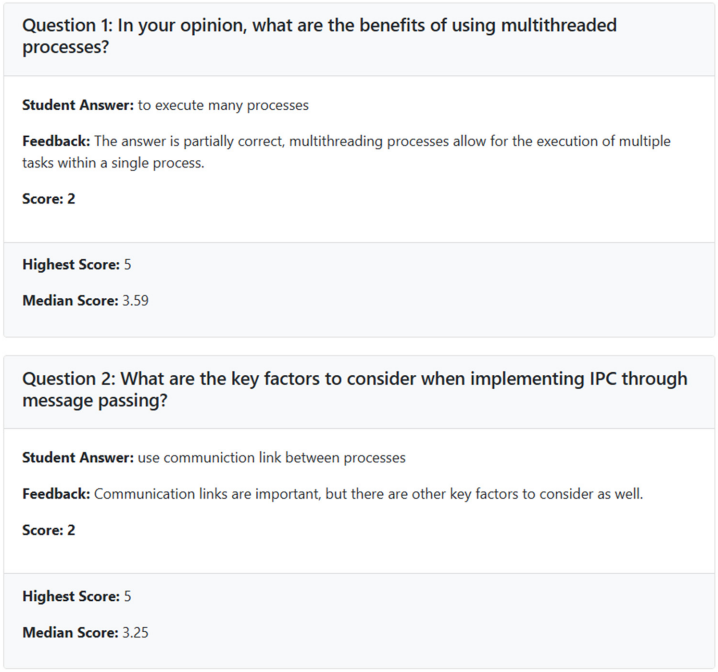


Figure 6. Student self-review screen displaying individual responses, AI feedback, and comparative performance insights.

5.2. The AI Evaluation Engine

The AI Evaluation Engine is the core computational component of VClass Engager, responsible for analyzing student responses and generating both individualized feedback and class-level instructional insights. Its primary objective is to automate the formative assessment process during synchronous online sessions by leveraging generative AI capabilities, thus reducing instructor workload and enabling real-time pedagogical adaptation.

The Evaluation Engine operates in two main phases: student-level evaluation and instructor-level synthesis.

5.2.1. Student-Level Evaluation

After the AI Evaluation Engine receives the session chat from the Formative Assessment Manager, it initiates the process described in Algorithm 1. The engine processes the chat transcript line by line, identifying each student's latest valid response while eliminating empty lines, parsing structured timestamps, and detecting repeated or copied content. Students who submit identical responses are flagged for duplication and automatically assigned a score of zero with an explanatory comment. For the unique and valid student responses, the AI Evaluation Engine retrieves the relevant marking criterion from the database, associated with the selected session and question. It then dynamically constructs

a structured prompt that includes the question, the students' answers, and the evaluation criteria. This prompt is submitted to the Gemini generative AI model, which returns both a numerical score and an individualized, explanation-based feedback message for each student. The system processes this output and attaches it to the corresponding student record, enabling instructors to review, approve, or adjust the AI-generated evaluations.

Algorithm 1. Student Response Assessment via Generative AI

```
Function EvaluateStudentResponses(chatText, sessionID):
responsesList ← [ ]
userName ← ""
currentQuestion ← ""
parsedLines ← SplitLines(chatText)
// Step 1: Parse chat line by line
For i from 0 to Length(parsedLines) − 1:
    line ← Trim(parsedLines[i])
    If line is empty:
      Continue
    If IsTimestamp(line):
      userName ← parsedLines[i − 1]
      chatEntry ← parsedLines[i + 1]
      // Step 2: Check if line contains instructor question
      If StartsWith(chatEntry, "Question"):
          currentQuestion ← chatEntry
      Else If IsStudent(userName):
          AddTo(responsesList, (userName, chatEntry))
    // Step 3: Keep only latest response per student
    finalResponses ← KeepLastResponsePerStudent(responsesList)
    // Step 4: Detect copied responses
    For each student in finalResponses:
      If IsCopied(student.response, finalResponses):
          student.score ← 0
          student.feedback ← "Copied response. No score assigned."
    // Step 5: Remove copied responses for evaluation
    filteredResponses ← FilterNonCopiedResponses(finalResponses)
    // Step 6: Retrieve evaluation criteria from DB
    criteria ← GetEvaluationCriteria(sessionID, currentQuestion)
    // Step 7: Prepare prompt for Gemini API
    prompt ← GeneratePrompt(currentQuestion, criteria, filteredResponses)
    //Step 8: Send prompt and process Gemini output
    geminiOutput ← CallGeminiAPI(prompt)
    For each result in geminiOutput:
      student ← result.name
      student.score ← result.score
      student.feedback ← result.feedback
    Return finalResponses
    processed_text ← reconstruct text from tokens
    append processed_text to new_dataset
end for
return new_dataset
```

5.2.2. Instructor-Level Synthesis

When the instructor requests summary feedback for a particular question, the AI Evaluation Engine performs an aggregate analysis of student responses. As described in Algorithm 2, The engine begins by retrieving all relevant responses and their associated scores for the selected session and question. Next, the engine constructs a structured prompt designed to guide the generative AI model (Gemini) to synthesize collective insights. This prompt includes the original question, the predefined evaluation criterion, and a curated list of student responses along with their scores. The AI is instructed to analyze recurring patterns, detect common strengths and misconceptions, and generate actionable instructional recommendations. The Gemini model returns a summary report that encapsulates the overall quality of student understanding, highlighting areas where learning objectives were met and where confusion persists. This synthesized feedback is then presented to the instructor, enabling data-informed instructional adjustments in real-time or during subsequent sessions.

Algorithm 2. Instructor-Level Summary Feedback Generation

```
Function GenerateInstructorFeedback(sessionID, questionID):
// Step 1: Retrieve student responses and scores from database
responses ← GetStudentResponsesWithScores(sessionID, questionID)
// Step 2: Retrieve question text and evaluation criterion
question ← GetQuestionText(questionID)
criterion ← GetEvaluationCriterion(sessionID, questionID)
// Step 3: Construct Gemini prompt
promptthe ← GenerateInstructorFeedbackPrompt(question, criterion, validResponses)
// Step 4: Submit prompt to Gemini API
summaryReport ← CallGeminiAPI(prompt)
// Step 5: Return structured summary for instructor
Return summaryReport
```

5.3. The Database

The backend of VClass Engager relies on a MySQL relational database to store, retrieve, and manage all required data. It maintains session records, formative question details, student responses, AI-generated feedback, and gamification metrics.

6. Evaluation Methodology

To assess the effectiveness of the VClass Engager system in enhancing student engagement and formative assessment practices during synchronous online sessions, a mixed-method evaluation approach was adopted. This section outlines the study's design, participant criteria, data collection procedures, and statistical methods used for analysis.

6.1. Target Group

The evaluation was conducted across four undergraduate courses offered by two Saudi universities, one public and one private. These courses were purposefully selected to represent varying academic levels and learning contexts: two were first-year introductory courses, one was a second-year intermediate course, and one was a third-year advanced course. This distribution ensured coverage of different cognitive demands and student experiences, allowing for a more comprehensive evaluation of the system's impact. The multi-institutional design aimed to capture diversity in learner profiles, instructional styles, and technical infrastructure.

Students enrolled in the selected courses were invited to participate. Eligibility was dependent upon their attendance in the relevant virtual sessions and their provision of informed consent. Ethical approval for the research was obtained from the Research Ethics Committee at Saudi Electronic University under approval number SEUREC-4521.

6.2. Research Design

A within-subject design was adopted to evaluate the system's impact on student engagement, ensuring that each participant serves as their own control. This approach enhances internal validity by minimizing inter-subject variability and increasing statistical power with a relatively smaller sample size [70]. It is widely regarded as an effective strategy for evaluating interventions in educational research, particularly when the intervention can be applied in a short time frame and across comparable learning conditions [71]. The study spanned two consecutive virtual sessions conducted via Blackboard Collaborate. Each session lasted approximately 1.5 h. The baseline and experimental sessions were conducted sequentially to avoid potential contamination or expectancy effects between conditions and to ensure that baseline participation reflected natural engagement before exposure to the AI-based intervention. Running the experimental session first could have caused students to adopt a competitive mindset from the gamification features or develop heightened expectations of receiving instant AI feedback. These effects would likely persist even after the tool was removed, preventing the baseline session from reflecting a true unassisted condition.

Both sessions covered equivalent content difficulty, followed the same instructional structure, and were delivered by the same instructor using identical materials. In both sessions, and in alignment with active learning principles, the instructor incorporated short engagement pauses approximately every 15–25 min. This timing was informed by prior literature suggesting that inserting pauses every 15 min [72] to 25 min [73] helps sustain learners' attention and engagement throughout longer sessions. At each pause, students were given 2–3 min to respond to a formative question (e.g., background knowledge probes, muddiest points) in the chat while the instructor temporarily paused delivery. These pauses were intentionally structured to promote reflection, application, and peer comparison. During these intervals, the instructor reviewed aggregated responses and feedback summaries, using them to identify misconceptions and facilitate targeted discussion immediately afterward.

6.2.1. Baseline Session (Without VClass Engager)

During the first session, the course instructor conducted a standard lecture, posing open-ended formative questions directly into the chat. Students responded in real time. Students were explicitly encouraged to participate. No AI evaluation or leaderboard features were employed. The frequency and quality of student responses were recorded manually to serve as a reference point for later comparison.

6.2.2. Experimental Session (With VClass Engager)

In the subsequent session, the VClass Engager system was integrated into the same instructional flow. The instructor employed a similar teaching strategy and posted comparable formative questions in the chat. Chat transcripts were entered into the VClass Engager system, which evaluated student responses using the built-in AI Evaluation Engine powered by Gemini. Gemini was selected based on recent peer-reviewed studies demonstrating its strong alignment with human expert evaluations in educational assessment contexts [74,75]. Scores and individualized feedback were generated based on pre-defined evaluation criteria. A live leaderboard displaying cumulative scores was presented throughout the session to promote motivation and peer-driven engagement. This

sequential design enabled a controlled comparison between the baseline and experimental conditions, while minimizing confounding variables such as changes in instructional approach or topic difficulty.

6.3. Data Collection and Instruments

Data were collected using three primary sources.

6.3.1. Chat Logs

During the baseline session, exported chat transcripts were used to capture student responses to the formative questions posed by the instructor. For each question, the number of distinct student responses was recorded to calculate participation rates. Additionally, the quality of each response was evaluated based on criteria established by the course instructor, aligned with the specific learning goal of the formative question. These instructor-defined criteria provided the basis for assessing changes in the quality of student responses during the experimental session.

6.3.2. AI Evaluation Output

In the experimental session, the number of distinct student responses to each question was recorded to calculate participation rates. In addition, the quality of student responses was evaluated using the AI Evaluation Engine integrated into the VClass Engager system. The engine generated scores and personalized feedback based on predefined evaluation criteria. To ensure reliability, the AI-generated scores were independently reviewed by the course instructor. This review process helped validate the accuracy and fairness of the automated evaluation.

6.3.3. Post-Session Survey

To complement the behavioral data with insights into student perceptions, a structured survey was administered immediately following the experimental session. The survey was distributed to all students who participated in both sessions and included a combination of 5-point Likert-scale items (e.g., ranging from "Strongly Disagree" to "Strongly Agree," or equivalent scale variations such as "Very Easy" to "Very Difficult") and open-ended questions. It aimed to assess students' perceptions of the system's usability, fairness, motivational impact, and overall contribution to their learning experience. The survey data were used to support and enrich the findings from chat logs and AI outputs, offering a better understanding of the system's educational value in real classroom settings.

6.4. Data Analysis

To evaluate the system's effectiveness, both descriptive and inferential statistical methods were applied. As the primary focus of the analysis was on student engagement, the participation rate was used as a key behavioral indicator. However, since participation alone does not fully reflect deeper cognitive or emotional dimensions, it was complemented by analyses of response quality and student perceptions collected through post-session surveys. This combined approach provides a more comprehensive understanding of engagement, encompassing behavioral (participation), cognitive (response quality), and affective (perceived motivation) dimensions of learner involvement.

6.4.1. Engagement Metrics

Student engagement was assessed based on the proportion of students who responded to each of three formative questions posed during each class session: one at the beginning (Q1), one in the middle (Q2), and one near the end (Q3). For each session, we calculated

the participation rate for each question by dividing the number of students who submitted an answer by the total number of students who attended that session, as follows:

$$Participation\ Rate_{ij} = \frac{R_{ij}}{N_j} \tag{1}$$

where

- R_{ij} is the number of students who responded to Question i in session j.
- N_j is the total number of students who attended session j.

To evaluate the effectiveness of the system, we calculated the change in participation rates from Q1 to Q2 ($\Delta Q2$) and from Q1 to Q3 ($\Delta Q3$) for each session:

$$\Delta Q2 = Participation\ Rate_{Q2} - Participation\ Rate_{Q1} \tag{2}$$

$$\Delta Q3 = Participation\ Rate_{Q3} - Participation\ Rate_{Q1}$$

These differences capture the increase or decrease in student engagement over the duration of the session. To determine whether the intervention led to a statistically significant increase in engagement, we performed paired t-tests comparing the $\Delta Q2$ and $\Delta Q3$ values between experimental and baseline sessions across the four courses, as follows:

$$t = \frac{\overline{d}}{s_d/\sqrt{n}} \tag{3}$$

where

- $\overline{d}$ is the mean of the difference scores ($\Delta Q2$ or $\Delta Q3$) across courses.
- s_d is the standard deviation of the difference scores.
- n is the number of paired observations (courses).

A p-value less than 0.05 was considered indicative of a statistically significant difference [76].

To complement the statistical significance analysis, Cohen's d was also calculated to evaluate the magnitude of the intervention's practical impact. It was computed using the following formulation for independent samples:

$$d = \frac{\overline{X}_{exp} - \overline{X}_{base}}{S_P} \tag{4}$$

where

- $\overline{X}_{\text{exp}}$ and $\overline{X}_{\text{base}}$ refer to the mean Δ engagement scores (e.g., $\Delta Q2$ or $\Delta Q3$) for the experimental (With Tool) and baseline (Without Tool) sessions, respectively.
- S_P is the pooled standard deviation across both sessions.

A Cohen's d value of approximately 0.2 was interpreted as a small effect, 0.5 as a medium effect, and 0.8 or above as a large practical effect, indicating substantial impact [77].

A post hoc power analysis was also conducted to ensure that the study had sufficient sensitivity to detect the observed effects. Statistical power was estimated based on the observed effect size (Cohen's d), the actual number of participants in each session, and a significance level of α = 0.05, using a two-tailed independent-samples t-test model. The statistical power $1 - \beta$ was computed using the following formulation:

$$1 - \beta = f(d, n1, n2, \alpha) \tag{5}$$

where

- d is the observed Cohen's d.
- $n1$ and $n2$ are the sample sizes in the baseline and experimental sessions.
- α is the Type I error rate.

Power values of 0.80 or higher are generally considered acceptable, indicating that the study has a sufficiently low probability of failing to detect a true effect (Type II error) [77].

6.4.2. Students' Responses Evaluation

Student responses were evaluated for quality across three formative questions presented during each class session (Q1, Q2, and Q3). As described earlier, each response was scored on a six-point ordinal scale ranging from 0 (very low quality) to 5 (high quality), based on the accuracy and completeness of the answer. Given the discrete, ordered nature of the scores, the bounded range between 0 and 5, and the nested data structure (students within courses), a Cumulative Link Mixed Model (CLMM) with a logit link function was used to assess the effect of the intervention [76,78,79].

The model included Session Type (Baseline or Experimental), Question Position (Q1, Q2, or Q3), and their interaction (Session Type × Question Position) as fixed effects. To account for the repeated measures' structure and potential dependency among students and courses, random intercepts for both student and course were included in the model [80]:

$$logit\left(P\left(Y_{ijk} \leq c\right)\right) = \theta c - \left(\beta 0 + \beta 1 \cdot SessionType_{ij} + \beta 2 \cdot Question_{ij} + \beta 3 \cdot (SessionType \times Question)_{ij} + u_k + v_i\right) \quad (6)$$

where

- Y_{ijk} is the ordinal score for student i in session j of course k.
- θc represents the threshold parameters between ordinal score categories.
- $\beta 0$ is the baseline log-odds for earning a higher score in Q1 of baseline sessions.
- $\beta 1$, $\beta 2$, $\beta 3$ represent the fixed effects of session type, question position, and their interaction, respectively.
- uk is the random intercept for course k.
- vi is the random intercept for student i.

This approach allowed us to examine whether the likelihood of higher-quality responses varied across the session and whether the AI-based gamified intervention moderated these changes over time. All effects were estimated on the log-odds scale, and statistical significance was evaluated at the $p < 0.05$ threshold. Post hoc comparisons of estimated marginal means were conducted using the Bonferroni adjustment to account for multiple comparisons [81]. All analyses were conducted using R (version 4.5.1), specifically the ordinal and emmeans packages.

6.4.3. Survey Analysis

Descriptive statistics were computed for each closed-ended item. Positive responses were defined based on top-scale selections (e.g., "Very Frequently," "Very Easy," "Strongly Agree," or "Excellent"), and response distributions were summarized to assess general trends. Open-ended responses were analyzed using thematic analysis, in which responses were read, coded, and grouped into key themes to identify recurring patterns.

7. Results

7.1. Participant Distribution

Table 1 presents the number of students who attended each of the two sessions, the baseline session (without VClass Engager) and the experimental session (with VClass Engager), across the four evaluated courses. For reporting purposes, the courses are anonymized as Course 1 through Course 4 and categorized by university type and academic level.

Table 1. Student attendance and formative question types used in baseline and experimental sessions across courses.

Course	University Type	Academic Level	Baseline Session Attendance	Experimental Session Attendance	Question Types Used (Q1, Q2, Q3)
Course 1	Private	First Year	15	17	Background Knowledge Probe, Misconception, Muddiest Point
Course 2	Public	First Year	7	6	Concept, Application Cards, Muddiest Point
Course 3	Private	Second Year	12	11	Concept, Problem Identification, Muddiest Point
Course 4	Public	Third Year	20	19	Concept, Application Cards, Muddiest Point
Total			**54**	**53**	

In total, 54 students attended the baseline sessions, while 53 students participated in the experimental sessions. Attendance remained relatively stable between the two types of sessions, which strengthens the validity of the within-subject comparison. There was also a relatively balanced distribution across institution types and academic levels, which ensures that the evaluation captures a range of learning contexts and student backgrounds.

To ensure fairness in the within-subject comparison, comparable formative question types were used at matching positions (Q1, Q2, Q3) in both the baseline and experimental sessions. These question types were intentionally selected based on their pedagogically appropriate timing.

7.2. Evaluation of Student Engagement

Although the intervention was implemented in four courses, the unit of statistical analysis was the individual student. Each of the 54 students contributed multiple responses across baseline and experimental sessions, yielding sufficient data points for paired and mixed-model analyses. Table 2 summarizes students' participation for each question (Q1, Q2, Q3) across baseline and experimental sessions for all four courses. On the other hand, Figure 7 illustrates the overall participation trends across questions (Q1 to Q3) for both session types, showing that participation increased throughout the session when VClass Engager was used, whereas it declined during baseline sessions.

Table 2. Summary of student participation for each question.

Course	Session	Total	Q1	Q2	Q3
Course 1	Baseline	15	4	5	5
	Experimental	17	12	15	17
Course 2	Baseline	7	4	4	3
	Experimental	6	3	6	6
Course 3	Baseline	12	7	4	6
	Experimental	11	6	9	9
Course 4	Baseline	20	17	8	6
	Experimental	19	12	14	10

Table 3 presents the participation rates for each question (Q1 Rate, Q2 Rate, Q3 Rate) and the corresponding changes in engagement (ΔQ2 and ΔQ3) across baseline and experimental sessions. In the experimental sessions where the AI-based intervention was applied, all courses showed a positive change in engagement from Q1 to both Q2 and Q3. In contrast, baseline sessions generally exhibited a decline or minimal change in participation.

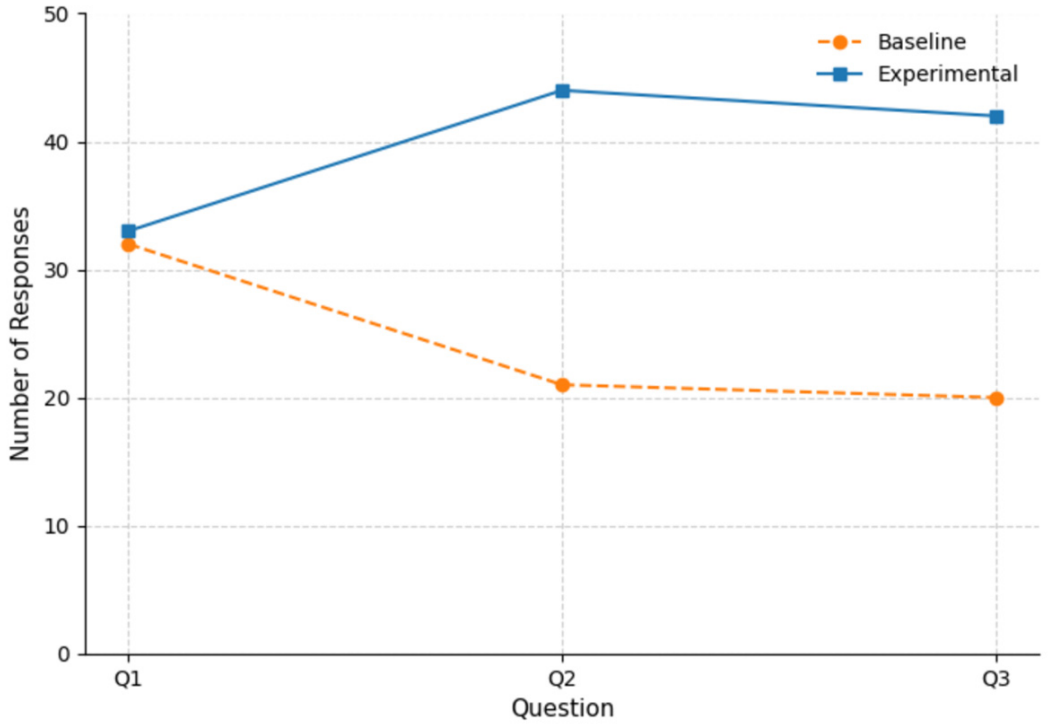


Figure 7. Overall student participation across questions (Q1 to Q3) in baseline and experimental sessions.

Table 3. Participation rates and engagement change per course.

Course	Session	Q1 Rate	Q2 Rate	ΔQ2	Q3 Rate	ΔQ3
Course 1	Baseline	0.267	0.333	0.067	0.333	0.067
	Experimental	0.706	0.882	0.176	1	0.294
Course 2	Baseline	0.571	0.571	0	0.429	−0.143
	Experimental	0.5	1	0.5	1	0.5
Course 3	Baseline	0.583	0.333	−0.250	0.5	−0.083
	Experimental	0.545	0.818	0.273	0.818	0.273
Course 4	Baseline	0.85	0.4	−0.450	0.3	−0.550
	Experimental	0.632	0.737	0.105	0.526	−0.105

Table 4 summarizes the overall differences in engagement changes between the two session types, while Figure 8 visually illustrates these differences. The average increase in participation from Q1 to Q2 (ΔQ2) was +0.26 in the experimental sessions, compared to a decrease of –0.16 in the baseline sessions. Similarly, the average increase from Q1 to Q3 (ΔQ3) was +0.24 for experimental sessions, while baseline sessions showed a mean decline of –0.18. Paired *t*-tests showed that the increases in both ΔQ2 and ΔQ3 were statistically significant between the experimental and baseline sessions, with *p*-values of 0.027 and 0.017, respectively.

Table 4. The overall differences in engagement changes between the two session types.

Metric	Experimental	Baseline
ΔQ2 Mean	0.264	−0.158
ΔQ3 Mean	0.24	−0.177
p-value(ΔQ2)	0.027	
p-value(ΔQ3)	0.017	

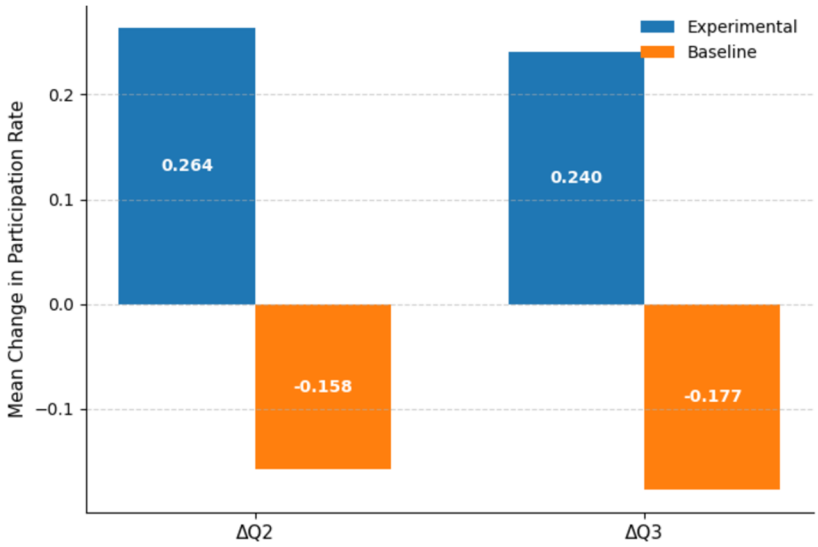


Figure 8. Mean changes in participation rates (ΔQ2 and ΔQ3) for baseline and experimental sessions.

Furthermore, as shown in Table 5, the corresponding effect sizes were substantial, with Cohen's d = 0.95 for ΔQ2 and 0.84 for ΔQ3, indicating large practical effects. These results confirm that the intervention was not only statistically significant but also meaningfully effective in sustaining and enhancing student engagement throughout the live session compared to sessions conducted without the system. In addition, the post hoc power analysis demonstrated exceptionally high statistical power ($\geq$0.99 for both ΔQ2 and ΔQ3), thereby confirming that the sample size of 54 participants was sufficient to detect the observed effects with high confidence.

Table 5. Effect size and statistical power for engagement changes (ΔQ2 and ΔQ3).

Metric	Cohen's d	Power ($1 - \beta$)
ΔQ2	0.95	0.998
ΔQ3	0.84	0.990

7.3. Assessment of Student Response Quality

As shown in Table 6, the fixed effects estimates indicate no significant difference in the quality of student responses between session types at Question 1 or Question 2. However, there was a statistically significant improvement in scores for Question 3 in the experimental session compared to the baseline. Specifically, while the effect of Session Type on Question 1 was not statistically significant (estimate = 0.255, $p = 0.564$), and Question 2 also showed no difference (estimate = −0.263, $p = 0.595$), Question 3 displayed a significant interaction effect (estimate = 2.378, $p = 0.001$), indicating a meaningful improvement in quality of responses due to the AI-based intervention.

Table 6. Estimated effects of session type and question on quality of responses (log-odds).

Effect	Estimate	*p*-Value
Session Type: Experimental (Q1)	0.255	0.564
Question 2 (vs. Q1 in Baseline)	−0.263	0.595
Question 3 (vs. Q1 in Baseline)	−2.412	<0.001
Session Type × Question 2	0.435	0.5
Session Type × Question 3	2.378	0.001

These findings were further supported by the estimated marginal means (see Table 7), which summarize the quality of student responses at each question position by session type. At the beginning (Question 1) and middle (Question 2) of the class, there was no statistically significant difference in response quality between experimental (with VClass Engager) and baseline (without VClass Engager) sessions. However, at the end of the class (Question 3), students in experimental sessions provided significantly better-quality responses than those in baseline sessions (log-odds difference = 2.633, $p < 0.0001$).

Table 7. Estimated marginal means by session type and question (log-odds).

Question	Baseline	Experimental	Difference	*p*-Value
Question 1	0.478	0.732	+0.255	0.564
Question 2	0.215	0.904	+0.69	0.155
Question 3	−1.935	0.698	+2.633	<0.001

Within-session quality of responses trends, as can be seen in Table 8, show that in baseline sessions (without VClass Engager), quality of student responses showed a significant decline from the beginning (Q1) to the end (Q3) of the class (log-odds difference for Q1–Q3 = 2.4124, $p = 0.0001$). A significant drop also occurred from Q2 to Q3 (log-odds difference for Q2-Q3 = 2.1497, $p = 0.0025$). In experimental sessions (with VClass Engager), the quality of student responses remained stable throughout the class, with no significant changes observed from Q1 to Q2, Q1 to Q3, or Q2 to Q3 (all p-values > 0.05).

Table 8. Within-session quality of responses changes.

Baseline Sessions:		
Comparison	**Difference**	***p*-Value**
Q1 vs. Q2	+0.263	1
Q1 vs. Q3	+2.412	<0.001
Q2 vs. Q3	+2.15	0.003
Experimental Sessions:		
Comparison	**Difference**	***p*-value**
Q1 vs. Q2	−0.172	1.000
Q1 vs. Q3	+0.034	1.000
Q2 vs. Q3	+0.207	1.000

7.4. Student Perceptions of the System

To complement the engagement data, a post-session survey was conducted to evaluate students' perceptions of VClass Engager. A total of 42 students responded to the survey. As presented in Table 9 and Figure 9, 62% of students reported participating "Very Frequently" or "Frequently" in the chat during the session with VClass Engager, reflecting a high level of activity. The system's usability was rated positively by 80% of respondents, with "Very Easy" and "Easy" being the two most frequently selected options.

Regarding motivational features, 80% of students either agreed or strongly agreed that the leaderboard encouraged them to participate more actively. When asked about the quality of the AI-generated feedback, 62% rated it as "High Quality" or better, although a notable 33% responded neutrally. Similarly, 67% of students agreed or strongly agreed that the AI feedback helped them better understand the material. With respect to comparative engagement, 62% reported feeling "More Engaged" or "Much More Engaged" than in sessions without the system. Finally, 82% of students rated the overall quality of the session as "Excellent" or "Good." The mean scores (ranging from 3.77 to 4.26) confirm an

overall strongly positive perception of VClass Engager, particularly in terms of usability, session quality, and its ability to enhance motivation and engagement. The relatively low to moderate standard deviations further indicate consistent agreement among students, with only minor variation in how the AI feedback was experienced across individuals.

Table 9. Statistical summary of student responses to the VClass Engager evaluation survey.

Question	Top 1 Response	Top 2 Response	Mean	Standard Deviation
Q1: How often did you participate in the chat during the virtual session with the app?	Very frequently (30.8%)	Frequently (30.8%)	4.00	1.19
Q2: How would you rate the ease of use of the app?	Very easy (48.7%)	Easy (30.8%)	4.26	0.85
Q3: Did the leaderboard motivate you to participate more actively in the session?	Agree (43.6%)	Strongly agree (35.9%)	4.08	0.90
Q4: How do you rate the quality of feedback provided by the generative AI on your responses?	High quality (41.0%)	Neutral (33.3%)	3.77	0.84
Q5: Did the generative AI feedback help you understand the material better?	Agree (41.0%)	Strongly agree (25.6%)	3.82	0.94
Q6: Compared to sessions without the app, how engaged did you feel during this session?	More engaged (33.3%)	Much more engaged (28.2%)	4.03	1.11
Q7: How would you rate the overall quality of the virtual session with the app?	Excellent (43.6%)	Good (38.5%)	4.23	0.81

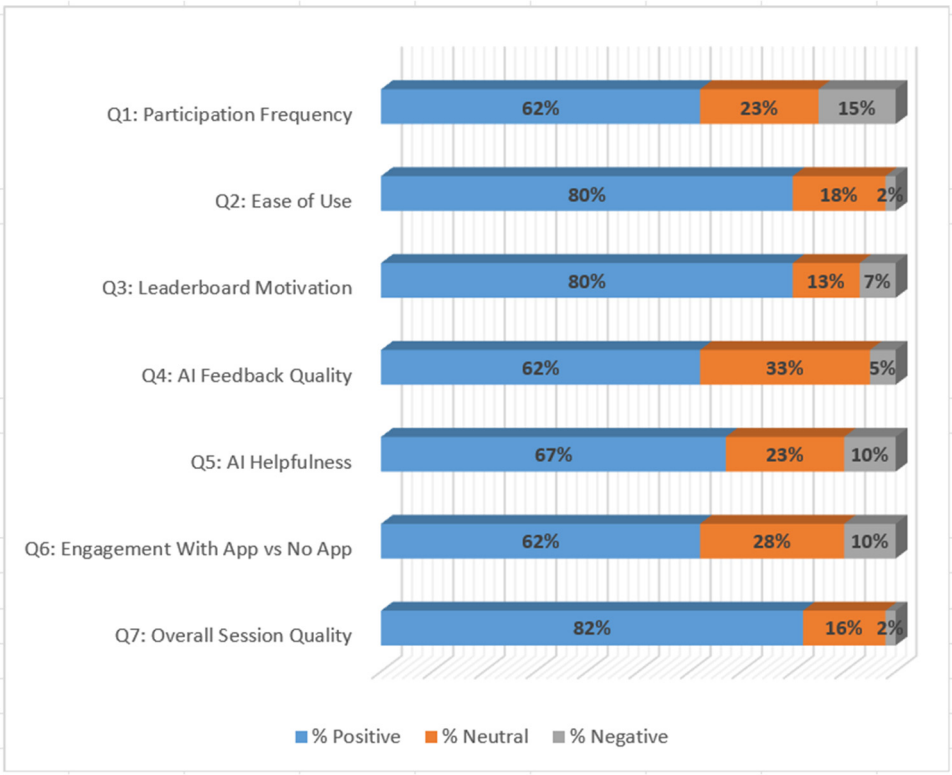


Figure 9. Distribution of student perceptions across survey questions (Positive, Neutral, and Negative Responses). Responses were originally collected on a five-point Likert scale (e.g., Strongly Agree → Strongly Disagree; Very Frequently → Never) but were grouped into three categories for visualization clarity: Positive (top two Likert levels), Neutral (middle level), and Negative (bottom two Likert levels).

Responses to the open-ended survey were also analyzed thematically. As shown in Table 10, the most frequently mentioned positive aspects of the system included gamification elements such as leaderboards and crowns, as well as the usefulness of AI-generated feedback and the ability to stay focused and engaged. The most common concerns centered on occasional vagueness in AI responses, the need to refresh or open external links to view feedback, and pressure related to public rankings. Suggestions for improvement include enhancing feedback accuracy, improving user interface design, and integrating the system more closely with the institutional platform.

Table 10. Summary of key themes from open-ended responses.

Question	Key Themes
What did you like most?	Gamification (leaderboard, crowns), useful feedback, engagement, focus, interaction
What did you like least?	Feedback issues, technical friction (refresh, link), pressure to participate, vague responses
Suggestions for improvement	Better feedback accuracy, platform integration, UI enhancements, and detection of AI-generated answers

8. Discussion

This study introduced and assessed the impact of a generative AI-based, gamified engagement system on student engagement and response quality in virtual classrooms. In terms of engagement, the analysis revealed a notable difference between baseline sessions (without VClass Engager) and experimental sessions (with the VClass Engager). In baseline sessions, student engagement declined significantly from the beginning (Question 1) to the end (Question 3) of the class, despite instructors' efforts to encourage participation. This pattern aligns with previous findings on attention drop-off in virtual learning environments [1,82]. Such a decline is often linked to screen fatigue, which makes it harder for students to concentrate and stay mentally engaged during long online sessions, as well as the lack of in-person cues that typically support active interaction and social connection in traditional classrooms [18].

In contrast, the experimental sessions showed a consistent increase in engagement throughout the class. All courses demonstrated higher participation in the second and third questions compared to the first, suggesting that VClass Engager not only sustained but actively enhanced student engagement as the session progressed. It seems that the gamified elements (e.g., leaderboard, scores) and real-time AI feedback may have maintained students' attention and motivation over time, preventing the typical participation drop observed in traditional sessions.

Regarding differences in response quality, students in the experimental sessions provided significantly higher-quality answers than those in baseline sessions by the end of the class (Question 3), as reflected in the statistically significant interaction effect. However, there were no significant differences in response quality between the two session types at the start (Question 1) or mid-point (Question 2) of the class. This lack of early difference may be due to the initial novelty of VClass Engager, which could require some time for students to become familiar with its features and adapt their behavior accordingly. It is also possible that the motivational effects of gamification accumulate gradually, becoming more impactful as the session progresses. This aligns with the findings of previous studies, which reported that the influence of game elements such as points, badges, and leaderboards tends to increase over time as users become more engaged with the system [83,84].

Consistent with this explanation, the quality of responses within the experimental sessions remained stable across the three questions, suggesting that once students engaged with VClass Engager, it helped sustain a consistent level of response quality throughout the session. In contrast, baseline sessions showed a marked decline in response quality from the beginning to the end of the class, highlighting the typical attention and engagement drop often seen in traditional virtual classes. This contrast reinforces the idea that the intervention was effective not only in improving the quality of end-of-class responses but also in sustaining cognitive engagement and mitigating the decline in response quality commonly observed over time.

The post-session survey results reinforce the observed positive impact of the VClass Engager on student engagement and provide new insights into student satisfaction. Students reported high levels of chat participation, ease of use, and overall satisfaction with VClass Engager. Notably, around 80% agreed that the leaderboard element encouraged more active participation, aligning with gamification literature that highlights the motivational benefits of competitive elements [85]. Similarly, a large majority rated the sessions as "Excellent" or "Good" compared to traditional sessions, suggesting a favorable shift in learner perceptions when the AI-enhanced system was in use.

However, findings on the perceived quality of AI-generated feedback were mixed. While 62% of students rated the feedback as "High" or "Very High," a significant minority (33%) provided neutral ratings, and a small number expressed dissatisfaction. This mixed sentiment is reflected in recent literature questioning the consistency, specificity, and pedagogical depth of AI-generated feedback [86]. Moreover, open-ended responses highlighted the need for clearer, more actionable feedback and raised concerns over occasional vagueness or technical glitches. A student commented, "I can't know what's wrong with my answer". These results suggest that while generative AI can be a valuable tool in maintaining engagement, it must be carefully refined to meet students' expectations for instructional support.

Another important observation relates to the motivational and cognitive impacts introduced by gamification. Students' comments about the leaderboard and scoring mechanism often reflected increased engagement and a sense of focus. However, a few students also mentioned feeling pressured to perform. One student noted "forced to participate" as what they liked the least about using VClass Engager. This aligns with ongoing debates about the potential impact of gamification in triggering performance anxiety or contributing to unequal participation [87]. This highlights the importance of balancing competitive mechanics with inclusive engagement strategies.

From an implementation perspective, VClass Engager was designed with scalability and practicality in mind. As a lightweight web-based system developed using widely supported technologies (PHP, HTML/CSS, JavaScript, and Bootstrap), VClass Engager can be deployed with minimal infrastructure requirements and integrated into existing LMS or video conferencing platforms without extensive technical overhead.

Despite the positive findings, several limitations must be acknowledged. First, the study involved a relatively small number of courses (n = 4) and student participants (n = 54), which may affect the generalizability of results. Although the analyses were conducted at the individual student level, providing sufficient data for reliable statistical modeling, the limited number of courses could still reflect localized instructional styles or student characteristics. Second, the study employed a sequential design in which the baseline session preceded the experimental session. Although both sessions covered comparable content and were matched in instructional delivery, the fixed order could introduce potential confounding factors, such as increased familiarity, topic variation, or novelty effects. Third, although the CLMM provided a robust method for analyzing ordinal outcomes, the study focused exclusively on short-term effects within single sessions and did not explore long-term patterns across multiple classes. As such, it remains unclear whether the observed benefits would persist over time or diminish due to novelty effects. Fourth, while the study captured observable indicators of engagement and quality of answers, it did not examine underlying cognitive or emotional dimensions, such as self-motivation, perceived autonomy, or a sense of belonging. These aspects are vital to sustained engagement and could offer a more comprehensive understanding of the VClass Engager's impact on virtual learning environments. Finally, the tool was not evaluated from the instructors' perspective, and certain features, such as the instructor summary feedback, were not assessed. This

leaves a gap in understanding how instructors perceive the tool's usability, effectiveness, and impact on instructional decision-making.

9. Conclusions

This study proposed and evaluated a novel experimental system, VClass Engager, designed to address the persistent challenges of student disengagement and passive participation in synchronous online sessions. The system integrates real-time generative AI feedback and game elements such as leaderboards and a scoring system to enhance interactivity, motivation, and focus during live virtual sessions. Unlike many existing approaches, this study contributes to the limited empirical research exploring how generative AI can be used as an instructional and engagement-support mechanism in synchronous online classrooms.

The effectiveness of VClass Engager was assessed using both quantitative and qualitative methods. Student engagement was measured by tracking participation across three in-class formative questions, while the quality of their responses was analyzed using a cumulative link mixed model (CLMM). In addition, a post-session survey captured students' perceptions of VClass Engager's usability, motivational features, and feedback quality. The results showed statistically significant improvements in engagement and quality of students' responses in the sessions that utilized VClass Engager. Survey responses further indicated high levels of satisfaction with VClass Engager usability and motivational aspects, though opinions on AI-generated feedback were more mixed.

Future work will proceed in several directions. First, long-term studies will examine the sustained impact of VClass Engager across multiple sessions, semesters, and institutions, to determine whether the observed benefits are maintained over time or influenced by novelty effects. These evaluations will incorporate larger and more diverse course samples with randomized or counterbalanced session designs to enhance generalizability, control for order effects, and ensure methodological robustness. Second, multi-condition experimental designs will be employed to differentiate the independent and synergistic effects of Generative AI and gamification on student engagement. Third, direct measures of learning impact, such as post-session comprehension tests or performance-based quiz scores, will be incorporated to determine whether increased engagement translates into measurable learning gains. Fourth, the influence of different formative question types and their timing within the session on both engagement and learning outcomes will be explored. Fifth, the AI evaluation mechanism will be validated against instructor ratings through formal inter-rater reliability analysis to establish the credibility and alignment of the automated scoring with expert judgment. Sixth, instructor interviews or surveys will be conducted to assess usability, pedagogical value, and the usefulness of features such as the instructor summary feedback, which was not covered in the current study. Seventh, plans are in place to integrate VClass Engager directly into widely used video conferencing platforms such as Zoom and Microsoft Teams. This integration will allow instructors and students to interact with VClass Engager simultaneously while attending virtual sessions, thereby creating a more cohesive and engaging learning experience within familiar digital environments. Finally, scalability to larger class sizes and the cost, maintenance, and deployment feasibility of the system in institutional settings will be investigated to determine its practicality for long-term adoption.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The author would like to express his sincere gratitude to the students and instructors from both participating universities who generously contributed their time and insights to this study.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
VClass Engager	Virtual Class Engagement Tool
CLMM	Cumulative Link Mixed Model
API	Application Programming Interface
IABCO	Improved Artificial Bee Colony Optimization
GANs	Generative Adversarial Networks
GAIDE	Generative AI for Instructional Development

References

1. Hollister, B.; Nair, P.; Hill-Lindsay, S.; Chukoskie, L. Engagement in Online Learning: Student Attitudes and Behavior During COVID-19. *Front. Educ.* **2022**, *7*, 851019. [CrossRef]
2. Buadu, Y.S.; Kear, K.; Donelan, H. Web Conferencing: The Technical and Social Challenges. *Proc. Int. Conf. Networked Learn.* **2024**, *14*, 1–4. [CrossRef]
3. Leenknecht, M.; Wijnia, L.; Köhlen, M.; Fryer, L.; Rikers, R.; Loyens, S. Formative assessment as practice: The role of students' motivation. *Assess. Eval. High. Educ.* **2021**, *46*, 236–255. [CrossRef]
4. Moorhouse, B.L.; Kohnke, L. Conducting formative assessment during synchronous online lessons: University teachers' challenges and pedagogical strategies. *Pedagog. Int. J.* **2023**, *18*, 431–449. [CrossRef]
5. Owoseni, A.; Kolade, O.; Egbetokun, A. Applications of Generative AI in Formative Learning and Assessment. In *Generative AI in Higher Education: Innovation Strategies for Teaching and Learning*; Owoseni, A., Kolade, O., Egbetokun, A., Eds.; Springer Nature: Cham, Switzerland, 2024; pp. 63–95.
6. Katona, J.; Gyonyoru, K.I.K. Integrating AI-based adaptive learning into the flipped classroom model to enhance engagement and learning outcomes. *Comput. Educ. Artif. Intell.* **2025**, *8*, 100392. [CrossRef]
7. Elsawah, W. Exploring the effectiveness of gamification in adult education: A learner-centric qualitative case study in a dubai training context. *Int. J. Educ. Res. Open* **2025**, *9*, 100465. [CrossRef]
8. Desai, V.; Lee, D.J.; Lin, R.J.; Alvinia, M.; MacNeill, H.; Chan, Y. Assessing Educator Burnout in Online Synchronous Teaching in Surgical Disciplines. *J. Surg. Educ.* **2024**, *81*, 1409–1417. [CrossRef] [PubMed]
9. Alammary, A.; Alshaikh, M.; Pratama, A.R. Awareness of security and privacy settings in video conferencing apps among faculty during the COVID-19 pandemic. *PeerJ Comput. Sci.* **2022**, *8*, e1021. [CrossRef]
10. Alammary, A.S. Blended Learning Delivery Methods for a Sustainable Learning Environment: A Delphi Study. *Sustainability* **2024**, *16*, 3269. [CrossRef]
11. Kuh, G.D. Assessing What Really Matters to Student Learning Inside The National Survey of Student Engagement. *Change Mag. High. Learn.* **2001**, *33*, 10–17. [CrossRef]
12. Astin, A.W. Student involvement: A developmental theory for higher education. *J. Coll. Stud. Dev.* **1999**, *40*, 518–529.
13. Fredricks, J.A.; Blumenfeld, P.C.; Paris, A.H. School Engagement: Potential of the Concept, State of the Evidence. *Rev. Educ. Res.* **2004**, *74*, 59–109. [CrossRef]
14. Piaget, J.; Cook, M. *The Origins of Intelligence in Children*; International Universities Press: New York, NY, USA, 1952; Volume 8.
15. Vygotsky, L.S.; Cole, M. *Mind in Society: Development of Higher Psychological Processes*; Harvard University Press: Cambridge, MA, USA, 1978; p. 159.
16. Ryan, R.M.; Deci, E.L. Self-Determination Theory. In *Encyclopedia of Quality of Life and Well-Being Research*; Maggino, F., Ed.; Springer International Publishing: Cham, Switzerland, 2023; pp. 6229–6235.
17. Bailenson, J.N. Nonverbal overload: A theoretical argument for the causes of Zoom fatigue. *Technol. Mind Behav.* **2021**, *2*, 1–6. [CrossRef]
18. Peper, E.; Wilson, V.; Martin, M.; Rosegard, E.; Harvey, R. Avoid Zoom fatigue, be present and learn. *NeuroRegulation* **2021**, *8*, 47–56. [CrossRef]
19. Castelli, F.R.; Sarvary, M.A. Why students do not turn on their video cameras during online classes and an equitable and inclusive plan to encourage them to do so. *Ecol. Evol.* **2021**, *11*, 3565–3576. [CrossRef]

20. Fabriz, S.; Mendzheritskaya, J.; Stehle, S. Impact of Synchronous and Asynchronous Settings of Online Teaching and Learning in Higher Education on Students' Learning Experience During COVID-19. *Front. Psychol.* **2021**, *12*, 733554. [CrossRef] [PubMed]
21. Alammary, A.S. Optimizing Components Selection in Blended Learning: Toward Sustainable Students Engagement and Success. *Sustainability* **2024**, *16*, 4923. [CrossRef]
22. Alammary, A.; Masoud, S. Towards Smarter Assessments: Enhancing Bloom's Taxonomy Classification with a Bayesian-Optimized Ensemble Model Using Deep Learning and TF-IDF Features. *Electronics* **2025**, *14*, 2312. [CrossRef]
23. Morris, R.; Perry, T.; Wardle, L. Formative assessment and feedback for learning in higher education: A systematic review. *Rev. Educ.* **2021**, *9*, e3292. [CrossRef]
24. Narciss, S.; Zumbach, J. Formative Assessment and Feedback Strategies. In *International Handbook of Psychology Learning and Teaching*; Zumbach, J., Bernstein, D., Narciss, S., Marsico, G., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 1–28.
25. Taylor, D.L.; Yeung, M.; Bashet, A.Z. Personalized and Adaptive Learning. In *Innovative Learning Environments in STEM Higher Education: Opportunities, Challenges, and Looking Forward*; Ryoo, J., Winkelmann, K., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 17–34.
26. Kohnke, L.; Moorhouse, B.L. Facilitating Synchronous Online Language Learning through Zoom. *RELC J.* **2022**, *53*, 296–301. [CrossRef]
27. Zhao, X. The impact of live polling quizzes on student engagement and performance in computer science lectures: A post-COVID19 study. In Proceedings of the 17th International Conference on Computer Supported Education, Porto, Portugal, 1–3 April 2025; Volume 1, pp. 291–298.
28. De Vita, M.; Verschaffel, L.; Elen, J. Interactive Whiteboards in Mathematics Teaching: A Literature Review. *Educ. Res. Int.* **2014**, *2014*, 401315. [CrossRef]
29. Alammary, A.S. Investigating the impact of pretraining corpora on the performance of Arabic BERT models. *J. Supercomput.* **2024**, *81*, 187. [CrossRef]
30. Sengar, S.S.; Hasan, A.B.; Kumar, S.; Carroll, F. Generative artificial intelligence: A systematic review and applications. *Multimed. Tools Appl.* **2024**, *84*, 23661–23700. [CrossRef]
31. Ooi, K.-B.; Wei-Han, T.G.; Mostafa, A.-E.; Al-Sharafi, M.A.; Alexandru, C.; Amrita, C.; Dwivedi, Y.K.; Tzu-Ling, H.; Kumar, K.A.; Voon-Hsien, L.; et al. The Potential of Generative Artificial Intelligence Across Disciplines: Perspectives and Future Directions. *J. Comput. Inf. Syst.* **2025**, *65*, 76–107. [CrossRef]
32. Sai, S.; Yashvardhan, U.; Chamola, V.; Sikdar, B. Generative AI for Cyber Security: Analyzing the Potential of ChatGPT, DALL-E, and Other Models for Enhancing the Security Space. *IEEE Access* **2024**, *12*, 53497–53516. [CrossRef]
33. Li, X.; Miao, Q.; Li, L.; Ni, Q.; Fan, L.; Wang, Y.; Tian, Y.; Wang, F.Y. Sora for Scenarios Engineering of Intelligent Vehicles: V&V, C&C, and Beyonds. *IEEE Trans. Intell. Veh.* **2024**, *9*, 3117–3122. [CrossRef]
34. De Vito, A.; Geremia, N.; Bavaro, D.F.; Seo, S.K.; Laracy, J.; Mazzitelli, M.; Marino, A.; Maraolo, A.E.; Russo, A.; Colpani, A.; et al. Comparing large language models for antibiotic prescribing in different clinical scenarios: Which performs better? *Clin. Microbiol. Infect.* **2025**, *31*, 1336–1342. [CrossRef]
35. Mittal, U.; Sai, S.; Chamola, V.; Sangwan, D. A Comprehensive Review on Generative AI for Education. *IEEE Access* **2024**, *12*, 142733–142759. [CrossRef]
36. Aperstein, Y.; Cohen, Y.; Apartsin, A. Generative AI-Based Platform for Deliberate Teaching Practice: A Review and a Suggested Framework. *Educ. Sci.* **2025**, *15*, 405. [CrossRef]
37. Atkins, C.; Girgente, G.; Shirzaei, M.; Kim, J. Generative AI tools can enhance climate literacy but must be checked for biases and inaccuracies. *Commun. Earth Environ.* **2024**, *5*, 226. [CrossRef]
38. Deterding, S.; Dixon, D.; Khaled, R.; Nacke, L. From game design elements to gamefulness: Defining "gamification". In Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, Tampere, Finland, 28–30 September 2011; pp. 9–15.
39. Ristianto, S.D.; Putri, A.; Rosmansyah, Y. Personalized Gamification a Technological Approach for Student Education: A Systematic Literature Review. *IEEE Access* **2025**, *13*, 55712–55726. [CrossRef]
40. Saleem, A.N.; Noori, N.M.; Ozdamli, F. Gamification Applications in E-learning: A Literature Review. *Technol. Knowl. Learn.* **2022**, *27*, 139–159. [CrossRef]
41. Kaufmann, D.A. Reflection: Benefits of gamification in online higher education. *J. Instr. Res.* **2018**, *7*, 125–132. [CrossRef]
42. Subhash, S.; Cudney, E.A. Gamified learning in higher education: A systematic review of the literature. *Comput. Hum. Behav.* **2018**, *87*, 192–206. [CrossRef]
43. Hanus, M.D.; Fox, J. Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Comput. Educ.* **2015**, *80*, 152–161. [CrossRef]
44. Dah, J.; Hussin, N.; Zaini, M.K.; Isaac Helda, L.; Senanu Ametefe, D.; Adozuka Aliu, A. Gamification is not Working: Why? *Games Cult.* **2025**, *20*, 934–957. [CrossRef]

45. Kim, S.; Song, K.; Lockee, B.; Burton, J. Legal and Ethical Issues. In *Gamification in Learning and Education: Enjoy Learning like Gaming*; Kim, S., Song, K., Lockee, B., Burton, J., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 109–116.
46. Baydas, O.; Cicek, M. The examination of the gamification process in undergraduate education: A scale development study. *Technol. Pedagog. Educ.* **2019**, *28*, 269–285. [CrossRef]
47. Dickey, E.; Bejarano, A. GAIDE: A Framework for Using Generative AI to Assist in Course Content Development. In Proceedings of the 2024 IEEE Frontiers in Education Conference (FIE), Washington, DC, USA, 13–16 October 2024; pp. 1–9.
48. Pesovski, I.; Santos, R.; Henriques, R.; Trajkovik, V. Generative AI for Customizable Learning Experiences. *Sustainability* **2024**, *16*, 3034. [CrossRef]
49. Becerra, Á.; Mohseni, Z.; Sanz, J.; Cobos, R. A Generative AI-Based Personalized Guidance Tool for Enhancing the Feedback to MOOC Learners. In Proceedings of the 2024 IEEE Global Engineering Education Conference (EDUCON), Kos Island, Greece, 8–11 May 2024; pp. 1–8.
50. Guo, S.; Latif, E.; Zhou, Y.; Huang, X.; Zhai, X. Using Generative AI and Multi-Agents to Provide Automatic Feedback. *arXiv* **2024**, arXiv:2411.07407. [CrossRef]
51. Muniyandi, A.P.; Balusamy, B.; Dhanaraj, R.K.; Ellappan, V.; Murali, S.; Sathyamoorthy, M.; Nkenyereye, L. Privacy Preserved Reinforcement Learning Model Using Generative AI for Personalized E-Learning. *IEEE Trans. Consum. Electron.* **2024**, *70*, 6157–6165. [CrossRef]
52. Li, Z.; Shi, L.; Wang, J.; Cristea, A.I.; Zhou, Y. Sim-GAIL: A generative adversarial imitation learning approach of student modelling for intelligent tutoring systems. *Neural Comput. Appl.* **2023**, *35*, 24369–24388. [CrossRef]
53. Liu, S.; Guo, X.; Hu, X.; Zhao, X. Advancing Generative Intelligent Tutoring Systems with GPT-4: Design, Evaluation, and a Modular Framework for Future Learning Platforms. *Electronics* **2024**, *13*, 4876. [CrossRef]
54. Balakrishnan, P.; Anny Leema, A.; Jothiaruna, N.; Arun Kumar, S. Intelligent Tutoring Systems Powered by Generative AI: Advancing Personalized Education and Overcoming Challenges. In Proceedings of the 2025 International Conference on Intelligent Systems and Computational Networks (ICISCN), Bidar, India, 24–25 January 2025; pp. 1–6.
55. Chen, E.; Lee, J.-E.; Lin, J.; Koedinger, K. GPTutor: Great Personalized Tutor with Large Language Models for Personalized Learning Content Generation. In Proceedings of the Eleventh ACM Conference on Learning @ Scale, Atlanta, GA, USA, 18–20 July 2024; pp. 539–541.
56. Wang, W.; Lin, X.; Feng, F.; He, X.; Chua, T.-S. Generative recommendation: Towards next-generation recommender paradigm. *arXiv* **2023**. [CrossRef]
57. Lin, X.; Luterbach, K.; Gregory, K.H.; Sconyers, S.E. A Case Study Investigating the Utilization of ChatGPT in Online Discussions. *Online Learn.* **2024**, *28*, 1–23. [CrossRef]
58. Dai, Y.; Xiao, J.-Y.; Huang, Y.; Zhai, X.; Wai, F.-C.; Zhang, M. How Generative AI Enables an Online Project-Based Learning Platform: An Applied Study of Learning Behavior Analysis in Undergraduate Students. *Appl. Sci.* **2025**, *15*, 2369. [CrossRef]
59. Zhang, L.; Jiang, Q.; Xiong, W.; Zhao, W. Effects of ChatGPT-Based Human–Computer Dialogic Interaction Programming Activities on Student Engagement. *J. Educ. Comput. Res.* **2025**, *63*, 988–1023. [CrossRef]
60. Kartika, S. AI and Gamification: Exploring the Integration of Generative AI in Interactive English Language Learning Apps. *Int. J. Lang. Humanit. Educ.* **2025**, *8*, 105–114. [CrossRef]
61. Aibar-Almazán, A.; Castellote-Caballero, Y.; Carcelén-Fraile, M.d.C.; Rivas-Campo, Y.; González-Martín, A.M. Gamification in the classroom: Kahoot! As a tool for university teaching innovation. *Front. Psychol.* **2024**, *15*, 1370084. [CrossRef]
62. Bucchiarone, A.; Martorella, T.; Frageri, D.; Colombo, D. PolyGloT: A personalized and gamified eTutoring system for learning modelling and programming skills. *Sci. Comput. Program.* **2024**, *231*, 103003. [CrossRef]
63. Song, Y.; Wu, K.; Ding, J. Developing an immersive game-based learning platform with generative artificial intelligence and virtual reality technologies—"LearningverseVR". *Comput. Educ. X Real.* **2024**, *4*, 100069. [CrossRef]
64. Zhao, X.; McClure, C.D. Gather.Town: A Gamification Tool to Promote Engagement and Establish Online Learning Communities for Language Learners. *RELC J.* **2024**, *55*, 240–245. [CrossRef]
65. Bonwell, C.C.; Eison, J.A. *Active Learning: Creating Excitement in the Classroom. 1991 ASHE-ERIC Higher Education Reports*; ERIC: Washington, DC, USA, 1991.
66. Vygotsky, L.S. *Mind in Society: The Development of Higher Psychological Processes*; Harvard University Press: London, UK, 1980; p. 175.
67. Angelo, T.A.; Zakrajsek, T.D. *Classroom Assessment Techniques: Formative Feedback Tools for College and University Teachers*; Wiley: Hoboken, NJ, USA, 2024.
68. Goldstein, G.S. Using Classroom Assessment Techniques in an Introductory Statistics Class. *Coll. Teach.* **2007**, *55*, 77–82. [CrossRef]
69. Barkley, E.F.; Major, C.H. *Learning Assessment Techniques: A Handbook for College Faculty*; Wiley: Hoboken, NJ, USA, 2016.
70. Gao, X.; De Hooge, I.E.; Fischer, A.R.H. Something underneath? Using a within-subjects design to examine schema congruity theory at an individual level. *J. Retail. Consum. Serv.* **2022**, *68*, 102994. [CrossRef]
71. Creswell, J.W.; Guetterman, T.C. *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*, 7th ed.; Pearson: Upper Saddle River, NJ, USA, 2024.

72. Bachhel, R.; Thaman, R.G. Effective use of pause procedure to enhance student engagement and learning. *J. Clin. Diagn. Res.* **2014**, *8*, XM01–XM03. [CrossRef] [PubMed]
73. The University of Melbourne. How Many Synchronous Sessions Should I Have in My Online Subject? Available online: https://www.unimelb.edu.au/tli/news/articles/how-many-synchronous-sessions-should-i-have-in-an-online-subject#:~:text=There%20is%20evidence%20that%20âĂŹZoom,brains%20are%20processing%20this%20information (accessed on 23 August 2025).
74. Benčina, K.; Nemet, L.D. Is Feedback In The Eye Of The Beholder? Chat Gpt Vs. Google Gemini in Giving Feedback on Students' Oral Presentation Scripts. *E-Methodology* **2024**, *11*, 115–124. [CrossRef]
75. Nguyen, D.L.; Le, P.T.T.; Le, T.T. Using Gemini for Formative Assessment in English Academic Writing—Critical Insights into The AI Tool's Efficacy. *AsiaCALL Online J.* **2025**, *16*, 328–343. [CrossRef]
76. Field, A. *Discovering Statistics Using IBM SPSS Statistics*; SAGE Publications: Thousand Oaks, CA, USA, 2024.
77. Lea, R.B.; Cohen, B.H. *Introductory Statistics for the Behavioral Sciences*; Wiley: Hoboken, NJ, USA, 2025.
78. Hedeker, D.; Gibbons, R.D. A Random-Effects Ordinal Regression Model for Multilevel Analysis. *Biometrics* **1994**, *50*, 933–944. [CrossRef]
79. Christensen, R.H.B. Cumulative link models for ordinal regression with the R package ordinal. *J. Stat. Softw.* **2018**, *35*, 1–46.
80. Agresti, A.; Kateri, M. Categorical Data Analysis. In *International Encyclopedia of Statistical Science*; Lovric, M., Ed.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 206–208.
81. Wieditz, J.; Miller, C.; Scholand, J.; Nemeth, M. A Brief Introduction on Latent Variable Based Ordinal Regression Models With an Application to Survey Data. *Stat. Med.* **2024**, *43*, 5618–5634. [CrossRef]
82. Martin, F.; Bolliger, D.U. Engagement matters: Student perceptions on the importance of engagement strategies in the online learning environment. *Online Learn.* **2018**, *22*, 205–222. [CrossRef]
83. Koivisto, J.; Hamari, J. The rise of motivational information systems: A review of gamification research. *Int. J. Inf. Manag.* **2019**, *45*, 191–210. [CrossRef]
84. Khoshnoodifar, M.; Ashouri, A.; Taheri, M. Effectiveness of Gamification in Enhancing Learning and Attitudes: A Study of Statistics Education for Health School Students. *J. Adv. Med. Educ. Prof.* **2023**, *11*, 230–239. [CrossRef]
85. Kassenkhan, A.M.; Moldagulova, A.N.; Serbin, V.V. Gamification and Artificial Intelligence in Education: A Review of Innovative Approaches to Fostering Critical Thinking. *IEEE Access* **2025**, *13*, 98699–98728. [CrossRef]
86. Er, E.; Akçapınar, G.; Bayazıt, A.; Noroozi, O.; Banihashem, S.K. Assessing student perceptions and use of instructor versus AI-generated feedback. *Br. J. Educ. Technol.* **2025**, *56*, 1074–1091. [CrossRef]
87. Almeida, C.; Kalinowski, M.; Uchôa, A.; Feijó, B. Negative effects of gamification in education software: Systematic mapping and practitioner perceptions. *Inf. Softw. Technol.* **2023**, *156*, 107142. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.