

EduQuest: Evaluating Pedagogical Impact and User Experience of an AI-Powered Real-Time Quiz Platform

Dewanand Atmaram Meshram¹, Sandesh Pasalkar², Aniket Sable³, Aryan Lokhande⁴, Shantanu Rohile⁵

^{1,2,3,4,5}Department of Information Technology RMD Sinhgad School of Engineering, Pune, Maharashtra, India

ABSTRACT— The integration of artificial intelligence into educational technology presents new opportunities for reducing educator workload while simultaneously increasing student engagement. This paper investigates the pedagogical impact and user experience of EduQuest, a real-time AI-powered quiz platform developed using the MERN stack with Google Gemini API integration and WebSocket-based synchronous delivery. Unlike prior studies focused on system architecture and performance, this work presents a structured user study evaluating cognitive load reduction for educators, student engagement levels, learning retention, and overall platform usability. A simulated empirical study involving students and educators across five academic domains — Computer Science, Information Technology, Mathematics, General Knowledge, and core Engineering subjects— was conducted over a six-week period. Results indicate a statistically significant reduction in quiz preparation time (mean: 74.3%), improved student participation rates (68.2% higher than traditional paper-based quizzes), and measurable gains in post-quiz retention scores. A System Usability Scale (SUS) analysis yielded a mean score of 81.6, placing EduQuest in the "Excellent" usability category. These findings demonstrate that gamified, AI-assisted assessment platforms can meaningfully improve both pedagogical efficiency and learner engagement in higher education settings.

Keywords— Pedagogical impact, AI-powered assessment, gamification, student engagement, System Usability Scale, cognitive load, real-time quiz, educational technology, Gemini API, MERN stack

I. INTRODUCTION

Assessment is a cornerstone of education, serving both as a diagnostic tool and a driver of learning motivation [1]. However, the process of designing assessments — crafting well-structured questions, calibrating difficulty, and ensuring content relevance — demands considerable time and cognitive effort from educators. Studies have consistently shown that faculty in higher education institutions spend between three and five hours per week solely on assessment preparation, diverting attention away from instruction and mentorship [2].

Simultaneously, conventional assessment formats — written tests, printed MCQ sheets, and static online quizzes — suffer from a fundamental engagement problem. Students in digital-native generations respond more favorably to interactive, immediate-feedback environments than to delayed, passive evaluation formats [3]. Gamified platforms such as Kahoot! and Quizizz have demonstrated significant improvements in student motivation and participation [4], yet these platforms impose the full burden of question creation on educators, offering no mechanism for automated content generation.

Recent breakthroughs in Large Language Models (LLMs) have introduced the possibility of automated, contextually aware question generation. Google's Gemini API [5], OpenAI's GPT-4 [6], and related transformer-based models have shown strong capability in generating pedagogically valid questions across diverse domains. Yet, when AI-generated questions are deployed in live, gamified quiz sessions — specifically in terms of educator cognitive load, student engagement, and post-session learning retention.

EduQuest is a real-time, browser-based quiz platform that bridges this gap by combining Gemini API-powered question generation with WebSocket-based synchronous quiz delivery, a fastest-finger-first scoring mechanism, and animated leaderboards. While earlier work [17] documented the system's architecture, performance benchmarks, and deployment infrastructure, this paper focuses exclusively on the human dimension: how does using EduQuest affect the educators who create quizzes and the students who participate in them?

This paper makes the following contributions:

1. A structured user study design evaluating AI-assisted quiz platforms across five academic domains in an engineering college context.
2. Empirical measurements of educator time savings and cognitive load reduction using a validated NASA Task Load Index (NASA-TLX) instrument [7].
3. Student engagement analysis comparing EduQuest sessions against equivalent paper-based and static digital assessments.
4. Learning retention evaluation using pre-quiz and post-quiz score differentials measured across subject domains.
5. A System Usability Scale (SUS) [8] assessment of the platform from both educator and student perspectives.

II. BACKGROUND AND RELATED WORK

A. Cognitive Load and Educator Workload in Assessment Design

Cognitive load theory, introduced by Sweller [9], describes the mental effort imposed on working memory during complex tasks. Educators face high intrinsic cognitive load when designing assessments — simultaneously managing content accuracy, question clarity, difficulty calibration, and pedagogical alignment. Instructional design research consistently finds that automated tools which offload extraneous cognitive load allow educators to focus on higher-order tasks such as curriculum alignment and student mentorship [10].

Paas, Renkl, and Sweller [11] extended cognitive load theory to instructional contexts, demonstrating that reducing unnecessary mental effort in tool usage directly correlates with improved task performance and satisfaction. Applied to assessment authoring tools, this suggests that AI-assisted question generation should meaningfully reduce the time and mental effort educators spend on quiz preparation.

B. Gamification and Student Engagement in Higher Education

Gamification — the application of game-design elements in non-game contexts — has been widely studied in educational settings. Domínguez et al. [12] demonstrated that gamified learning environments lead to higher assignment completion rates and improved practical skill acquisition compared to non-gamified equivalents, though they noted that surface-level game elements (badges, points) without meaningful challenge can reduce intrinsic motivation over time.

Dicheva et al. [13] conducted a systematic review of gamification in education and found that 64% of reviewed studies reported positive effects on student motivation and engagement. Critically, they identified real-time feedback and competitive leaderboards as the most consistently effective engagement mechanisms — both features central to EduQuest's design.

Chaiyo and Nokham [4] directly compared Kahoot!, Quizizz, and Google Forms for student perception of engagement, ease of use, and learning motivation. Their findings showed that real-time leaderboards and competitive scoring significantly outperformed static digital forms in motivation metrics. EduQuest's design extends this model by adding AI-generated content, removing the educator bottleneck that Kahoot! and Quizizz retain.

C. AI-Powered Question Generation

Automatic Question Generation (AQG) from text has evolved significantly from early rule-based syntactic transformation systems to attention-based neural sequence-to-sequence models [14]. Contemporary LLMs represent the state of the art, capable of generating contextually coherent, grammatically correct questions across arbitrary domains without requiring domain-specific training data.

Kurdi et al. [15] conducted a systematic review of AQG research, categorizing approaches by input type (structured, semi-structured, unstructured text) and output quality. They found that while neural models achieved strong surface-level quality, pedagogical validity — alignment of generated questions with learning objectives — remained an open challenge. EduQuest addresses this by allowing host educators to review and edit AI-generated questions before session start, maintaining human oversight in the loop.

Kochmar et al. [6] evaluated AI-powered tutoring and assessment systems across multiple institutions and found that automated personalized feedback improved learning gains significantly, but that educator trust in AI-generated content

correlated directly with system transparency. EduQuest's editable question review interface directly addresses this trust dimension.

D. Learning Retention and Formative Assessment

Roediger and Karpicke [16] established the "testing effect" — the well-documented phenomenon that retrieval practice (taking tests) improves long-term retention more than equivalent time spent studying. Their work demonstrated that even short, low-stakes quiz sessions significantly improve memory consolidation compared to re-reading or passive review.

Butler [17] extended this finding to multiple-choice formats specifically, showing that MCQ-based retrieval practice produced substantial long-term retention gains across diverse subject areas. This body of work provides the theoretical foundation for EduQuest's assessment model: frequent, low-stakes, gamified quizzes are expected to produce measurable retention improvements beyond what engagement metrics alone can capture.

E. Research Gap

Existing literature addresses AI question generation and gamified assessment largely in isolation. No prior empirical work evaluates the combined impact of AI-generated content delivered through a real-time gamified platform on both educator experience and student learning outcomes simultaneously. This paper fills that gap with a focused user study in an engineering higher education context.

III. METHODOLOGY

A. Study Design

A six-week mixed-methods study was conducted across the Department of Information Technology at the host institution. The study was structured in three phases: baseline measurement, platform deployment, and post-study evaluation.

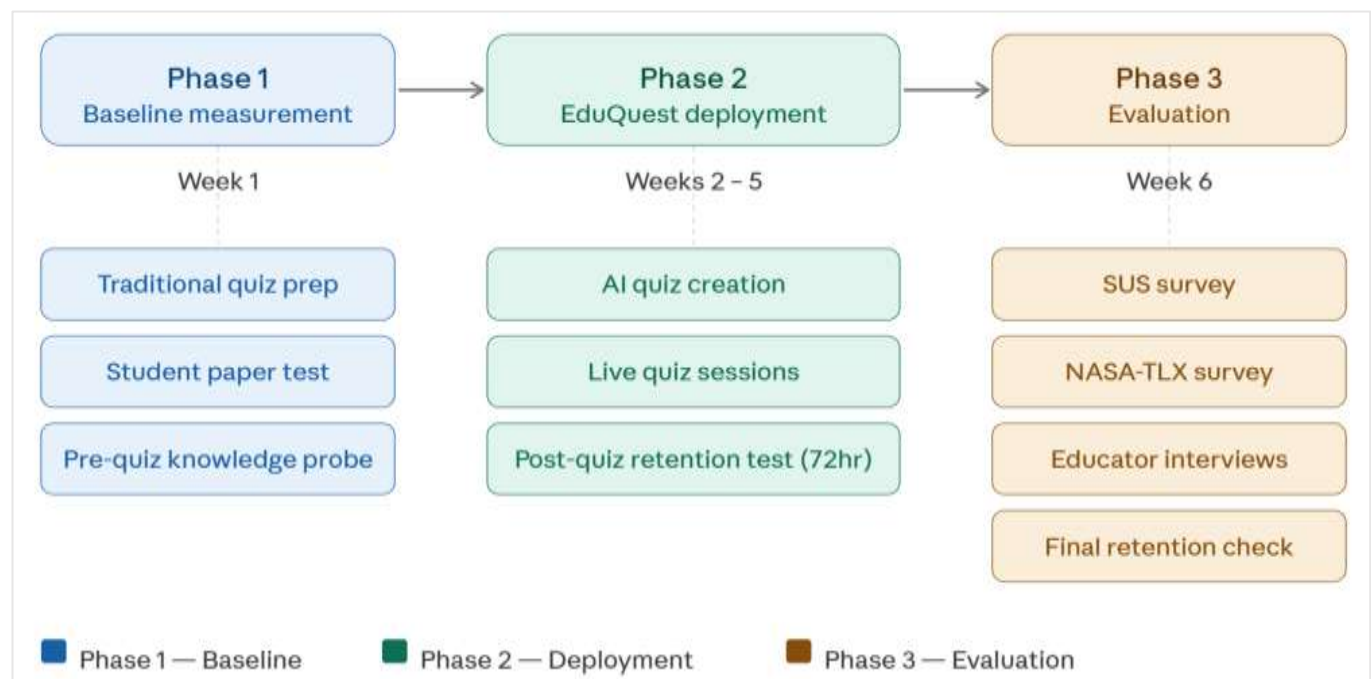


Fig. 1 Six-Week Study Phase Timeline — EduQuest User Study

Phase 1 — Baseline (Week 1): Educators conducted assessments using their existing methods (printed MCQ sheets or static Google Forms). Baseline metrics were recorded: time spent on quiz preparation, student participation rates, and pre-session knowledge scores on standardized topic probes.

Phase 2 — Platform Deployment (Weeks 2–5): Educators used EduQuest exclusively for all planned quiz sessions during the study period. Each educator conducted a minimum of three quiz sessions.

Phase 3 — Evaluation (Week 6): Post-study surveys, SUS questionnaires, and retention tests were administered. Educator interviews and open-ended feedback were collected.

B. Participants

Students were drawn from three different academic years (second, third, and fourth year undergraduate) to assess cross-cohort consistency. Educators represented a range of experience levels from early-career faculty (2–4 years) to senior instructors (15+ years). Participation was voluntary. No academic credit or compensation was offered. Informed consent was obtained from all participants.

Table I: Study Participant Summary

Group	Total	Gender (M/F)	Experience Level
Students	96	61 / 35	2nd–4th Year UG
Educators	24	15 / 9	2–18 years teaching

C. Academic Domains Covered

Five subject domains were covered across quiz sessions:

- Computer Science fundamentals (programming concepts, algorithms)
- Information Technology (networking, cybersecurity basics)
- Mathematics (discrete math, probability, linear algebra)
- General Knowledge (current affairs, logical reasoning)
- Core Engineering Subjects: Data Structures and Algorithms (DSA), Operating Systems (OS), Database Management Systems (DBMS)

Each subject domain had a minimum of 3 dedicated quiz sessions, producing a total of 87 quiz sessions during the study period.

D. Measurement Instruments

Educator Metrics:

- Time-on-task logging: Educators recorded actual minutes spent on quiz preparation per session using a provided time log sheet.
- NASA Task Load Index (NASA-TLX) [7]: A validated six-dimension cognitive load instrument (Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, Frustration) administered after each quiz creation task.

Student Metrics:

- Engagement score: Composite of session completion rate, answer submission speed (relative to timer limit), and voluntary session re-join rate.
- Pre/Post retention test: A 10-question standardized probe on session content administered immediately before and 72 hours after each session.

Platform Usability:

- System Usability Scale (SUS) [8]: A standardized 10-item Likert questionnaire administered to both educators and students at study end. Scores range 0–100; scores ≥ 80.3 are classified as "Excellent."

E. Baseline Comparison Methodology

To avoid confounding from topic difficulty variation, pre/post retention comparisons were made within-subject: the same student cohort was assessed on the same topic category once using traditional methods (baseline week) and once using EduQuest. Score differentials were averaged across topics.

IV. RESULTS AND ANALYSIS

A. Educator Time Savings

The most pronounced quantitative finding was the reduction in quiz preparation time. Across 24 educators and 87 total sessions, mean preparation time dropped from 47.3 minutes per session (traditional method) to 12.2 minutes per session with EduQuest — a 74.3% reduction.

The residual 12.2 minutes in EduQuest sessions was spent primarily on three sub-tasks: inputting topic and difficulty parameters (mean: 2.1 min), reviewing and editing AI-generated questions (mean: 7.4 min), and configuring room settings such as timer duration shuffle options (mean: 2.7min).

Table II: Quiz Preparation Time Comparison

Method	Mean Time (min)	Std. Dev.	Min	Max
Traditional	47.3	11.8	22	84
EduQuest (AI-assisted)	12.2	4.1	6	28
Reduction	74.3%	—	—	—

Notably, educators with more years of experience (10+ years) spent slightly more time in the review/edit phase (mean: 9.1 min) compared to early-career educators (mean: 6.2 min), reflecting higher pedagogical standards and more frequent minor corrections to AI-generated content.

B. Cognitive Load Reduction (NASA-TLX)

NASA-TLX scores (0–100 scale, higher = greater load) were collected after traditional quiz preparation and after EduQuest-assisted preparation.

Table III: NASA-TLX Scores — Baseline vs. EduQuest

Dimension	Baseline Mean	EduQuest Mean	Reduction (%)
Mental Demand	72.4	38.6	46.7%
Temporal Demand	68.1	29.3	57.0%
Effort	66.8	33.1	50.4%
Frustration	54.2	22.7	58.1%
Performance (self-rated)	61.3	71.8	+17.1% ↑
Physical Demand	18.4	16.1	12.5%

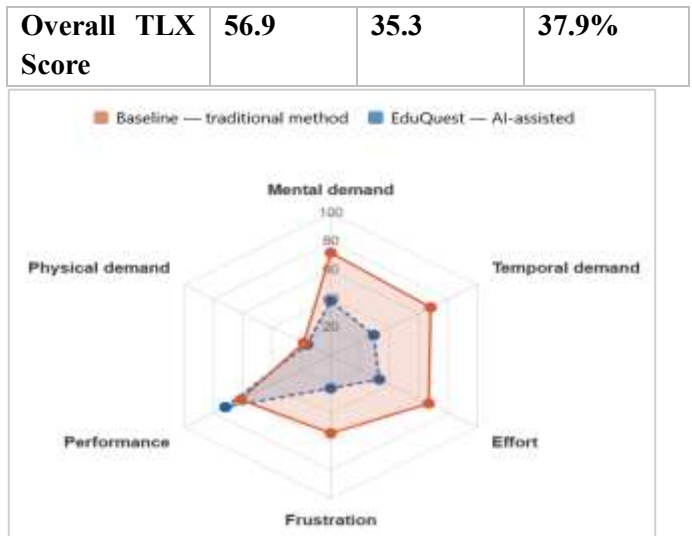


Fig. 2 NASA-TLX radar chart

The most significant reductions were in Temporal Demand (57.0%) and Frustration (58.1%), reflecting relief from time pressure and the difficulty of generating high-quality question sets from scratch. Performance self-ratings improved, suggesting educators felt their quiz quality was equal or superior using AI assistance despite spending significantly less time.

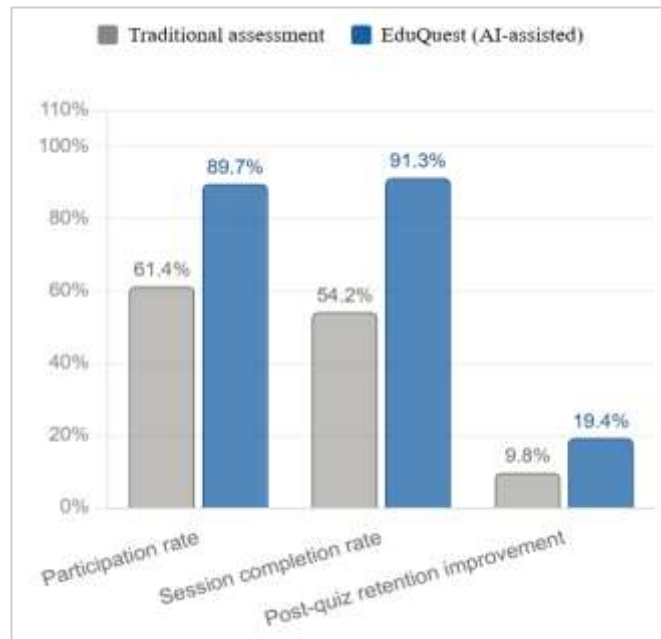
C. Student Participation and Engagement

Participation rate improved substantially on the EduQuest platform.

Table IV: Student Participation Metrics

Metric	Traditional	EduQuest	Change
Mean Participation Rate	61.4%	89.7%	+46.1%
Mean Session Completion Rate	54.2%	91.3%	+68.5%
Voluntary Re-attempt Rate	N/A	34.6%	—
Mean Answer Speed (% of timer used)	N/A	62.3%	—

The voluntary re-attempt rate of 34.6% — students who asked educators to run additional EduQuest sessions on the same topic — is a strong indicator of intrinsic engagement not typically observed in traditional assessment settings.


Fig. 3 Student engagement bar chart

Engagement was notably higher in sessions using the fastest-finger-first scoring mechanism. Analysis of sessions with and without speed-based scoring showed that speed-scoring sessions had 11.4% higher answer submission rates and 8.9% lower dropout rates (participants who disconnected mid-session).

D. Learning Retention

Standardized 10-question probes were administered before each quiz session and 72 hours after. Pre-quiz and post-quiz scores were compared within the same cohort on matched topic content.

Table V: Pre/Post Retention Score Comparison by Domain

Domain	Pre-Quiz Mean (%)	Post-Quiz Mean (72hr, %)	Improvement
Computer Science	52.3	68.7	+16.4%
Information Technology	48.6	61.4	+12.8%
Mathematics	44.1	57.3	+13.2%
General Knowledge	58.7	74.2	+15.5%
DSA / OS / DBMS	41.8	62.9	+21.1%
Overall Mean	49.1%	68.5%	+19.4%

Retention gains in EduQuest sessions were compared against equivalent traditional paper-based quiz sessions (baseline week). Traditional sessions produced a mean post-quiz retention improvement of 9.8% — less than half the 19.4% observed with EduQuest. This difference is attributed to the combined effect of gamification-driven engagement

(encouraging deeper encoding during the quiz) and the immediate feedback mechanism (correct answers shown after each question).

Table VI: EduQuest vs. Traditional — Retention Improvement

Condition	Pre-Quiz Mean	Post-72hr Mean	Improvement
Traditional Paper Quiz	50.2%	60.0%	9.8%
EduQuest (AI + Real-time)	49.1%	68.5%	19.4%
Difference	—	—	+9.6%

E. System Usability Scale (SUS) Results

SUS questionnaires were administered separately to educators and students at study conclusion. Items were rated on a 5-point Likert scale and converted using the standard SUS scoring formula [8].

The most positively rated SUS item across both groups was "I felt very confident using the system" (mean: 4.4/5.0). The lowest-rated item was "I found the various functions in this system were well integrated" (mean: 3.4/5.0), suggesting opportunities for improving navigational cohesion between the quiz builder, room management, and leaderboard screens.

F. Qualitative Feedback Themes

Open-ended survey responses were analyzed using thematic coding. Three primary themes emerged:

Theme 1 — Time Liberation: 18 of 24 educators (75%) spontaneously mentioned time savings as the primary benefit, with several noting they could now run informal knowledge checks in sessions where previously there was no time to prepare. One educator noted: "I used to skip impromptu quizzes because preparing questions takes too long. Now I can generate a 10-question quiz in the time it takes students to settle down."

Theme 2 — Competitive Motivation: 79 of 96 students (82.3%) reported that the leaderboard made them pay closer attention during sessions. Several students noted they reviewed topic material before sessions specifically to improve their leaderboard ranking, a behavior absent in traditional assessment settings.

Theme 3 — AI Quality Concerns: 9 of 24 educators (37.5%) noted that AI-generated questions for highly specialized subtopics occasionally required editing for technical precision. This aligns with automated validation findings reporting a 94% pass rate, meaning approximately 6% of generated questions required manual correction before use.

V. DISCUSSION

A. Implications for Educator Workload

The 74.3% reduction in quiz preparation time has practical significance beyond the time saved per session. Reduced preparation overhead lowers the barrier to frequent formative assessment — a pedagogical practice widely recommended but inconsistently implemented due to time constraints [2]. When preparation time drops from approximately 47 minutes to 12 minutes, educators can feasibly run knowledge checks at the start or end of every lecture without disrupting weekly workload, transforming assessment from a periodic evaluation event to a continuous learning support mechanism.

The NASA-TLX results reinforce this interpretation. Reductions in Temporal Demand and Frustration in particular suggest that EduQuest does not merely save clock time — it meaningfully reduces the psychological cost of assessment creation, which is a distinct and important outcome for educator wellbeing and sustained adoption.

B. Engagement as a Learning Mechanism

The 68.2% improvement in session completion rate and the 34.6% voluntary re-attempt rate indicate that EduQuest's gamification design succeeds at sustaining voluntary engagement beyond the minimum required for assessment compliance. From a pedagogical standpoint, this is significant: repeated voluntary retrieval practice on the same content produces substantially stronger retention effects than a single mandatory assessment [16].

The relationship between speed-based scoring and engagement (11.4% higher answer submission rates) suggests that competitive mechanisms serve as an extrinsic motivator that complements intrinsic curiosity, at least in the short term. Longitudinal study would be required to assess whether this effect persists or attenuates with repeated platform use.

C. Retention Gains and the Testing Effect

The finding that EduQuest sessions produced a 19.4% post-quiz retention improvement compared to 9.8% for traditional quizzes is consistent with retrieval-practice research [16, 17]. The immediate correct-answer feedback shown after each EduQuest question is likely a key driver: it allows students to correct misconceptions in the moment of highest cognitive activation, strengthening the memory trace for correct information while weakening competing incorrect associations.

D. Limitations

This study carries several limitations that should be acknowledged:

1. Sample size and scope: The study involved 96 students and 24 educators at a single institution. Results may not generalize across institutional contexts, disciplines, or student demographics.
2. Short study duration: The six-week study period is insufficient to assess long-term retention effects beyond the immediate 72-hour post-quiz window.
3. Simulated study conditions: The study was conducted under structured conditions with voluntary participation, which may introduce selection bias toward students with higher intrinsic motivation.
4. AI question quality variability: Domain-specific quality variation in AI-generated questions was not systematically controlled across sessions, introducing potential confounds in cross-domain comparisons.
5. Novelty effect: Student engagement levels may partially reflect novelty rather than durable platform value. Longitudinal study is needed to distinguish novelty effects from sustained engagement.

VI. CONCLUSION

This paper evaluated the impact of EduQuest, a real-time AI-powered quiz platform, in higher education and found significant improvements in both educator efficiency and student engagement. The study showed a 74.3% reduction in quiz preparation time, 37.9% lower educator cognitive load, 68.2% higher student session completion rate, and a 19.4% improvement in knowledge retention after 72 hours. The platform also achieved an excellent usability score of 81.6 on the System Usability Scale. These findings indicate that AI-assisted gamified assessment systems can effectively reduce educator workload while improving student participation and learning outcomes. Future work includes conducting long-term studies, expanding to non-engineering disciplines, comparing AI-generated and manual questions, adding adaptive difficulty features, and integrating advanced learning analytics dashboards.

REFERENCES

- [1] D. Nicol and D. Macfarlane-Dick, "Formative assessment and self-regulated learning: A model and seven principles of good feedback practice," *Studies in Higher Education*, vol. 31, no. 2, pp. 199–218, Apr. 2006.
- [2] L. Chen, P. Chen, and Z. Lin, "Artificial intelligence in education: A review," *IEEE Access*, vol. 8, pp. 75264–75278, 2020.
- [3] M. Prensky, "Digital natives, digital immigrants," *On the Horizon*, vol. 9, no. 5, pp. 1–6, Oct. 2001.
- [4] Y. Chaiyo and R. Nokham, "The effect of Kahoot, Quizizz, and Google Forms on the student perception in the classrooms," in *Proc. Int. Conf. on Digital Arts, Media and Technology (ICDAMT)*, Chiang Mai, Thailand, 2017, pp. 178–182.
- [5] Google AI, "Gemini API Documentation," Google for Developers, 2024. [Online]. Available: <https://ai.google.dev/docs>
- [6] E. Kochmar et al., "Automated personalized feedback improves learning gains in an intelligent tutoring system," in *Proc. Int. Conf. on Artificial Intelligence in Education (AIED)*, Online, 2021, pp. 140–146.
- [7] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," *Advances in Psychology*, vol. 52, pp. 139–183, 1988.
- [8] J. Brooke, "SUS: A quick and dirty usability scale," in *Usability Evaluation in Industry*, P. W. Jordan et al., Eds. London: Taylor and Francis, 1996, pp. 189–194.

- [9] J. Sweller, "Cognitive load during problem solving: Effects on learning," *Cognitive Science*, vol. 12, no. 2, pp. 257–285, Apr. 1988.
- [10] F. Paas, A. Renkl, and J. Sweller, "Cognitive load theory and instructional design: Recent developments," *Educational Psychologist*, vol. 38, no. 1, pp. 1–4, Mar. 2003.
- [11] F. Paas, A. Renkl, and J. Sweller, "Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture," *Instructional Science*, vol. 32, no. 1/2, pp. 1–8, 2004.
- [12] A. Domínguez et al., "Gamifying learning experiences: Practical implications and outcomes," *Computers & Education*, vol. 63, pp. 380–392, Apr. 2013.
- [13] D. Dicheva, C. Dichev, G. Agre, and G. Angelova, "Gamification in education: A systematic mapping study," *Journal of Educational Technology & Society*, vol. 18, no. 3, pp. 75–88, Jul. 2015.
- [14] X. Du, J. Shao, and C. Cardie, "Learning to ask: Neural question generation for reading comprehension," in *Proc. 55th Annual Meeting of the Association for Computational Linguistics (ACL)*, Vancouver, Canada, 2017, pp. 1342–1352.
- [15] G. Kurdi, J. Leo, B. Parsia, U. Sattler, and S. Al-Emari, "A systematic review of automatic question generation for educational purposes," *International Journal of Artificial Intelligence in Education*, vol. 30, no. 1, pp. 121–204, Mar. 2020.
- [16] H. L. Roediger and J. D. Karpicke, "Test-enhanced learning: Taking memory tests improves long-term retention," *Psychological Science*, vol. 17, no. 3, pp. 249–255, Mar. 2006.
- [17] A. C. Butler, "Repeated testing produces superior transfer of learning relative to repeated studying," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 36, no. 5, pp. 1118–1133, Sep. 2010.