International Journal of Evaluation and Research in Education (IJERE)
Vol. 15, No. 3, June 2026, pp. 2249~2260
ISSN: 2252-8822, DOI: 10.11591/ijere.v15i3.39041

Evaluating gamified learning strategies in internet of things-based software engineering education

Amneh Shaban[1], Arar Al Tawil[2]
[1]Department of Software Engineering, Faculty of Information Technology, Applied Science Private University, Amman, Jordan
[2]Department of Computer Science, Faculty of Information Technology, Applied Science Private University, Amman, Jordan

Article Info

Article history:

Received Feb 9, 2026
Revised May 5, 2026
Accepted May 7, 2026

Keywords:

Formative assessment
Gamification
Internet of things education
Learning analytics
Student engagement

ABSTRACT

This study examines the effectiveness of gamified formative assessment in undergraduate internet of things (IoT) education, focusing on how content complexity and question format influence student performance. A quasi-experimental comparative design was employed, administering two gamified quizzes to 75 undergraduate students enrolled in two IoT-related courses DevOps for IoT (n=41) and human computer interaction in IoT (n=34) during spring 2025. The gamified platform incorporated visual feedback, progress indicators, and interactive components. Results revealed statistically significant differences in student performance between the two quiz conditions, with human–computer interface (HCI) students substantially outperforming DevOps students. Question-level analysis further indicated that fill-in-the-blank formats impaired performance more than multiple-choice formats, and a pronounced ceiling effect was observed in the HCI assessment. These findings suggest that gamification effectiveness is contingent on alignment between content complexity, question format, and students' prior knowledge. Educators are advised to calibrate assessment difficulty and question types carefully when designing gamified learning experiences in technical education.

Corresponding Author:

Arar Al Tawil
Department of Computer Science, Faculty of Information Technology, Applied Science Private University
Amman, Jordan
Email: ar_altawil@asu.edu.jo

1. INTRODUCTION

In the last decade, gamification has gained considerable traction in engineering and computing education as a means of improving engagement and motivation in challenging technical disciplines [1]. The addition of game elements in non-game situations is referred to as gamification. According to a study by modelling learner responses in software engineering [2], adding game elements like role-plays had a good impact on cognitive skills, problem-solving and knowledge of the subject domain. Despite these benefits, it may not be effective in all learner populations or academic contexts, with various studies reporting varying effects, depending on the characteristics of students and the complexities of courses [3]. The measurement of sustainable impact of gamified interventions on learning outcomes is a still open methodological challenge [4]. Concerns like these are especially important when we look at rapidly evolving technical domains like internet of things (IoT) and DevOps which have more abstract concepts, hardware-software dependencies and continuous integration and continuous delivery/continuous deployment (CI/CD) pipeline complexities [5]. Confronting these obstacles seems to require novel teaching methods, which link theory and practical application while maintaining student engagement throughout the learning process [5].

The domain of IoT comprising embedded systems, sensor networks, cloud, computing, and data analytics require integrative understanding involving a number of disciplines [6]. In the same way, students' ability to internalize continuous integration workflows and infrastructure automation, which is hard to achieve via lectures, is critical in DevOps education [5]. As curricula are unable to keep pace with rapidly advancing technology, the need for experiential adaptive learning to reduce conceptual complexity will grow, without losing the attention of the learner [4]. The aim-oriented educational methods help to trigger the intrinsic motivators such as achievement, competition, and teamwork [5] and this can be achieved through game-based learning. The use of elements like points, badges, and leaderboards have been shown to enhance continuous motivation in software engineering courses [5]. However, there's not much empirical evidence regarding their effectiveness in IoT and DevOps domains in particular [7]. The formative assessment plays an important role in continuous feedback of students and monitoring learning [8]. Experiential assessments enhance learning while determining students' comprehension of curricular provisions. Also, through learning analytics techniques, faculty members can systematically collect and analyze data from the students' interactions, such as response times, attempt patterns, and accuracy levels, to identify areas of difficulty that can trigger adaptive lessons [9]. This analysis establishes a relationship between gamified behaviors and learning outcomes, thus increasing the validity and responsiveness of assessment design [10]. Nonetheless, the majority of current studies focus on the gamification effect on the general engagement metrics and do not look rigorously at how the complexity of the content and the type of questions interact and affect the outcome of a gamified assessment in a specific area of engineering [11]. This gap is particularly significant in IoT and DevOps education, where the tightly coupled hardware-software-network stack and the varied cognitive demands of different topic areas may substantially moderate gamification effectiveness. This study addresses this gap by empirically comparing gamified formative assessments across two IoT-related undergraduate courses of differing complexity levels, using learning analytics to examine how content difficulty, question type, and assessment format jointly influence student performance and engagement. The novelty of this work lies in its context-specific focus on IoT and DevOps curricula, its integration of fine-grained learning analytics with gamified assessment, and its explicit investigation of whether gamification attenuates or amplifies performance differences across students of varying ability levels. Therefore, this study is guided by the following research questions and corresponding hypotheses:

- How does gamified assessment impact student learning outcomes across IoT-related courses with different content complexity levels? (RQ1)
- What is the relationship between question difficulty and student performance in gamified assessment environments? (RQ2)
- How does question type (multiple-choice vs. fill-in-the-blank vs. open-ended) affect response accuracy in gamified assessments? (RQ3)
- Does gamification reduce performance variability among students of varying ability levels, thereby promoting more equitable learning outcomes? (RQ4)
- A positive association exists between gamified formative assessment engagement and learning outcomes, with the strongest gains occurring at moderate complexity levels (H1).
- Higher topic complexity is associated with lower response accuracy, though this relationship is attenuated within gamified environments through increased attempts and time-on-task (H2).
- Multiple-choice items yield higher mean accuracy and shorter response times than fill-in-the-blank and open-ended items, while open-ended items demonstrate greater discriminatory power for higher-order understanding (H3).
- Performance variability between lower- and higher-achieving students is reduced in gamified assessment settings, indicating a narrowing of achievement gaps relative to non-gamified contexts (H4).

2. LITERATURE REVIEW

2.1. Theoretical foundations of gamification

Scholars link gamification with various established theories that explains its motivational mechanism. According to the self-determination theory (SDT), gamification stimulates intrinsic motivation by satisfying the three basic psychological needs–autonomy, competence, and relatedness [12], [13]. In gamified learning environments, specific game elements are intentionally aligned with these needs, such as progress indicators and points which provide feedback on competence, branching learning pathways for autonomy, and leaderboards and collaborative challenges for relatedness [14]–[16]. When thoughtfully designed, gamified interventions have the potential to make done boring learning tasks more rewarding and sustain learner persistence [17]. These motivational constructs are further operationalized to relate to tangible measures of success, such as badges and experience points, which are linked to learners' internal goals and feedback about performance [18], [19].

Cognitive load theory can be seen as a complementary approach in which the mental effort required to engage in a learning task is managed. In the technical environments incorporated into games, the incorrect adjustment of levels of challenges or unknown question format leads to an extraneous cognitive load being imposed on the students, which can eradicate the motivational advantage of the game-like elements that were intended by the developers [5]. Flow theory adds to the framework with an optimal psychological state, which occurs when task difficulty matches learner skill level [20]. In combination, the three theoretical perspectives explicitly employed SDT, cognitive load theory and flow theory–propose a conceptualization to understand how gamification influences the complexity of the content and the design of assessments in technical education.

2.2. Gamification and student motivation in higher education

Across the different contexts of higher education, there is a consistent link between gamified learning strategies and an increase in student motivation and engagement [14], [21]. Research has shown that game elements such as points, badges, and leaderboards elicit competitive and cooperative drives, and this increased time-on-task and persistence [22]. A meta-analysis of gamification studies further confirmed that game-based interventions consistently yield positive effects on learning outcomes and student engagement across diverse educational settings [23]. A systematic review by Jaramillo-Mediavilla *et al*. [24] confirmed positive associations between gamification and academic motivation across disciplines. To further support the efficacy of gamification, Hellín *et al*. [14] report how university students perceived gamified environments to provide a greater value in their learning. Similarly, students also felt that gamified environments enhanced their intrinsic motivation. Gamification implementation leads to an increase in motivation, engagement, and achievement [15]. Wijaya *et al*. [16] similarly noted improvements in science. According to a systematic review of high school and higher education's gamified strategies [21], gamification leads to the right presence of motivation as long as it is considerate of course objectives and students' expectations and goals.

2.3. Gamification in IoT and software engineering education

Although the evidence base for gamification in higher education is growing, we know little about its use in technically complex areas like IoT and DevOps [25]. Past research studies on software engineering education have mainly examined generic engagement metrics without a strong link between gamified interactions and domain-specific competencies or learning outcomes [25]. Alhammad and Moreno [25] did a systematic mapping of gamification in software engineering education and identified a continuing gap between reported benefits in engagement and technical knowledge acquisition. Ngandu *et al*. [5] similarly pointed out that the software engineering gamification literature appears to disproportionately focus on the more surface aspects of engaging students and developing skills.

The education of IoT and DevOps has its own unique set of challenges unlike other computing fields. The multi-layer CI/CD workflows at the center of DevOps involve tight coupling of the hardware-software-network stack in IoT which is difficult to develop by merely using traditional assessment formats [5], [6]. Gordillo and López-Fernández [1] showed that active learning interventions in software engineering education had significantly better outcomes than traditional lectures, indicating that gamification has potential in these technically demanding contexts. Empirical assessments in engineering education have further confirmed that gamification can yield measurable improvements in both learning performance and student engagement when appropriately designed [26]. Nonetheless, according to researchers, penetrating assessment must be carried out in order to judge whether gamification can improve the transfer of complex problem-solving skill and practical knowledge in IoT and DevOps specifically [22], [27].

2.4. Assessment formats and gamified learning

The form of assessment items was found to be a significant moderating factor in gamified learning. Multiple-choice questions (MCQs) are often compatible with the fast feedback cycles of gamified systems, as the system can quickly reinforce scores and move the user forward [5]. Open-response types like fill-in-the-blank and open-ended questions place an increased cognitive load on users due to the requirement of recalling, rather than recognizing information, which may also interfere with the motivational flow of the gamified assessment [20]. According to Willig *et al*. [28], which type of question students are asked impacts how engaged students are in the context of gamification and clinical education. If the question was not relevant to the mechanics of the gamification, it decreased performance rather than improving it. Moreover, the reward system must be balanced: items that are too easy lose their diagnostic value, while items that are too difficult may cause frustration and lack of involvement, especially with high cognitive loads [5]. In this regard, AI-generated feedback has recently been proposed as a promising mechanism to enhance the quality and timeliness of assessment feedback in technology-enhanced learning environments [29].

Although much has been written about gamification in education, there are still crucial gaps. Most of the existing studies investigate gamification in general or introduction computing courses. There are not many studies concerning specific technical domains like IoT and DevOps [25]. Second, there have been few studies that systematically investigate the interactions that content complexity has on gamification effectiveness and how question format moderates gamified assessment in technical education [5], [20]. Furthermore, fine-grained learning analytics such as the response time, attempt pattern and question-wise accuracy have not been sufficiently characterized in terms of gamification effectiveness in an IoT education context [11]. The fourth issue related to equity outcomes i.e., whether gamification reduces or exacerbates performance differences across students of different ability levels has not received much attention in engineering education. This research tackles these voids head-on with an empirical comparative study of gamified formative assessment across 2 IoT-related courses of different content complexities.

3. RESEARCH METHOD

3.1. Research design

Utilizing a quasi-experimental comparative design, this study will examine whether a gamified formative assessment is effective for two undergraduate IoT-related courses of differing content complexity. In spring 2025, two different gamified quizzes were run with distinct groups of students using a between-subjects comparison approach. The design used was due to three complementary theories. First, SDT was used to understand how gamification mechanics can address the psychological needs of students with respect to autonomy, competence and relatedness. Second, cognitive load theory was useful in understanding how the type and level of content difficulty can influence learner performance under assessment conditions. Finally, the flow theory was used to evaluate the balance of challenge-skill in a gamified environment. By means of these frameworks, a multidimensional examination of both motivational and cognitive aspects of gamified assessment was possible.

3.2. Participants

The study included 75 undergraduate students (41 DevOps for IoT and 34 human–computer interface (HCI) in IoT students) from a medium-sized university that offers a dedicated IoT specialization track. Undergraduate students from fields of computer science, software engineering, and information technology (IT) were participants. Students enrolled into the course through convenience sampling. The inclusion criteria stipulated that student had to: i) enroll in the IoT specialization track; ii) register in one of the two target courses in spring 2025; and iii) complete the gamified quiz as part of the course. We excluded students that did not submit or did not consent to the research. All participants consented to the study and all data collection was anonymous with pseudonymous identifiers in use. Demographic factors such as gender were not systematically analyzed in this study; however, prior research suggests that gender may moderate gamified learning outcomes in computing education [30].

3.3. Course descriptions

The DevOps for IoT course educated students about advanced software engineering practices for the development and deployment of complex IoT architectures covering the use of CI/CD pipelines, over-the-air (OTA) updates, containerization, infrastructure as code, security integration. As per Bloom's taxonomy, application and analysis-level competencies were emphasized in the course and students had to integrate their knowledge in hardware, software, and network. The HCI in IoT course focused on designing user interfaces and experiences for IoT systems. It included accessibility, voice-enabled interfaces, gesture recognition and universal design. The course was heavily weighted to knowledge and comprehension level objectives and selected higher order thinking tasks, along with more concrete content students already interact with through technology.

3.4. Instruments

The two quizzes used a gamified learning application that was commercial off-the-shelf and had point scoring, visualization of progress in real time, automatic feedback on submission and time warnings. We disabled the leaderboard feature to avoid social comparison anxiety. The platform could be accessed from any device on standard web browsers with basic computer literacy.

The DevOps quiz consisted of ten questions, of which nine were multiple-choice (with four options each) which involved recognizing CI/CD practices, DevOps tools, and containerization concepts, and one fill-in-the-blank which required recall of technical terms. The HCI quiz consisted of eleven questions, ten of which were multiple choice covering areas such as the definition of HCI and principles of accessibility, and the interaction modalities of the IoT. The final word cloud question was an open-ended question that asked students what they already knew about the benefits related to gesture recognition. The instructors ensured that the quiz contents corresponded with course learning objectives. Due to resource constraints, we could not get the expert panel review, which is a limitation of the study.

3.5. Procedure

The quizzes were conducted during scheduled classes in computer labs with technical support at the university. Students were given brief standardized instructions on the quiz format, navigation, and time. Students did not have a specific time limit; thus, they were assessed under similar conditions at their own pace. Due to the first-response commitment design of the platform, students could not revisit the submitted questions. The formative assessment criteria for both quizzes contributed to students' final course grades providing enough motivational incentive without high-stakes anxiety.

3.6. Data collection

The platform recorded individual student performance and engagement data automatically. The total score obtained by the respondents in the quiz, expressed as a percentage of the questions answered correctly, was the main outcome measure. Other measures were the per-question accuracy (correct/incorrect/unattempt), time-on-task per question (in seconds, from display to submission), completion rates, and response sequence patterns. Details for individual questions (mean accuracy, mean completion time), as well as distractor selection statistics for MCQs, were exported in a format suitable for analysis.

3.7. Variables

The course condition (DevOps vs. HCI) served as the main independent variable, acting as a proxy for content complexity. The second independent variables question type (multiple-choice, fill-in-the-blank, open-ended) and question-level difficulty operationalized according to empirical accuracy rates. In terms of both the outcome and predictive measures, "time on task" functioned as that. The overall quiz score, the proportion of accurate responses per question, performance variability (coefficient of variation or $CV=\sigma/\mu\times100$), and pass rate (ratio of students who scored ≥60% are students who fail the quiz was our dependent variables).

3.8. Data analysis

Data analysis was conducted using Python 3.x with NumPy, Pandas, SciPy, and Matplotlib libraries. Statistical significance was set at $\alpha=0.05$ for all tests.

3.8.1. Descriptive statistics

Central tendency and dispersion were characterized through mean (μ), median (M), standard deviation (σ), variance (σ^2), range, interquartile range (IQR), and standard error (SE). Distribution properties were assessed via skewness and kurtosis coefficients. The CV was calculated as in (1). Pass rates were determined as the percentage of students achieving scores ≥ 60%.

$$CV = (\sigma / \mu) \times 100 \tag{1}$$

3.8.2. Normality and homogeneity tests

Distributional normality was evaluated using the Shapiro-Wilk test, under the null hypothesis that data follow a normal distribution. Variance homogeneity between groups was assessed using Levene's test, testing H_0: $\sigma^2_1=\sigma^2_2$.

3.8.3. Inferential statistics

Mean differences between quiz conditions were examined using the independent samples t-test in (2). The Mann-Whitney U test provided non-parametric validation. Effect size was quantified using Cohen's d in (3), where the pooled standard deviation is in (4). Confidence intervals (CI) (95%) were constructed using the t-distribution.

$$t = (\mu_1 - \mu_2) / \sqrt{}(\sigma^2{}_1/n_1 + \sigma^2{}_2/n_2) \tag{2}$$

$$d = (\mu_2 - \mu_1) / \sigma_{\text{poole}}{}^{\text{d}} \tag{3}$$

$$\sigma_{\text{poole}}{}^{\text{d}} = \sqrt{}[((n_1 - 1)\sigma^2{}_1 + (n_2 - 1)\sigma^2{}_2) / (n_1 + n_2 - 2)] \tag{4}$$

3.8.4. Question-level analysis

To understand the item level difficulties, accuracy rates and mean completion times were analyzed across all questions. To examine the relationship between time on task and response accuracy, Pearson's r and Spearman's p coefficients were calculated comparison across the different types of questions was done using stratified analysis.

3.9. Methodological limitations

It is important to note the various limitations of the research design. The findings may not be generalizable to other institutions owing to the convenience sampling approach. The two course groups were not only different in the complexity of content, but also instructor, teaching methodology, and student background, which introduces unmanageable confounds. Due to no pre-test measure of knowledge, it prevents us from attributing causes for differences in performance specifically to gamification. Furthermore, due to the limited demographic information collected, it was not possible to analyze the data by gender or academic year.

4. RESULTS AND DISCUSSION

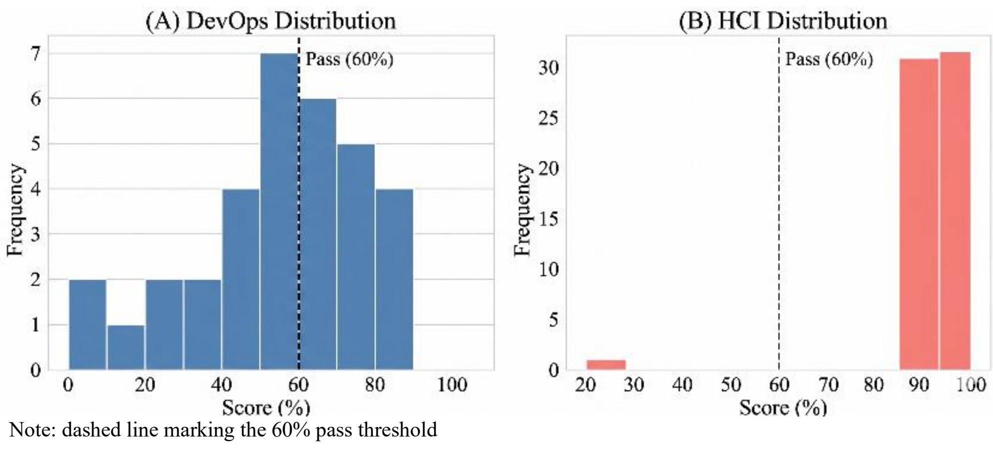
This part compares the gamified quiz outcomes of two IoT undergraduate courses. Statistical analyses detected significant differences in student scores on the DevOps and HCI assessments, which we discuss in relation to gamification design and content complexity.

4.1. Descriptive statistics and performance overview

The DevOps quiz was completed by 41 students and the HCI quiz by 34 students. As indicated in Table 1, the mean rate difference was highly significant. The average score on the DevOps test was 48.54% (SD=25.16, SE=3.93) and the scores ranged from 0% to 90% with a pass score of 41.46%. The distribution was moderately positively skewed (0.17) and platykurtic (kurtosis=-0.89), indicating a relatively flat spread with few extremes. Conversely, the average score for the HCI quiz was significantly high at 96.76% (SD=14.08, SE=2.42), with scores being between 20% and 100% and a pass percentage of 97.06%. The distribution was very negatively skewed (−4.98), indicative of a considerable ceiling effect where most students scored nearly perfectly. Figure 1 shows distributional differences in frequency histograms. In panel A, we show that the distribution of DevOps is roughly normal and center near the 60% threshold. In contrast, in panel B heavy concentration for HCI above 90% shows that 31 of 34 students score above this level.

Table 1. Descriptive statistics for DevOps and HCI quizzes

Statistic	DevOps	HCI
N	41	34
Mean	48.54%	96.76%
Median	50.00%	100.00%
Standard deviation	25.16	14.08
Variance	632.83	198.31
Minimum	0.00%	20.00%
Maximum	90.00%	100.00%
Range	90.00%	80.00%
Q1 (25th percentile)	30.00%	90.00%
Q3 (75th percentile)	70.00%	100.00%
IQR	40.00%	10.00%
Skewness	0.17	−4.98
Kurtosis	−0.89	24.22
CV	51.84%	14.55%
SE	3.93	2.42
Pass rate (≥60%)	41.46%	97.06%



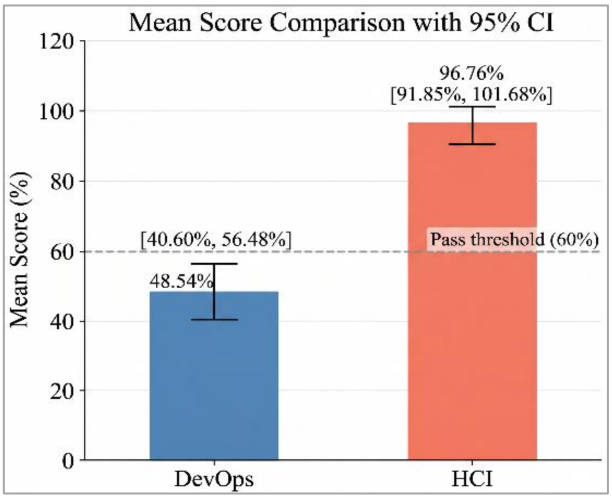
Note: dashed line marking the 60% pass threshold

Figure 1. Frequency histograms for DevOps (panel A) and HCI (panel B) scores

The CI of the DevOps and HCI means were found to be [40.60, 56.48] and [91.85, 101.68] respectively. There was a small overlap between the mean CI which suggests that there could be substantial differences at the population level, as shown in Table 2. The higher CV for DevOps (51.84%) compared to HCI (14.55%) shows that the scores in the DevOps condition exhibit more heterogeneity than the HCI condition. As shown in Figure 2, a bar chart with 95% confidence interval error bars indicates that the difference between the two conditions is 48.23 percentage points.

Table 2. Mean comparison with 95% CI

Quiz	Mean	Sd	Se	CI lower	CI upper	N
DevOps	48.54%	25.16	3.93	40.60%	56.48%	41
HCI	96.76%	14.08	2.42	91.85%	101.68%	34
Difference	48.23%	N/A	N/A	N/A	N/A	N/A



Non-overlapping error bars confirm statistically significant differences

Figure 2. Mean score comparison with 95% CI. DevOps mean: 48.54% (CI: 40.60–56.48%); HCI mean: 96.76% (CI: 91.85–101.68%)

4.2. Inferential statistical analysis

Assumptions about distribution were tested before comparing means. The Shapiro-Wilk test showed that the DevOps score can be assumed to be normal ($W=0.958$, $p=0.137$). The HCI score was significantly not normally distributed ($W=0.248$, $p<0.001$), consistent with the ceiling effect found in the descriptive analyses, as shown in Table 3. A Levene's test for homogeneity of variance of the three groups reveals a significant violation of the assumption ($F=27.41$, $p<0.001$) as shown in Table 4.

To support inferences, parametric analysis and non-parametric analysis were used. The t-test for independent samples indicated a statistically significant difference in group means ($t=-9.95$, $p<0.001$), with a mean difference of 48.23 percentage points in favor of the HCI condition. The Mann-Whitney U test confirmed that the performance differed without any assumptions on the distribution ($U=42.5$, $p<0.001$).

According to the report, Cohen's $d=2.31$ was substantially greater than the standard of a large effect size which is equal to or greater than 0.8. Thus, the analysis denotes substance and suggests practical significance. As shown in Figure 3, both the box plots in panel A and the violin plots in panel B indicate the distribution of the two quiz scores. The box plot shows large IQR, Q1=30%, Q3=70% and median at 50%.

Many outliers are also visible on the left. The HCI box plot reveals a narrow IQR close to 100% with notches tightly clustered. Violin plots show the complete probability density of the distributions. The DevOps violin is broad and approximately symmetric while the HCI violin is heavily peaked at 100%. This confirms the observed ceiling effect.

Table 3. Normality tests (Shapiro-Wilk)

Quiz	Test	Statistic	P-value	Normal	Alpha
DevOps	Shapiro-Wilk	0.9583	0.1366	Yes	0.05
HCI	Shapiro-Wilk	0.2482	<0.001	No	0.05

Table 4. Inferential statistics

Test	Statistic	P-value	Interpretation
Independent t-test	−9.953	<0.001	Significant
Mann-Whitney U	42.5	<0.001	Significant
Levene's test (Variance)	27.409	<0.001	Unequal variances
Effect size (Cohen's d)	2.309	N/A	Large

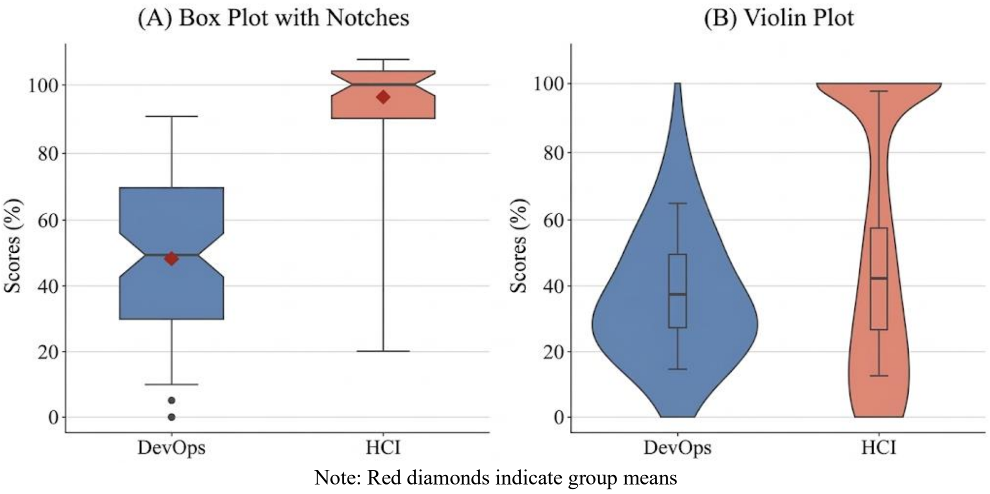


Figure 3. Score distributions for DevOps and HCI quizzes: box plots (panel A) and violin plots (panel B)

4.3. Question-level performance analysis

Insights into the difficulty and proficiency areas for both quizzes could be obtained through analysis. Table 5 provides accurate questions and timing of the DevOps quiz. The accuracy rates were between 41% and 66%, with a mean completion time of 15.44 seconds per question. This item (Q10), the single fill-in-the-blank item performed significantly worse, with 7% correct with mean time to complete was 85 seconds, which was the highest of all items. This item of a DevOps quiz was responsible for a significant drop in overall scores. This indicates that open-response type is particularly challenging in gamified technical assessment contexts.

Table 5. DevOps question-level performance

Question	Accuracy (%)	Avg time (sec)	Type	Correct	Incorrect
Q1	41.0	12	Multiple choice	17	22
Q2	51.0	12	Multiple choice	21	17
Q3	54.0	22	Multiple choice	22	17
Q4	51.0	17	Multiple choice	21	17
Q5	59.0	12	Multiple choice	24	14
Q6	51.0	12	Multiple choice	21	16
Q7	44.0	19	Multiple choice	18	20
Q8	66.0	15	Multiple choice	27	11
Q9	61.0	18	Multiple choice	25	12
Q10	7.0	85	Fill-in-the-blank	3	36

The data of HCI question is shown in Table 6. The performance on each item was consistently substantial, with accuracies for each multiple-choice ranging from 94% to 100% and a mean time of 23.50 seconds. The higher average completion time on DevOps multiple-choice items indicates that students were engaged in the items, not just guessing. All 34 students provided perfectly accurate answers to Q2, achieving 100% accuracy. The open-ended word cloud question (Q11) had 94% accuracy, taking on average 53 seconds to complete. This shows that engaging alternative question formats can work well if they are well-designed and are suitable to the course.

The performance by type of question on each quiz is summarized in Table 7. Average accuracy for the DevOps multiple-choice item was 53.11% while HCI was 96.70%. The fill in the blank item related to DevOps (7% accuracy) performed much worse than the word cloud item related to HCI (94% accuracy). This suggests that question format interacts with content complexity and student preparation. It does not mean one questions is more difficult than another in isolation. Figure 4 compares question-level accuracy on the two quizzes visually. The persistently elevated HCI bars juxtaposed with the fluctuating DevOps bars clearly highlight the significant dip observed at Q10 (fill-in-the-blank), reaffirming the negative influence of open-response format under conditions of high content complexity.

Table 6. HCI question-level performance

Question	Accuracy (%)	Avg time (sec)	Type	Correct	Incorrect
Q1	94.0	15	Multiple choice	32	0
Q2	100.0	20	Multiple choice	34	0
Q3	97.0	21	Multiple choice	33	1
Q4	97.0	17	Multiple choice	33	0
Q5	97.0	23	Multiple choice	33	0
Q6	97.0	27	Multiple choice	33	0
Q7	97.0	44	Multiple choice	33	0
Q8	94.0	20	Multiple choice	32	0
Q9	97.0	25	Multiple choice	33	0
Q10	97.0	23	Multiple choice	33	0
Q11	94.0	53	Word cloud	0	0

Table 7. Question type analysis

Quiz	Question type	N questions	Mean accuracy (%)	Mean time (sec)
DevOps	Multiple choice	9	53.11%	15.44
DevOps	Fill-in-the-blank	1	7.00%	85.00
HCI	Multiple choice	10	96.70%	23.50
HCI	Word cloud	1	94.00%	53.00

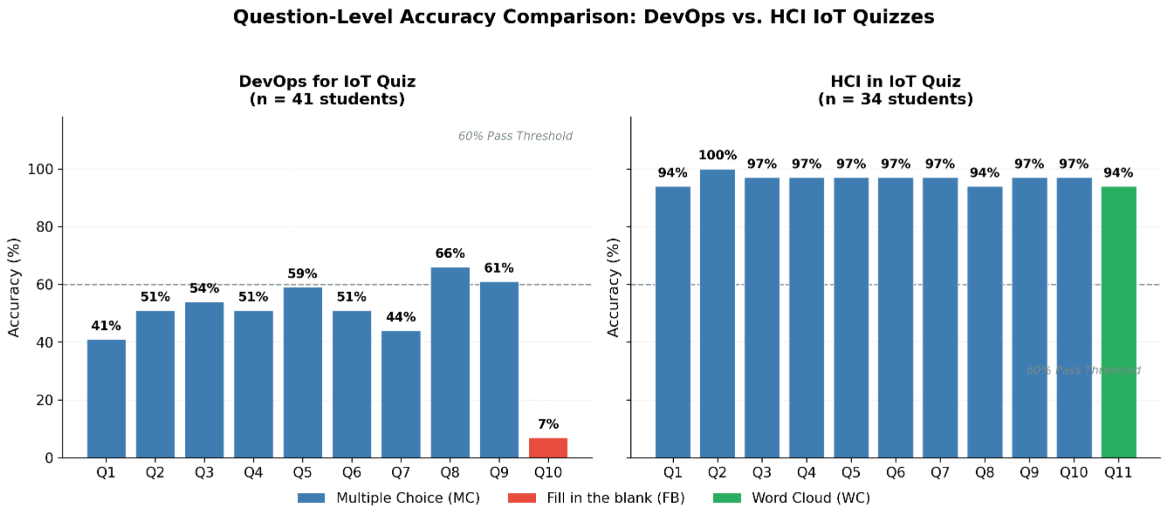


Figure 4. Question-level accuracy for DevOps (Q1–Q10) and HCI (Q1–Q11), highlighting Q10 fill-in-the-blank performance and the HCI ceiling effect

4.4. Relationship between time investment and accuracy

The correlation analyses of mean time on task and response accuracy are demonstrated in Table 8. There was a statistically significant negative Pearson correlation for the DevOps quiz ($r=-0.872$, $p=0.001$). This means that as the time to complete a question increases, the accuracy decreases. This pattern indicates how the extended response time reflects the difficulty, the cognitive struggle and not deliberation. The Spearman correlation ($\rho=-0.158$, $p=0.662$) was not significant which indicated a non-monotonic relationship likely due to the outlier fill-in-the-blank item (Q10) shown in Figure 4.

The relationships between punctuality and HCI quiz score as assessed by Pearson ($r=-0.221$, $p=0.513$) and Spearman correlation ($\rho=0.027$, $p=0.938$) were not statistically significant. The ceiling effect has caused the apparent absence of a time-accuracy relationship in a HCI task. Because most students achieved a maximum or near-maximum score no matter how quickly they responded, there was little variation in the accuracy scores and, therefore, no meaningful relationships could be detected as shown in Figures 2-4.

Table 8. Correlation analysis: time vs. accuracy

Quiz	Correlation type	Coefficient	P-value	Significant
DevOps	Pearson	−0.8717	0.0010	Yes
DevOps	Spearman	−0.1583	0.6624	No
HCI	Pearson	−0.2214	0.5129	No
HCI	Spearman	0.0268	0.9377	No

4.5. Discussion and implications

This study revealed that gamified assessment does not yield the same learning outcome in IoT courses of different content complexity. There was a performance gap of 48.23 percentage (Cohen's d=2.31) between the DevOps and HCI conditions, suggesting that content complexity is a strong moderating factor of gamification effectiveness. The findings disprove that gamification can act as a motivational intervention for all students, and call for context-sensitive assessment design in technical education.

The better performance of HCI students can be attributed to the relatively concrete, human-centered content of the HCI subject. The concepts of HCI are pulled from students' everyday experiences with technology interfaces to help students activate their prior knowledge. In contrast, the DevOps content needs abstract multi-domain knowledge from CI/CD pipelines, containerization, and infrastructure automation-areas where students had a high variance in their prior knowledge and their preparation, as shown in the wide distribution of scores in Figure 1. The performance difference could have been driven by instruction quality and student preparation, and not gamification. Those students may have more content familiarity and self-efficacy regarding the task, allowing them to benefit more fully from the gamified platform's motivational affordances. The restricted variance of HCI scores as shown in Figure 3, is supportive of this interpretation together with the fact that the cohort appeared to master the assessed material prior to being assessed.

The way questions are framed moderates the outcomes. The fill-in-the-blank item on DevOps (Q10) performed far worse than the multiple-choice items (7% vs. 41–66% correct) which is in accordance with theorizing derived from cognitive load theory, which predicts that open-response items impose a greater extraneous cognitive load as they require recall and active generation in time-pressured gamified environments. Conversely, open-ended formats can help keep student engagement high without increasing cognitive load when the material is neither too easy nor too difficult, and when designed with the relevant learning goal in mind.

These findings have significant implications for the design of gamified assessments in technical education. Instructors should calibrate the complexity of questions according to the level of knowledge expected of the students so that the gamification mechanisms can actually work as motivator supports and not just as a source of frustration. The strong negative correlation of time and accuracy for the DevOps condition ($r=-0.872$, $p=0.001$) suggests that the use of adaptive difficulty mechanisms, such as hint systems or tiered scaffolding, could enhance the learning experience remarkably in high-complexity gamified contexts. On the other hand, the ceiling effect exhibited in the HCI as shown in Figure 1, shows that it is not diagnostic behavior to use too easy an assessment, as the reduced variance in scores limits the ability to differentiate the mastery level of students. An effective gamified assessments could yield distributions showing both high scores and enough scatter for meaningful pedagogical decisions.

5. CONCLUSION

This research examined gamified formative assessment across two IoT undergraduate courses of varying content complexity. Findings indicate that gamification yields inconsistent outcomes depending on context. Content complexity, question format, and instructional preparation were identified as critical moderating factors. Results showed that higher content complexity negatively impacts student performance, as evidenced by low scores on technical DevOps items. Open-response formats proved unsuitable for high-complexity content, while multiple-choice and word cloud formats produced better outcomes. Gamification alone did not reduce performance variability without adequate prior knowledge. Effective gamified assessment requires careful calibration of difficulty, question format, and instructional alignment. Educators should incorporate adaptive features such as hint systems and progressive scaffolding, particularly in technically demanding courses. Study limitations include convenience sampling, uncontrolled confounds, absence of pre-testing, and a small single-semester sample. Future research should employ controlled experiments, adaptive difficulty algorithms, longitudinal designs, and qualitative methods to better understand learner experience and equity in gamified technical education.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Amneh Shaban	✓	✓		✓		✓	✓			✓		✓		
Arar Al Tawil	✓	✓	✓		✓	✓		✓	✓	✓	✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The authors confirm that the data supporting the findings of this study are available within the article.

REFERENCES

[1] A. Gordillo and D. López-Fernández, "Are educational escape rooms more effective than traditional lectures for teaching software engineering? A randomized controlled trial," *IEEE Transactions on Education*, vol. 67, no. 5, pp. 660–668, Oct. 2024, doi: 10.1109/TE.2024.3403913.

[2] A. Fatima *et al.*, "Exploring the use of gamification in human-centered agile-based requirements engineering," *Frontiers in Computer Science*, vol. 6, p. 1442081, Aug. 2024, doi: 10.3389/fcomp.2024.1442081.

[3] S. Tonhão *et al.*, "Gamification in software engineering education: a tertiary study," in *Proceedings of the XXXVII Brazilian Symposium on Software Engineering*, May 2023, pp. 358–367, doi: 10.1145/3613372.3614193.

[4] H. Cigdem, M. Ozturk, Y. Karabacak, N. Atik, S. Gürkan, and M. H. Aldemir, "Unlocking student engagement and achievement: the impact of leaderboard gamification in online formative assessment for engineering education," *Education and Information Technologies*, vol. 29, no. 18, pp. 24835–24860, Dec. 2024, doi: 10.1007/s10639-024-12845-2.

[5] M. R. Ngandu, D. Risinamhodzi, G. P. Dzvapatsva, and C. Matobobo, "Capturing student interest in software engineering through gamification: a systematic literature review," *Discover Education*, vol. 2, no. 1, p. 47, Nov. 2023, doi: 10.1007/s44217-023-00069-4.

[6] I. Alves and C. Rocha, "Qualifying software engineers undergraduates in DevOps-challenges of introducing technical and non-technical concepts in a project-oriented course," in *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET)*, 2021, pp. 144–153, doi: 10.1109/ICSE-SEET52601.2021.00024.

[7] I. K. Romanova, "Methodological aspects of the use of software in the teaching of engineering disciplines: tasks, problems and prospects," in *ITM Web of Conferences*, Dec. 2020, vol. 35, p. 04020, doi: 10.1051/itmconf/20203504020.

[8] G. Skedsmo and S. G. Huber, "Enhancing feedback practices and formative assessment in education: insights and challenges from diverse contexts," *Educational Assessment, Evaluation and Accountability*, vol. 36, no. 2, pp. 141–144, May 2024, doi: 10.1007/s11092-024-09436-7.

[9] T. A. B. Tretow-Fish and M. S. Khalid, "Methods for evaluating learning analytics and learning analytics dashboards in adaptive learning platforms: a systematic review," *Electronic Journal of e-Learning*, vol. 21, no. 5, pp. 430–449, Dec. 2023, doi: 10.34190/ejel.21.5.3088.

[10] E. Baran, D. AlZoubi, and A. S. Morales, "Design and implementation of an automated classroom analytics system: stakeholder engagement and mapping," *TechTrends*, vol. 67, no. 6, pp. 945–954, Nov. 2023, doi: 10.1007/s11528-023-00905-2.

[11] A. Giannakoulas and S. Xinogalos, "Studying the effects of educational games on cultivating computational thinking skills to primary school students: a systematic literature review," *Journal of Computers in Education*, vol. 11, no. 4, pp. 1283–1325, Dec. 2024, doi: 10.1007/s40692-023-00300-z.

[12] V. binti F. A. Chon, S. bte Sharif, and N. binti M. Saleh, "Gamification for enhancing students' learning motivation: a systematic review," *Journal of Cognitive Sciences and Human Development*, vol. 10, no. 2, pp. 19–41, 2024, doi: 10.33736/jcshd.7231.2024.

[13] M. Fodale, "Gamification in educational technology: a scoping review of trends and effectiveness," *Universal Journal of Educational Research*, vol. 4, no. 3, pp. 250–257, 2025, doi: 10.64637/629052.

[14] C. J. Hellín, F. Calles-Esteban, A. Valledor, J. Gómez, S. Otón-Tortosa, and A. Tayebi, "Enhancing student motivation and engagement through a gamified learning environment," *Sustainability*, vol. 15, no. 19, Sep. 2023, doi: 10.3390/su151914119.

[15] B. Mitchell and M. J. Co, "The impact of implementing gamification elements on motivation, engagement and academic achievement," in *Proceedings of the International Conference on Education Research (ICER 2024)*, pp. 184–193, Nov. 2024, doi: 10.34190/icer.1.1.3113.

[16] G. A. Wijaya, A. Z. Tantowi, and E. Nurizzah, "The effectiveness of gamification in enhancing student engagement in science learning," *International Journal of Mathematics and Science Education*, vol. 1, no. 2, pp. 33–37, May 2024, doi: 10.62951/ijmse.v1i2.90.

[17] Z. Zainuddin, S. K. W. Chu, M. Shujahat, and C. J. Perera, "The impact of gamification on learning and instruction: a systematic review of empirical evidence," *Educational Research Review*, vol. 30, p. 100326, Jun. 2020, doi: 10.1016/j.edurev.2020.100326.

[18] J. Luo, "Validating the impact of gamified technology-enhanced learning environments on motivation and academic performance: enhancing TELEs with digital badges," *Frontiers in Education*, vol. 9, p. 1429452, Nov. 2024, doi: 10.3389/feduc.2024.1429452.

[19] D. R. Wulan, D. M. Nainggolan, Y. Hidayat, T. Rohman, and A. Y. Fiyul, "Exploring the benefits and challenges of gamification in enhancing student learning outcomes," *Global International Journal of Innovative Research*, vol. 2, no. 7, pp. 1657–1674, Jul. 2024, doi: 10.59613/global.v2i7.238.

[20] H. Ateş and C. Ş. Kölemen, "Integrating theories for insight: an amalgamated model for gamified virtual reality adoption by science teachers," *Education and Information Technologies*, vol. 30, no. 2, pp. 2123–2153, Feb. 2025, doi: 10.1007/s10639-024-12892-9.

[21] E. Ratinho and C. Martins, "The role of gamified learning strategies in student's motivation in high school and higher education: a systematic review," *Heliyon*, vol. 9, no. 8, p. e19033, Aug. 2023, doi: 10.1016/j.heliyon.2023.e19033.

[22] M. Jun and T. Lucas, "Gamification elements and their impacts on education: a review," *Multidisciplinary Reviews*, vol. 8, no. 5, p. 2025155, Dec. 2025, doi: 10.31893/multirev.2025155.

[23] R. Mellado, C. Cubillos, R. M. Vicari, and G. Gasca-Hurtado, "Leveraging gamification in ICT education: examining gender differences and learning outcomes in programming courses," *Applied Sciences*, vol. 14, no. 17, 2024, doi: 10.3390/app14177933.

[24] L. Jaramillo-Mediavilla, A. Basantes-Andrade, M. Cabezas-González, and S. Casillas-Martín, "Impact of gamification on motivation and academic performance: a systematic review," *Education Sciences*, vol. 14, no. 6, p. 639, Jun. 2024, doi: 10.3390/educsci14060639.

[25] M. M. Alhammad and A. M. Moreno, "Gamification in software engineering education: a systematic mapping," *Journal of Systems and Software*, vol. 141, pp. 131–150, Jul. 2018, doi: 10.1016/j.jss.2018.03.065.

[26] J. Díaz-Ramírez, "Gamification in engineering education–an empirical assessment on learning and game performance," *Heliyon*, vol. 6, no. 9, p. e04972, Sep. 2020, doi: 10.1016/j.heliyon.2020.e04972.

[27] M. Covrig, S. I. G. Agoston, R. Ș. Igreț, C. V. Marinaș, A. D. Miron, and M. Roman, "Students' engagement and motivation in gamified learning," *Amfiteatru Economic*, vol. 25, no. SI 17, pp. 709–730, Nov. 2023, doi: 10.24818/EA/2023/S17/1003.

[28] J. H. Willig *et al.*, "Gamification and education: a pragmatic approach with two examples of implementation," *Journal of Clinical and Translational Science*, vol. 5, no. 1, p. e181, Jun. 2021, doi: 10.1017/cts.2021.806.

[29] I. Letteri and P. Vittorini, "Enhancing student feedback in data science education: harnessing the power of AI-generated approaches," *International Journal of Artificial Intelligence in Education*, vol. 35, no. 5, pp. 3071–3094, Dec. 2025, doi: 10.1007/s40593-025-00492-8.

[30] M. Li, S. Ma, and Y. Shi, "Examining the effectiveness of gamification as a tool promoting teaching and learning in educational settings: a meta-analysis," *Frontiers in Psychology*, vol. 14, p. 1253549, Oct. 2023, doi: 10.3389/fpsyg.2023.1253549.

BIOGRAPHIES OF AUTHORS

Amneh Shaban brings a wealth of experience and a proven track record of success in leading and enhancing educational programs as an esteemed BTEC lecturer specializing in software engineering. With a robust background as a lead internal verifier (IV), she ensures quality assessments, manages curriculum development, and implements innovative teaching methodologies. She holds a master's degree (M.Sc.) in Computer Science from the prestigious Applied Science Private University in Jordan, where her research focused on integrating ontology and data mining in smart homes. Passionate about integrating research-driven insights into education, she aims to empower students with the skills necessary to thrive in software engineering. She looks forward to contributing to the esteemed Faculty of Information Technology at Applied Science Private University in Amman. She can be contacted at email: a_shaban@asu.edu.jo.

Arar Al Tawil received his B.Sc. in Computer Science from Al-Hussein Bin Talal University, Jordan, in 2018, and his M.Sc. from the University of Jordan in 2021. He is currently a Ph.D. candidate in Computer Science at Applied Science Private University, Amman, Jordan, in collaboration with Universiti Sains Malaysia (USM). He serves as a lecturer and developer specializing in Virtual Reality and Game Design at the Faculty of Information Technology, Applied Science Private University. His research interests include virtual reality, augmented reality, machine learning, deep learning, natural language processing, and gamified learning environments. He can be contacted at email: ar_altawil@asu.edu.jo; arartawil@gmail.com.