

November – 2025

Assessment of Learner Engagement and Expert Evaluations of AI-Generated Versus Human-Created Interactive Content in an Online Course

Hamza Aydemir¹ and Şeyda Kır²

¹Faculty of Communication, Kahramanmaraş İstiklal University, Kahramanmaraş, Türkiye; ²Technology Transfer Office, Yozgat Bozok University, Yozgat, Türkiye

Abstract

Generative artificial intelligence (GenAI) has introduced a novel aspect to educational methodologies and sparked fresh dialogues regarding the creation and evaluation of instructional resources. This project seeks to investigate the impact of GenAI on the development and assessment of online course materials and learners' engagement with these materials in the online learning environment. The study analyzed GenAI-generated multiple-choice questions, fill-in-the-blank exercises, and true-false activities during 3 weeks of a 14-week online course. Subject matter experts assessed these documents in regards to content, relevance, and clarity. Data was collected through an online form with open-ended questions. The interactions of learners with the GenAI-created learning activities were analyzed using log records of the learning management system and compared to the content provided by the course instructor regarding interaction levels. The study's conclusions elucidate the capability of GenAI technologies to produce course-specific content and their efficacy in education. We stress that human specialists' critical evaluations play a crucial part in improving the pedagogical validity of GenAI-powered learning materials. Further research into topics including the ethical dimension, the effect on academic achievement, and student motivation is recommended.

Keywords: generative AI, online education, learner engagement, higher education, case study

Introduction

Education as well as the economic, social, political, and cultural spheres of life are all impacted in different manners by artificial intelligence (AI) technologies. AI tools are becoming increasingly prevalent in education for design, production, distribution, and access to learning, and the potential in these areas continues to grow dramatically. The use of chatbots in education, learning analytics, intelligent tutoring systems, virtual learning assistants, and applications offering personalized learning experiences are just a few examples (Aydemir & Seferoğlu, 2024). Additionally, it is widely acknowledged that AI technologies are frequently used to assist educators, administrators, and learners in various ways, especially in settings that enable face-to-face or online instruction (Kir & Şenocak, 2022). The advancement of AI technologies to generative AI (GenAI), as well as the gradual integration of these tools into people's daily lives, has brought an unanticipated dimension to the applications and projects planned for the very near future. Reflections on GenAI in the field of education have grown to be a hot subject in the past few months, and this prominence has contributed to the widespread implementation of many kinds of AI technologies in educational environments. It is projected that enhanced development and organization using GenAI, in addition to the development of the capacity it provides, will usher in an era of innovation in instructional design (Bozkurt, 2023; Haleem et al., 2022; Kasneci et al., 2023).

As GenAI technologies evolve, they can enhance educational practices by providing personalized learning experiences, facilitating content creation, and promoting innovative teaching strategies. However, the integration of these technologies raises ethical concerns and requires careful consideration of their impact on traditional educational paradigms. GenAI can significantly transform instructional design by enabling the rapid creation of educational materials. For example, AI tools can generate quizzes, lesson plans, and multimedia content based on specific learning objectives, thus allowing educators to focus more on pedagogical strategies rather than content creation (DaCosta & Kinsell, 2024; Wood & Moss, 2024). However, productive AI use in education is not without its challenges. A key concern is the potential for bias towards AI-generated content (Ferrara, 2023). It is crucial for educators to critically evaluate the outputs of AI systems to ensure they align with ethical standards and promote inclusivity.

Many have highlighted the necessity to concentrate on the positive outcomes that the adoption of these technologies would bring for humanity with the arrival of GenAI technology at the end of 2022 (Lee, 2023). It has been suggested that it would be advantageous to adopt a more balanced approach in future studies by taking into account both the possible advantages and disadvantages of GenAI technology in the field of education (Bozkurt, 2023). Furthermore, research into the favorable effects of GenAI and ChatGPT in educational settings has suggested that this new technology will elevate traditional teaching methodologies to a new level and induce a paradigm shift (Tili et al., 2023). This paradigm shift requires an examination of the influence of GenAI on the instructional design and learning processes inside a course. This study primarily investigated the significance of GenAI technologies in the design of learning materials and their impact on learning efficiency.

Purpose of the Study

The purpose of this research was to assess the content quality of interactive learning materials created by GenAI for an online course. This series of research questions were put forward:

1. At what level would learners engage with GenAI-generated and human-created interactive content?
2. What are the expert opinions on the questions generated by GenAI in terms of content, clarity, and relevance?
 - a. How do experts evaluate multiple-choice questions?
 - b. How do experts evaluate true-false questions?
 - c. How do experts evaluate the fill-in-the-blank questions?
3. What are the expert opinions on the quality of the questions generated by GenAI?

The efficacy of instructional materials hinges on a triad of core principles: content, clarity, and relevance. Content encompasses the substance of the material—its information, concepts, and skills. To be effective, this content must be accurate, comprehensive, and aligned with learning objectives, while also being engaging enough to sustain student motivation (Sun, 2010). Equally important is clarity, which pertains to the lucidity and logical organization of the presentation. Information must be articulated in a straightforward and unambiguous manner to ensure comprehension and prevent confusion (Sundari et al., 2023). Finally, relevance bridges the material to the learner's world. By connecting to students' lives, interests, and cultural contexts, learning becomes more meaningful (Jamilah et al., 2024), a practice shown to significantly boost learner engagement and motivation (Zabala-Vargas et al., 2021). This synergy of rich content, clear presentation, and contextual relevance is what ultimately supports instructional goals and enhances student learning outcomes.

Background

The integration of GenAI into open and distance learning (ODL) demands critical examination through established pedagogical frameworks. As a technology that leverages deep learning to produce novel, human-like content (Lim et al., 2023), GenAI has reached a tipping point, sparking a wide spectrum of views on its potential impact on education (Lodge et al., 2023). While the use of AI in education is not new (Alasadi & Baiz, 2023; Chen et al., 2022), GenAI's advanced capabilities necessitate a deeper understanding of its pedagogical implications.

The Community of Inquiry (CoI) framework (Garrison et al., 2000), which emphasizes the interplay of cognitive, social, and teaching presence, provides a valuable lens. The pedagogical quality of GenAI-generated content must be assessed on its capacity to foster meaningful cognitive engagement and replicate core instructional functions. This is particularly relevant in assessment, a long-standing area of AI research (Zawacki-Richter et al., 2019), where GenAI can automate the creation of sophisticated questions using machine learning and natural language processing (Bachiri & Mouncif, 2023; Skanda et al., 2020; Tran et al., 2021). Furthermore, Anderson's (2003) Interaction Equivalency Theorem has suggested that robust learning can occur if one form of interaction, such as learner-content, is highly developed. GenAI can enhance this interaction by personalizing learning materials (Lodge et al., 2023), employing advanced transformer models to create context-aware queries (Kriangchaivech & Wangperawong, 2019). However, this must align with cognitive load theory (Sweller, 1988), ensuring that AI-generated content is designed to support, not overwhelm, learners' cognitive processing.

While GenAI offers powerful tools for developing tailored course materials and even assisting in academic writing (Ali & OpenAI Inc., 2023), its application is fraught with challenges. Balancing the scalability and benefits of AI with profound ethical, privacy, and security considerations is paramount (Lim et al., 2023; Nguyen et al., 2023). Therefore, responsible integration into ODL systems depends on rigorous evaluation and continuous human oversight to ensure both pedagogical effectiveness and ethical use.

Method

In this research, a case study, one of the qualitative research methods, was adopted. A case study can be a single situation selected from life or it can consist of multiple situations that are limited to a certain period. At this point, the case with well-defined boundaries can be a concrete entity such as a person, institution, or group, as well as an abstract concept such as a process or a relationship (Creswell, 2013; Yin, 2014). Yin (2014) argued that the most important point to focus on in case studies is to seek answers to how and why questions.

Research Context

The study was conducted in the distance computer programming department of a state university in Turkey. Forty-four students enrolled in the Mobile Programming course participated in the study. The course was about mobile application development. In the course, the theoretical background of the mobile application development process was discussed, the Dart programming language (<https://dart.dev/>) was explained practically, the creation of a mobile application interface in Flutter (<https://flutter.dev/>) was explained, and sample mobile applications were programmed in Flutter. The theoretical and practical parts of the course were completely synchronous and remote.

Within the scope of the research, tasks were defined for the students in the learning management system (LMS) for the first 8 weeks after the theoretical course. After the tasks, including the file upload task, assessment activities using the H5P tool (<https://h5p.org/>) were prepared by the researcher and defined to the students weekly over the course of 3 weeks. In the following 3 weeks, H5P activities prepared by the GenAI were defined. The learning materials included multiple-choice questions, fill-in-the-blank problems, and true-false activities. The process is shown in Figure 1.

Figure 1

Research Process



Note. Generative artificial intelligence (GenAI) is a type of artificial intelligence that uses patterns from existing data to create new and original content. H5P is a free and open-source tool primarily used for creating and sharing interactive content (<https://h5p.org/>).

Finally, the learning materials generated by GenAI were sent to experts and their opinions were collected through an online form. These content experts were asked to evaluate these materials in terms of content, relevance, and clarity through a qualitative data collection form, and the evaluation based on these three criteria was combined and interpreted under the concept of “appropriateness.” In addition, to provide in-depth answers to the research questions, students’ interactions in the LMS were analyzed and used to interpret the qualitative data.

Participants

The participants consisted of 10 experts, with some specializing in instructional technology and others in computer engineering. The selection of experts from technical sciences was influenced by the case study’s focus on mobile programming. Table 1 presents the demographic information of these experts.

Table 1

Demographic Information of the Ten Experts Participating in the Study

Expert	Expertise	Degree	Experience, <i>n</i> of years
1	Instructional technologies	Master’s degree	5–10
2	Instructional technologies	Master’s degree	5–10
3	Instructional technologies	Bachelor’s degree	5–10
4	Instructional technologies & computer engineering	Master’s degree	10–15
5	Computer engineering	Master’s degree	1–5
6	Instructional technologies	Master’s degree	1–5
7	Computer engineering	Master’s degree	5–10
8	Computer engineering	Bachelor’s degree	5–10
9	Electrical and electronic engineer	Master’s degree	10–15
10	Instructional technologies	Bachelor’s degree	5–10

The majority of the experts were in the field of instructional technologies, and the majority of them had obtained a master’s degree. Except for two, they were also engaged in pursuing doctoral education. In terms of years of experience, the most common group was those with 5–10 years of experience. The distribution of computer engineering specialists appeared to be balanced, with an additional specialist in electrical and electronic engineering.

Generation of H5P Activities With Generative Artificial Intelligence

Our study employed Nolej AI (<https://www.nolej.io/>), a GenAI tool capable of generating H5P activities, albeit with certain limitations in its free usage. This platform, which can accept document or

multimedia content as input, subsequently transforms that content into an H5P activity. In this study, the researcher provided a document containing the theoretical course content to the system during the relevant weeks. The content for each week consisted of an average of 2,000 words. Following the generation of the H5P activities, a review and testing phase was initiated. Once it was determined that there were no technical errors, the content H5P activities was integrated into the course's LMS.

Research Design

In the research, the interaction of learners with the learning activities created by GenAI and the appropriateness of those activities to the course context were accepted as a case, and the case study method was applied (Yin, 2014). Interaction data were obtained from the records of the course on the LMS, while qualitative data were collected by asking open-ended questions to the participants through an online form. The study group included in the qualitative part of the research consisted of experts with knowledge and experience about the content of the course.

Ethical Issues

Before the data collection phase of the study, a declaration of ethical compliance was obtained by a state university ethics committee commission numbered 179885 and dated November 20, 2023, and the research process was initiated after this declaration of ethical compliance. In addition, participant selection criteria were created, and those who voluntarily participated were included in the data collection process. In addition, students' interaction data in the LMS were anonymized before being analyzed and reported.

Data Analysis and Learning Analytics

A multifaceted approach was adopted to address the research questions. In this context, the opinions of experts were collected using an online form. The data obtained were analyzed quantitatively with descriptive statistics and qualitatively with a thematic approach. To understand the result of the analysis in-depth, the interaction data of students in the LMS were also analyzed and visualized using descriptive statistics. RStudio (Version 2024.12.1+563) was employed for the analysis of quantitative data.

During the research period, 25,595 records were logged in the LMS. Before the initiation of data analysis, the records underwent a pre-process of anonymization. Subsequently, H5P interactions included in weeks 9–14 were filtered out. Within the context of this study, two metrics—viewing and completing the relevant activity—were evaluated based on student-based time. This process yielded a total of 925 logs that were subsequently analyzed and reported.

Limitations of the Study

The findings of this study are constrained to the learning materials produced within the scope of the online course titled Mobile Programming and the data collected from experts in the context of these materials. Furthermore, the situation revealed in the research findings and discussion is limited to the opinions and data obtained from the participants involved in the study. In terms of understanding student interaction and engagement, the study is limited to the user interactions recorded by the Moodle system. It is acknowledged that the LMS data may not be sufficient to make a full assessment of students' learning processes and engagement. Therefore, interaction data were used to understand expert opinions in more depth.

Results and Discussion

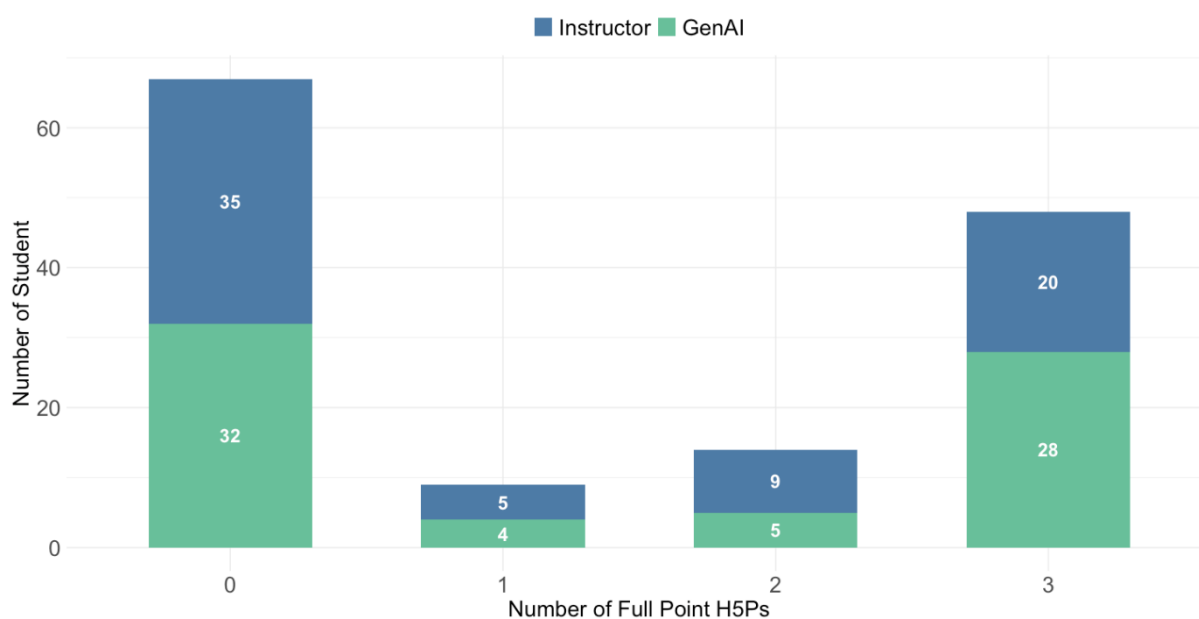
This study explored the engagement levels of students with interactive content generated by GenAI and traditional human-created content. Additionally, it delved into expert assessments of GenAI-generated inquiries, with a particular emphasis on pivotal factors such as content quality, clarity, and relevance. The analysis encompassed a range of question types, including multiple-choice, true-false, and fill-in-the-blank questions. By exploring expert opinions and student engagement, this study has the potential to provide insights into the effectiveness and educational value of interactive content created by GenAI.

Learner Engagement Levels With GenAI-Generated and Human-Created Interactive Content

Analysis of student engagement with the H5P activities offered insights into performance patterns based on content origin. Out of the 69 students in this distance course, which does not require participation in online activities or completion of assignments, 25 did not participate in these activities. Figure 2 illustrates the performance distribution for the remaining students across three activities created by the instructor and three generated by GenAI. The figure displays the number of students who achieved full points on zero, one, two, or all three activities. The blue bars represent performance on instructor-created content, while the green bars represent performance on GenAI-generated content.

Figure 2

Distribution of Getting Full Points From Instructor-Versus GenAI H5P Activities



Note. Generative artificial intelligence (GenAI) is a type of artificial intelligence that uses patterns from existing data to create new and original content. H5P is a free and open-source tool primarily used for creating and sharing interactive content (<https://h5p.org>).

As illustrated in Figure 2, there were distinct performance distributions between the two content types. A primary observation was the substantial cohort of students who did not achieve full points in any activity, a figure slightly higher for instructor-created content than for GenAI-created content. In cases

of partial completion (one or two activities), the instructor-created materials appeared to have facilitated incremental success for a slightly larger group of students.

The most significant divergence, however, was evident in the category of complete mastery. A considerably larger group of students successfully obtained full points on all three GenAI-generated activities compared to the instructor-created set. This suggests a polarization of outcomes, particularly with GenAI content; students who engaged successfully were more likely to achieve complete mastery. Conversely, the instructor-created content, while leading to fewer instances of complete mastery, showed a slightly greater capacity to support students in achieving partial success. These findings underscore that GenAI- and instructor-led approaches may produce distinct learning trajectories, with GenAI potentially creating a more uniform “all-or-nothing” performance pattern, while the instructor’s activities might possess a more varied difficulty curve. The finding that GenAI-generated content produced a more polarized performance pattern aligns with research on AI-driven personalized learning. As Kasneci et al. (2023) suggested, such tools can create customized learning paths that allow some students to accelerate toward complete mastery, which may explain the higher number of students achieving full points in our study.

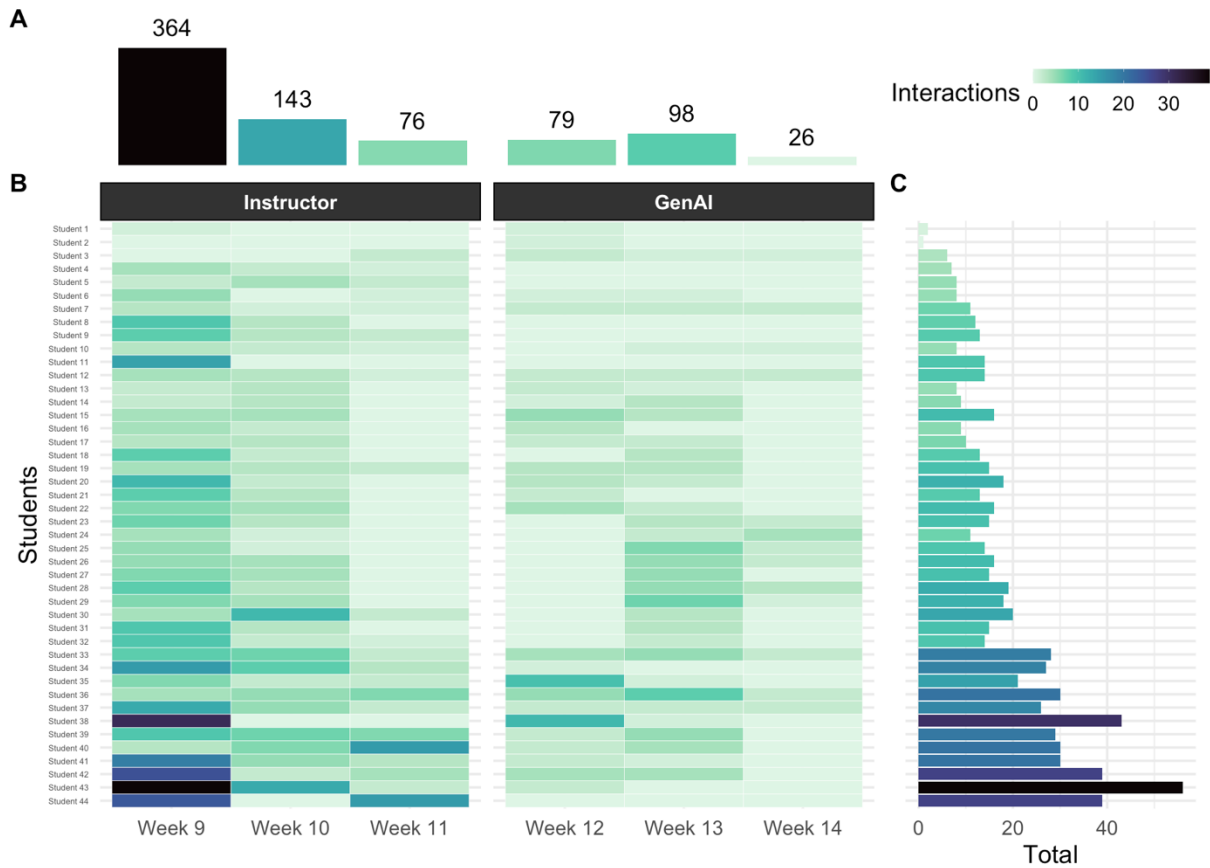
Figure 3 provides a granular, student-level visualization of interaction patterns with course materials across 6 weeks, distinguishing between content created by the instructor (weeks 9–11) and by GenAI (weeks 12–14). The central heatmap displays the interaction intensity for each of the 44 students who participated in the activities, where lighter shades of green indicate fewer interactions and darker shades of blue and purple signify a higher number of interactions, as detailed in the color-scale legend. The figure is composed of three interconnected parts:

1. A bar chart at the top summarizes the total number of interactions across all students for each week.
2. The main heatmap illustrates individual student engagement week-by-week.
3. A horizontal bar chart on the right aggregates the total interactions for each student over the entire 6-week period, sorted from least to most engaged.

The interaction data represents the sum of all viewing and completion events recorded in the LMS for each student on the H5P activities.

Figure 3

Distribution of Students' Number of H5P Activity Interactions



Note. Panel A summarizes the total number of interactions across all students for each week. Panel B illustrates individual student engagement week-by-week. Panel C aggregates the total interactions for each student over the entire 6-week period of the activities, sorted from least to most engaged.

Analysis revealed a multi-faceted narrative of student engagement as illustrated in Figure 3. Visually, the most prominent feature was the dramatic peak in interactions seen during week 9, the initial week of the instructor-led activities. This was followed by a sharp and sustained decrease in engagement for all subsequent weeks, showing that the timing within the course or the novelty of the first activities had a significant impact on the overall volume of interactions. Furthermore, stratification of the student body was suggested by the figure: there was a small cohort of highly engaged students (i.e., students 35-44) accounting for a disproportionate number of total interactions, driving the peak in week 9 and remaining the most active throughout the study period.

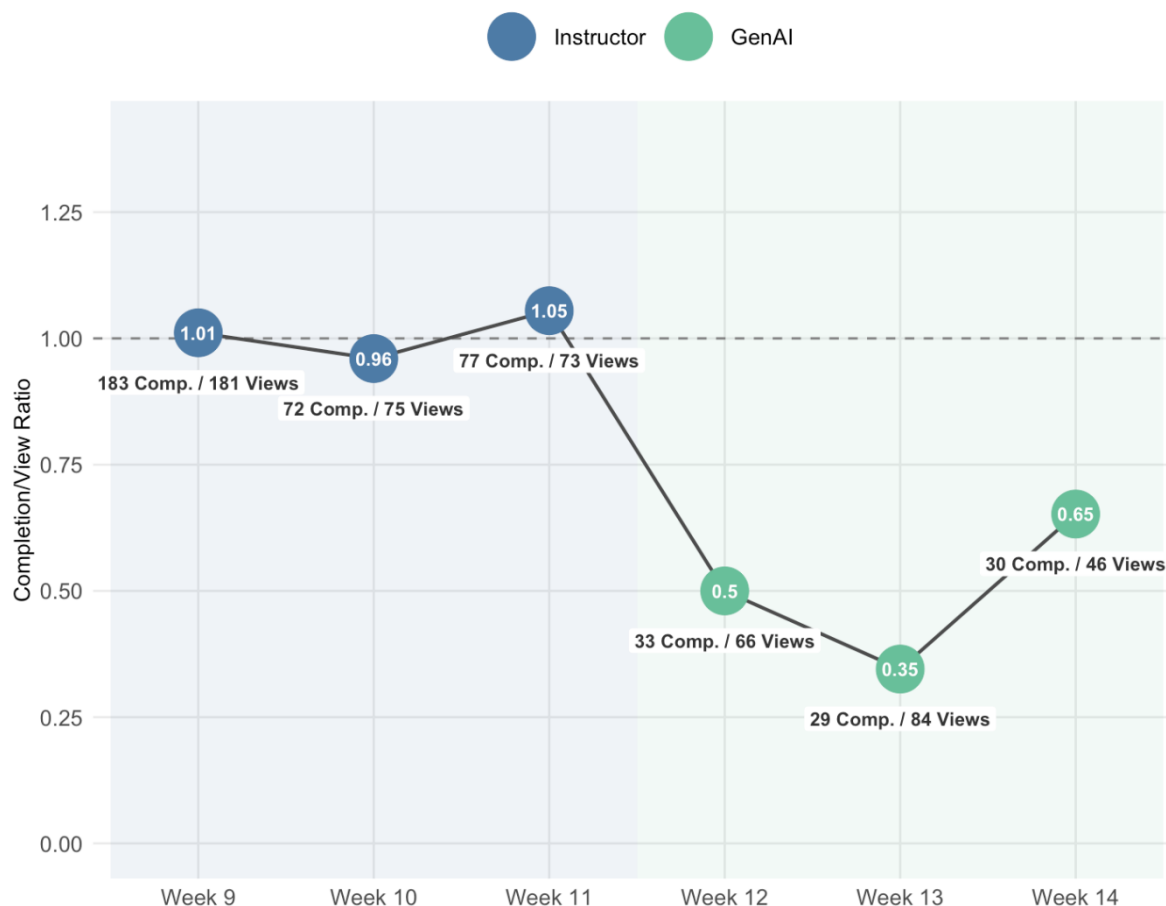
Beyond these visual patterns of timing and distribution, a deeper analysis of the underlying log data exposed critical differences in the nature of these engagements. For the instructor-created content, the ratio of views to completions was 0.99, and activities were often completed more than once. In contrast, for GenAI-generated content, the view-to-completion ratio was 2.13, and activities were typically completed only once.

When viewed together, these two sets of findings provided a comprehensive picture. The high interaction volume in week 9, combined with the efficient completion rate and high repetition, strongly

suggested that the engaged cohort used the instructor’s initial activities for mastery-oriented learning. Conversely, the less efficient view-to-completion ratio for GenAI content suggested that students may have struggled with instructional clarity or motivation, leading to more superficial engagement (Figure 4). This dichotomy underscores that while Generative AI is a powerful content-authoring tool, it requires expert human oversight to ensure pedagogical alignment and effectiveness. However, it is crucial to acknowledge that these differing engagement qualities are set against a backdrop of declining overall interaction, reinforcing that student engagement is highly dynamic and influenced by factors beyond just the content’s origin.

Figure 4

Weekly Student Interaction Completion/View Ratio of Instructor and GenAI H5P Materials



Note. Generative artificial intelligence (GenAI) is a type of artificial intelligence that uses patterns from existing data to create new and original content. The ratio is calculated by dividing the total number of completion events by the total number of view events for each week. Data labels below each point indicate the raw counts for completions (Comp.) and views used to calculate the ratio for that week. The dashed line at a ratio of 1.0 serves as a baseline, indicating an equal number of completions and views.

Figure 4 reveals a difference in the quality of student engagement between the two phases. The instructor-led H5P activities (Weeks 9-11) are characterized by a high engagement efficiency. The completion-to-view ratio consistently hovers around 1.0, indicating that nearly every view led to a completion. This pattern, particularly the instances where completions exceed views, suggests that students were deeply engaged with the material, likely using it for mastery-oriented learning rather than a cursory review. While the overall interaction volume peaked in Week 9 and then declined, the

efficiency of these interactions remained high, pointing to the pedagogical effectiveness of the instructor-authored content. The initial high volume may indeed be partially attributed to a novelty effect, as students encountered their first H5P activity. Research has indicated that heightened engagement derived from novelty diminishes over time (Liang et al., 2020). Figure 4 is consistent with this finding. However, comparative experimental interventions are needed for a more accurate conclusion.

In contrast, the transition to the GenAI-generated content (Weeks 12-14) is marked by a sharp drop in engagement efficiency. The ratio falls significantly below 1.0, revealing that students viewed the content far more often than they completed it. This suggests a more superficial or exploratory form of engagement, where students may have struggled with instructional clarity or lacked the motivation to complete the activities. While the total number of interactions in this phase appears more stable at a lower level, the quality of these interactions, as measured by the completion ratio, is substantially lower than in the instructor-led phase. However, it is crucial to acknowledge that this apparent difference in performance could be influenced by a confounding variable—namely, that the instructor H5Ps were previously available in the system. Therefore, a dedicated empirical study is needed to isolate the causal impact of content authorship and allow for robust generalizations.

In summary, the findings indicate that a hybrid application of instructor-created and GenAI-generated materials may address different aspects of student engagement. The materials created by the instructor were associated with higher completion-to-view ratios, a pattern consistent with mastery-oriented learning. Conversely, content generated by GenAI corresponded with lower engagement efficiency, suggesting a need for human oversight to enhance its pedagogical alignment and effectiveness. Although AI can be leveraged to reduce unnecessary cognitive load and increase intrinsic motivation (Guo et al., 2023; Xu & Ismail, 2024), the data suggest these outcomes are contingent on careful implementation. The integration of these two methods aligns with tenets of engagement theory, which posits that collaborative interactions can be amplified through technological applications to foster meaningful educational experiences (Bachiri et al., 2023; Huang et al., 2021).

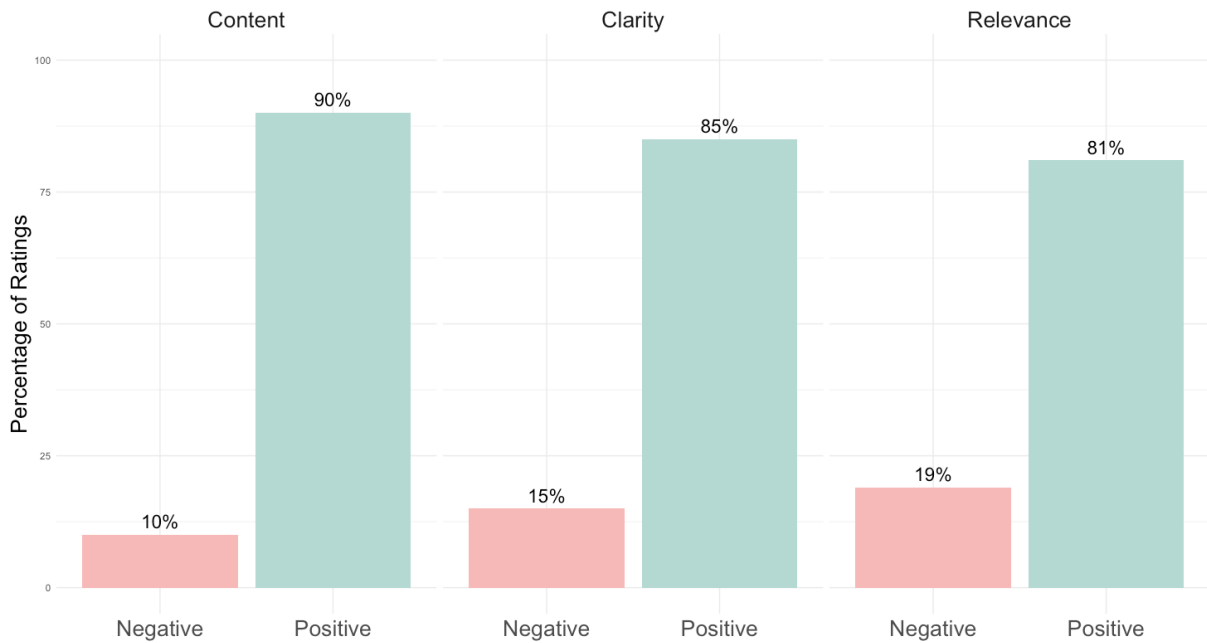
While these quantitative interaction patterns from the LMS provided valuable insight into student engagement, they do not, on their own, explain the underlying reasons for these behaviors. The data revealed what students did, but not intrinsic quality or pedagogical soundness of the materials they interacted with. To evaluate GenAI-generated content, clarity, and relevance, and thus to contextualize the engagement findings, an assessment by subject matter experts was conducted. The following sections present the results of this expert evaluation, which scrutinized GenAI materials from a pedagogical and content-validity perspective.

Expert Opinions on GenAI-Generated Questions

The expert opinions on the content, clarity, and relevance of the multiple-choice questions generated by Gen AI were analyzed. To evaluate the GenAI-generated questions, subject matter experts rated multiple-choice, true-false, and fill-in-the-blank questions across three criteria: content, clarity, and relevance. For each question and criterion, experts provided a binary rating of 'Negative' (inappropriate) or 'Positive' (appropriate). The percentages shown in the figures represent the aggregate distribution of these ratings. Figure 5 shows the percentage of positive versus negative opinions in each category under consideration.

Figure 5

Expert Evaluation of GenAI-Generated Multiple-Choice Questions



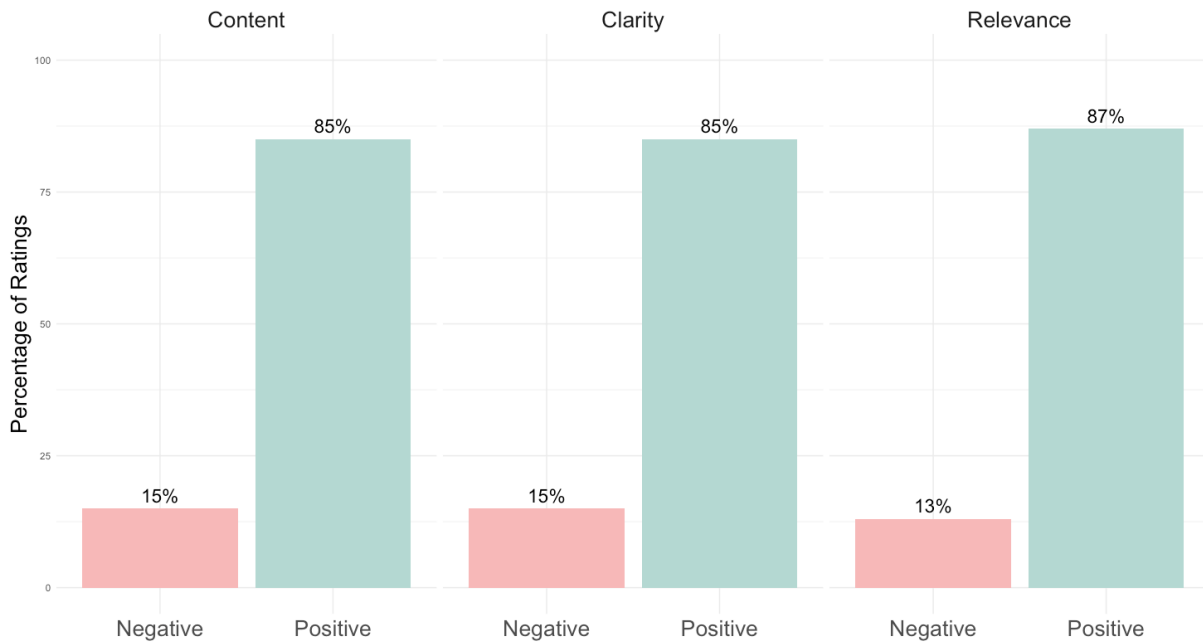
Note. Generative artificial intelligence (GenAI) is a type of artificial intelligence that uses patterns from existing data to create new and original content.

As can be seen in Figure 5, the ratings for multiple-choice questions created by GenAI were, on average, 85% positive across the three criteria. Among the criteria for multiple-choice questions, 'Content' received the highest percentage of positive ratings (90%), while 'Relevance' received the lowest (81%).

The ratio of the content, clarity, and relevance levels of the true-false questions that received expert opinion within the scope of the research is presented in Figure 6. It shows consistency and that approximately 85% of the true-false questions prepared by GenAI were found to be appropriate.

Figure 6

Expert Evaluation of GenAI-Generated True-False Questions

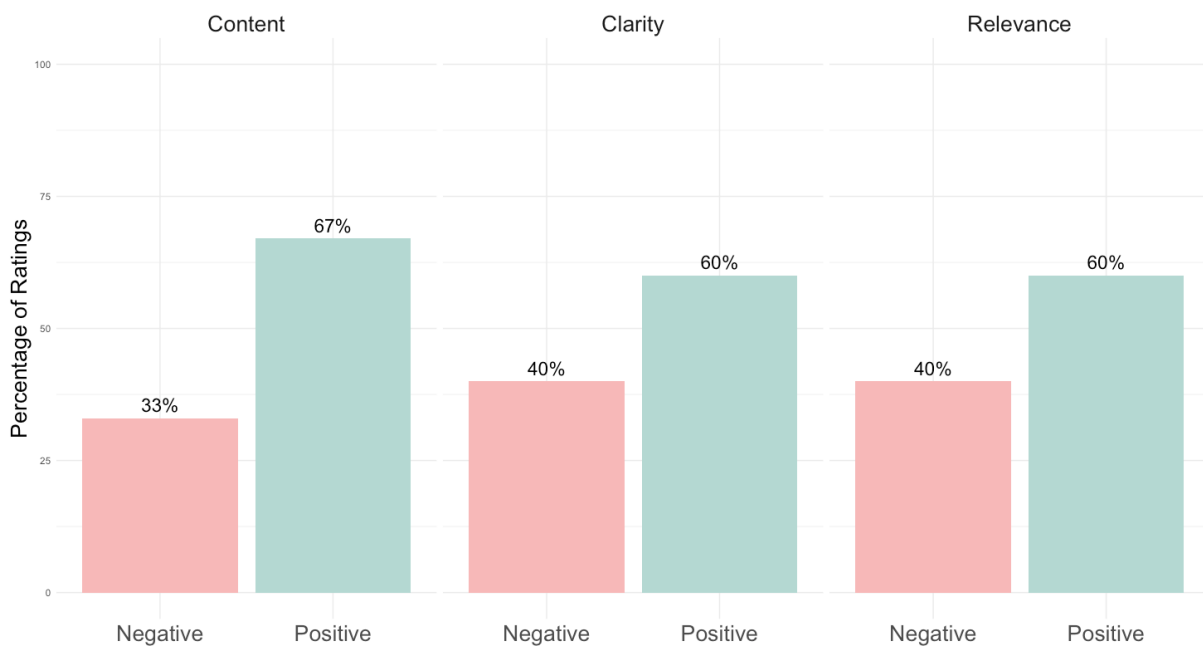


Note. Generative artificial intelligence (GenAI) is a type of artificial intelligence that uses patterns from existing data to create new and original content.

The ratio of the content, clarity, and relevance levels of the fill-in-the-blank questions that received expert opinion within the scope of the research is presented in Figure 7.

Figure 7

Expert Evaluation of GenAI-Generated Fill-in-the-Blank Questions



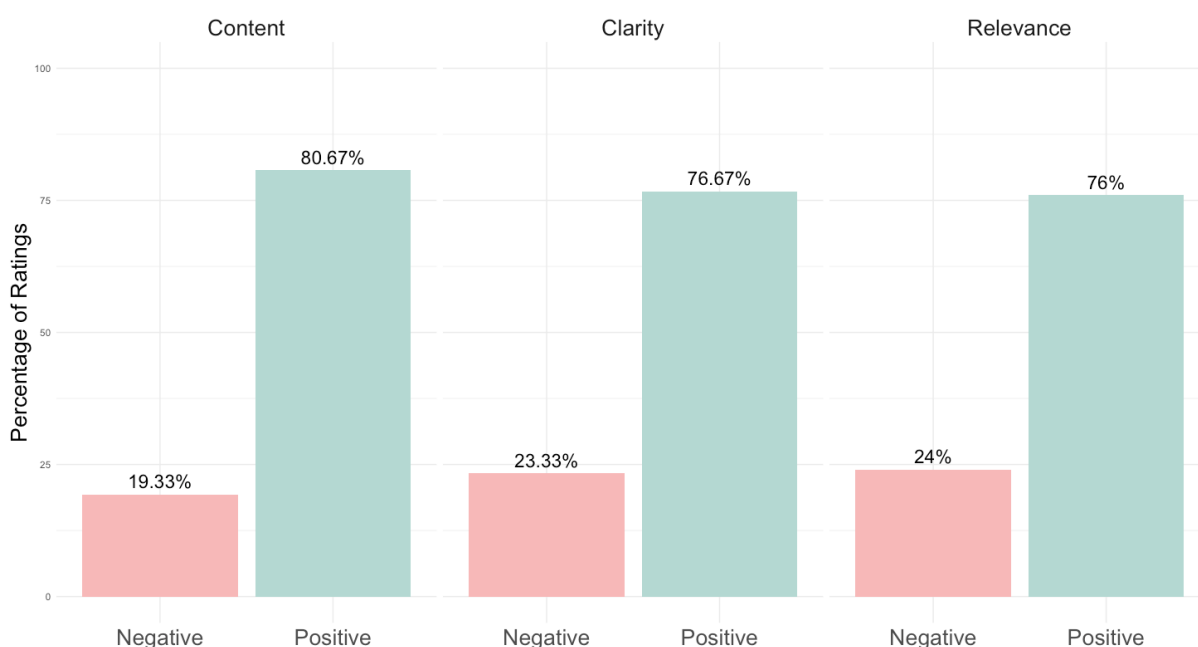
Note. Generative artificial intelligence (GenAI) is a type of artificial intelligence that uses patterns from existing data to create new and original content.

The fill-in-the-blank questions created by GenAI were found appropriate at an average rate of 62%. Experts evaluated the content quality of the questions as the highest and the clarity and relevance of the questions as the lowest. However, in contrast to the other question types, the majority of ratings for fill-in-the-blank questions were not positive, indicating a general inadequacy in all three criteria.

To provide a holistic view of the quantitative data, the expert ratings for all three question types (multiple-choice, true-false, and fill-in-the-blank) were aggregated. Figure 8 presents this general summary, illustrating the overall distribution of 'Positive' and 'Negative' ratings across the core criteria of content, clarity, and relevance.

Figure 8

Aggregated Distribution of Expert Evaluation Across All Question Types



As shown in Figure 8, which aggregates all ratings, the overall positive rating percentage was highest for 'Content' (80.67%) and lowest for 'Relevance' (76%).

The findings indicated that the content quality of interactive learning materials produced by GenAI was inconsistent and highly dependent on the assessment format. GenAI was proficient at generating structured question types, with both multiple-choice and true-false questions being found appropriate by experts at an average rate of approximately 85%. For these formats, which are often used in online courses to assess foundational knowledge, the AI demonstrated a strong ability to produce factually accurate and clear content, supporting the conclusions of other researchers that GenAI can be a valuable resource for educators. Research by Sihite et al. (2023) showed that GenAI-generated questions can cover a range of cognitive levels, thus providing a valuable resource for educators seeking to improve their assessment methods. Nasution (2023) assessed the validity and reliability of AI-generated biology

questions and showed that GenAI can produce high-quality educational materials. This suggests that for creating a baseline of standard assessment items, GenAI is a capable and efficient tool.

However, our study revealed significant limitations in GenAI's ability to generate content requiring deeper semantic understanding. Fill-in-the-blank questions were found to be largely inadequate, with an average appropriateness rate of only 62%, and experts rated them poorly across all criteria. Furthermore, even for the highly-rated multiple-choice questions, relevance was the least satisfactory criterion. This discrepancy suggests that while GenAI can assemble correct information, it struggles to ensure that the content is contextually aligned with specific instructional goals, a critical component of quality in educational materials.

The most effective educational outcomes are achieved through partnership rather than replacement. The critical role of subject matter experts in this study underscores a central theme in recent literature: while Generative AI is proficient at creating content, its outputs require human evaluation to ensure pedagogical nuance and contextual accuracy (Baidoo-Anu & Ansah, 2023). GenAI should be regarded as a powerful assistant for content creation in online learning, rather than a fully autonomous solution. While students report positive perceptions of AI tools and appreciate their support (Aydemir & Seferoğlu, 2024), our findings emphasized that this enthusiasm must be met with rigorous quality control. The variability in output quality necessitates a "human-in-the-loop" model, where instructional designers and subject matter experts play an indispensable role in validating, refining, and ensuring the pedagogical relevance of all AI-generated materials. While technology can accelerate the development of formative assessments, human oversight remains crucial to guarantee the high quality and instructional integrity required for effective online courses.

Qualitative Analysis of Expert Feedback on the Quality of GenAI-Generated Questions

Following the quantitative evaluation where experts provided ratings for the GenAI-generated questions, a qualitative analysis was conducted to explore the reasoning behind their assessments. In addition to the ratings, experts were asked to provide open-ended feedback on the overall quality, advantages, disadvantages, and suggestions for improving the H5P activities. This section presents a thematic analysis of their written responses, providing deeper context for the quantitative results presented previously.

In the qualitative data collection phase of this study, open-ended questions were asked to the participants. In their first three questions, participants were asked to evaluate the GenAI-generated questions on the concepts of content, clarity, and relevance by rating them as good, fair, or poor. Results are shown in Figure 9.

Figure 9

Evaluation of the Content, Clarity, and Relevance of the Questions Generated by GenAI



Note. Generative artificial intelligence (GenAI) is a type of artificial intelligence that uses patterns from existing data to create new and original content.

The evaluation by ten experts indicated that the 'Content' and 'Clarity' of the GenAI-generated questions were rated as "Good" at an identical proportion of 70%, while the 'Relevance' of the questions received a lower proportion of "Good" ratings. Notably, for both 'Content' and 'Clarity', no ratings were in the "Poor" category. However, experts expressed greater concern regarding the 'Relevance' of the questions, where ratings were distributed across "Fair" (36.4%) and "Poor" (9.1%) categories, in addition to "Good". These findings suggested that while GenAI is proficient in generating clear and factually sound content, its ability to align that content with specific pedagogical contexts—its relevance—requires further improvement. Jouault et al. (2016) conducted a comparative study between automatically generated questions and questions generated by human experts. They found that GenAI-generated questions were able to cover a significant portion of knowledge acquisition objectives, thus confirming the effectiveness of GenAI in educational settings (Aydemir & Seferoğlu, 2024).

In the other three questions in the qualitative data collection process, the experts were asked for their opinions about the quality, advantages-disadvantages, and suggestions for improvement of the questions developed by GenAI, respectively.

Figure 10 shows their opinions about quality, categorized by theme.

Figure 10

Experts' Opinions on the Quality of Questions Developed by GenAI

Themes	Experts
Content appropriateness	E3, E5, E6, E8, E10
Human intervention required	E2, E4, E5, E6
Suitability to learner level	E6, E8, E9
Uncertainties	E4, E1

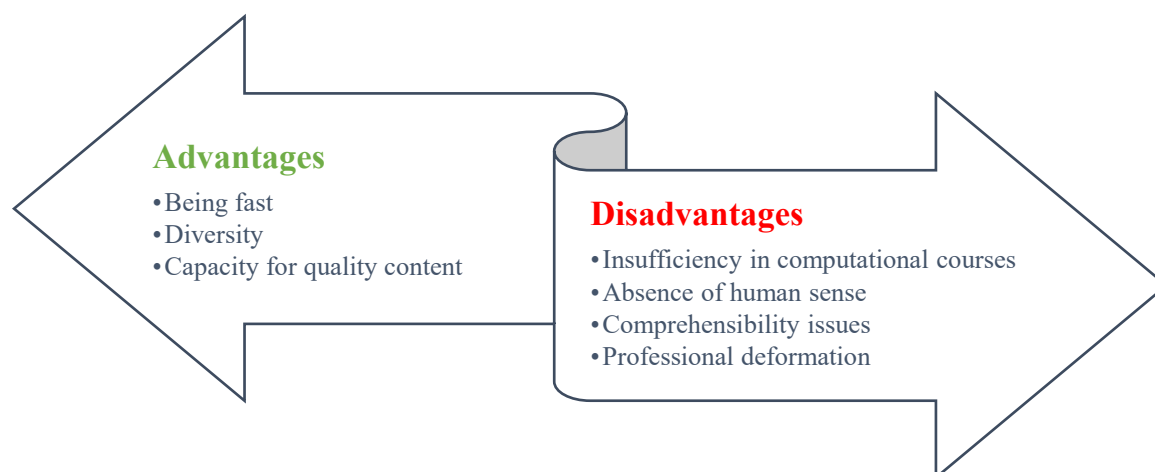
Note. E = expert participant.

While *content appropriateness* was the most emphasized theme, identified by five experts, four experts emphasized *human intervention required*. One expert wrote, "I see no harm in using it as long as it is re-evaluated by the instructor. If it is not revised, it can still make simple mistakes." The theme, *suitability to learner level*, was a point that drew the attention of three experts. Finally, two experts drew attention to the fact that some questions contained incomplete information or had poor comprehensibility, and the theme of *uncertainties* emerged.

In the next question, experts were asked about the advantages and disadvantages of the questions developed by GenAI. Figure 11 shows the results of their answers.

Figure 11

Potential Advantages and Disadvantages of AI-Generated Questions

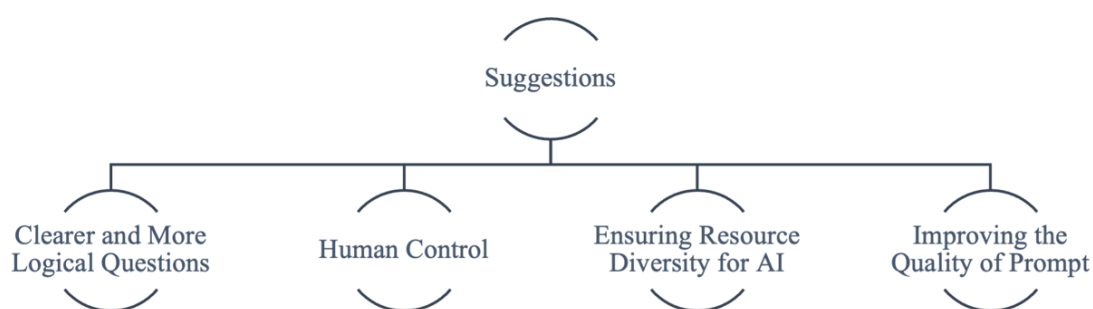


As seen in this table, there were criticisms, such as the issue that GenAI is insufficient in some subjects (such as computational courses) and that some questions contain ambiguity. In terms of advantages, themes of being fast, diversity, and capacity for quality content stood out. Among the disadvantages, comprehensibility issues would have a significant role.

In the last open-ended question, experts were asked for suggestions to improve the questions created by GenAI. The qualitative data obtained from this question were analyzed under four main themes: human control, clearer and more logical questions, ensuring resource diversity for GenAI, and improving the quality of prompts. The connections between these themes is shown in Figure 12.

Figure 12

Suggestions for Enhancing the Questions Formulated by GenAI



Note. AI = Artificial intelligence.

The most frequently mentioned suggestions of the experts were *human control* and *producing clearer and more logical questions*. One participant drawing attention to human control stated: “In my opinion, human control is important to improve the questions produced by GenAI. Human control of the content, appropriateness, and comprehensibility of questions can improve the quality of the questions.”

Furthermore, to enhance learning experience and learner success, augmenting the resources provided to GenAI and formulating clearer, more stimulating questions were also highlighted as recommendations for improvement. Enhancing the capacity to formulate questions through the generation of additional inquiries by GenAI was also included among recommendations. A further recommendation was to refine the terminology employed. One expert reported: “To preemptively train GenAI, particularly to uphold and preserve linguistic purity. It may be useful to train and test it in isolation without confusing it about topics to be produced [as] questions.” This idea could be considered while giving prompts.

Conclusion and Future Directions

In this research, course assessment activities for some weeks of a distance course were developed by GenAI, and learners’ interactions with these course assessment activities through a LMS were analyzed. In addition, the suitability of these assessment activities to the context of the course was evaluated by 10 experts in course content. In this evaluation process, the questions developed by GenAI were analyzed in terms of content, clarity, and relevance, and the experts were asked for their opinions with open-ended questions and a quantitative evaluation.

This study demonstrates that learners' engagement with content generated by GenAI and produced by instructors may fluctuate based on contextual variables. Results indicate that carefully integrating these two material categories may effectively address learners' diverse demands. More precisely, within the realm of distance education, materials produced by GenAI could serve as adapted instructional materials at the micro-instructional level and enhance educators' pedagogical approaches to design. To fully actualize this potential, it is essential to create professional development programs that empower educators to proficiently integrate AI content and enhance their AI literacy. Additionally, extensive systems such as massive open online course (MOOC) platforms need to establish criteria that define quality requirements for content generated through AI-human collaboration and allocate resources for data-driven enhancements to adaptive learning systems. Forthcoming research will substantially improve the area by elucidating these findings using experimental approaches that investigate learner interaction dynamics and their impact on learning outcomes.

References

- Alasadi, E. A., & Baiz, C. R. (2023). Generative AI in education and research: Opportunities, concerns, and solutions. *Journal of Chemical Education*, 100(8), 2965–2971. <https://doi.org/10.1021/acs.jchemed.3c00323>
- Ali, F., & OpenAI, Inc., C. (2023). Let the devil speak for itself: Should ChatGPT be allowed or banned in hospitality and tourism schools? *Journal of Global Hospitality and Tourism*, 2(1), 1–6. <https://doi.org/10.5038/2771-5957.2.1.1016>
- Anderson, T. (2003). Getting the mix right again: An updated and theoretical rationale for interaction. *The International Review of Research in Open and Distributed Learning*, 4(2). <https://doi.org/10.19173/irrodl.v4i2.149>
- Aydemir, H., & Seferoğlu, S. S. (2024). Investigation of users' opinions, perceptions and attitudes towards the use of artificial intelligence in online learning environments. *European Journal of Open, Distance and E-Learning*, 26(1), 70–92. <https://doi.org/10.2478/eurodl-2024-0007>
- Bachiri, Y.-A., & Mouncif, H. (2023). Artificial intelligence system in aid of pedagogical engineering for knowledge assessment on MOOC platforms: Open EdX and Moodle. *International Journal of Emerging Technologies in Learning (iJET)*, 18(05), 144–160. <https://doi.org/10.3991/ijet.v18i05.36589>
- Bachiri, Y.-A., Mouncif, H., & Bouikhalene, B. (2023). Artificial intelligence empowers gamification: Optimizing student engagement and learning outcomes in e-learning and MOOCs. *International Journal of Engineering Pedagogy (iJEP)*, 13(8), 4–19. <https://doi.org/10.3991/ijep.v13i8.40853>
- Baidoo-anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, 7(1), 52–62. <https://doi.org/10.61969/jai.1337500>
- Bozkurt, A. (2023). ChatGPT, üretken yapay zekâ ve algoritmik paradigma değişikliği [ChatGPT, generative artificial intelligence and algorithmic paradigm shift]. *Alanyazın* 4(1), 63–72. <https://doi.org/10.59320/alanyazin.1283282>
- Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C. (2022). Two decades of artificial intelligence in education. *Educational Technology & Society*, 25(1), 28–47. <https://www.jstor.org/stable/48647028>
- Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). Sage.
- DaCosta, B., & Kinsell, C. (2024). Investigating media selection through ChatGPT: An exploratory study on generative artificial intelligence in the aid of instructional design. *Open Journal of Social Sciences*, 12(4), 187–227. <https://doi.org/10.4236/jss.2024.124014>
- Ferrara, E. (2023). Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Sci*, 6(1), Article 3. <https://doi.org/10.3390/sci6010003>

- Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, 2(2–3), 87–105. [https://doi.org/10.1016/S1096-7516\(00\)00016-6](https://doi.org/10.1016/S1096-7516(00)00016-6)
- Guo, Z., Jing, T., Zhu, H., & Xu, X. (2023). Analysis of the impact of artificial intelligence on college students' learning. *The Frontiers of Society, Science and Technology*, 5(13), 110–116. <https://doi.org/10.25236/fsst.2023.051318>
- Haleem, A., Javaid, M., & Singh, R. P. (2022). An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, 2(4), Article 100089. <https://doi.org/10.1016/j.tbench.2023.100089>
- Huang, J. Shen, G., & Ren, X. (2021). Connotation analysis and paradigm shift of teaching design under artificial intelligence technology. *International Journal of Emerging Technologies in Learning (iJET)*, 16(05), 73–86. <https://doi.org/10.3991/ijet.v16i05.20287>
- Jamilah, Jayadi, K., Yatim, H., Sahnir, N., Djirong, A., & Abduh, A. (2024). The integration of local cultural arts in the context of teaching materials on the implementation of the Merdeka Belajar curriculum. *Journal of Education Research and Evaluation*, 8(2), 404–413. <https://doi.org/10.23887/jere.v8i2.78359>
- Jouault, C., Seta, K., & Hayashi, Y. (2016). Content-dependent question generation using LOD for history learning in open learning space. *New Generation Computing*, 34(4), 367–394. <https://doi.org/10.1007/s00354-016-0404-x>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, Article 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kır, Ş., & Şenocak, D. (2022). Açık ve uzaktan öğrenme sistemlerinde yapay zekânın öğrenen destek hizmeti bağlamında kullanımı [The use of artificial intelligence in the context of learner support services in open and distance learning systems]. *Dijital Teknolojiler Ve Eğitim Dergisi*, 1(1), 39–56. <https://doi.org/10.5281/zenodo.6647642>
- Kriangchaivech, K., & Wangperawong, A. (2019). *Question generation by transformers*. ArXiv. <https://doi.org/10.48550/arxiv.1909.05017>
- Lee, E. (2023). Is ChatGPT a false promise? *Berkeley Blog*. <https://coilink.org/20.500.12592/hzd6k>
- Liang, Z., Zhao, Q., Zhou, Z., Yu, Q., Li, S., & Chen, S. (2020). The effect of “novelty input” and “novelty output” on boredom during home quarantine in the COVID-19 pandemic: The moderating effects of trait creativity. *Frontiers in Psychology*, 11, Article 601548. <https://doi.org/10.3389/fpsyg.2020.601548>

- Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education*, 21(2), Article 100790. <https://doi.org/10.1016/j.ijme.2023.100790>
- Lodge, J. M., Thompson, K., & Corrin, L. (2023). Mapping out a research agenda for generative artificial intelligence in tertiary education. *Australasian Journal of Educational Technology*, 39(1), 1–8. <https://doi.org/10.14742/ajet.8695>
- Nasution, N. E. A. (2023). Using artificial intelligence to create biology multiple choice questions for higher education. *Agricultural and Environmental Education*, 2(1), Article em002. <https://doi.org/10.29333/agrenvedu/13071>
- Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B.-P. T. (2023). Ethical principles for artificial intelligence in education. *Education and Information Technologies*, 28, 4221–4241. <https://doi.org/10.1007/s10639-022-11316-w>
- Sihite, M. R., Meisuri, M., & Sibarani, B. (2023). Examining the validity and reliability of ChatGPT 3.5-generated reading comprehension questions for academic texts. *Randwick International of Education and Linguistics Science Journal*, 4(4), 937–944. <https://doi.org/10.47175/rielsj.v4i4.835>
- Skanda, V. C., Jayaram, R., Bukitagar, V. C., & Kumar, N. S. (2020). Automatic questionnaire and interactive session generation from videos. In A. Chandrabose, U. Furbach, A. Ghosh, & M. A. Kumar (Eds.), *IFIP advances in information and communication technology* (Vol. 578; pp. 205–212). Springer. https://doi.org/10.1007/978-3-030-63467-4_16
- Sun, Z. (2010). Language teaching materials and learner motivation. *Journal of Language Teaching and Research*, 1(6), 889–892. <https://doi.org/10.4304/jltr.1.6.889-892>
- Sundari, W., Siahaan, S. M., & Ismet, I. (2023). Digital teaching materials based on attention relevance confidence satisfaction substance pressure material. *Journal of Curriculum Indonesia*, 6(1), 69–75. <https://doi.org/10.46680/jci.v6i1.77>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10(1), Article 15. <https://doi.org/10.1186/s40561-023-00237-x>
- Tran, S., Krishna, P., Pakuwal, I., Kafle, P., Singh, N., Lynch, J., & Drori, I. (2021). *Solving machine learning problems*. ArXiv. <https://doi.org/10.48550/arxiv.2107.01238>
- Wood, D., & Moss, S. H. (2024). Evaluating the impact of students' generative AI use in educational contexts. *Journal of Research in Innovative Teaching & Learning*, 17(2), 152–167. <https://doi.org/10.1108/jrit-06-2024-0151>

Xu, B., & Ismail, H. H. (2024). The impact of artificial intelligence-assisted learning applications on oral English ability: A literature review. *International Journal of Academic Research in Progressive Education and Development*, 13(4), 1118–1134.

<https://doi.org/10.6007/ijarped/v13-i4/23352>

Yin, R. K. (2014). *Case study research design and methods*. Sage Publications.

Zabala-Vargas, S., García-Mora, L. H., Arciniegas-Hernández, E., Reina-Medrano, J. I., de Benito-Crosetti, B., & Darder-Mesquida, A. (2021). Strengthening motivation in the mathematical engineering teaching processes—A proposal from gamification and game-based learning. *International Journal of Emerging Technologies in Learning (iJET)*, 16(06), 4–19.

<https://doi.org/10.3991/ijet.v16i06.16163>

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators?

International Journal of Educational Technology in Higher Education, 16, Article 39.

<https://doi.org/10.1186/s41239-019-0171-0>

