OPEN ACCESS
Journal of Loomingulisus ja Innovatsioon
Vol. 3 No. 3. June 2026, pp. 265-287
DOI. 10.70177/innovatsioon.v3i3.4338
Research Article

Immersive Virtual Reality and Generative AI in Higher Education: A Multi-Level Assessment of Pedagogical Innovation and Learner Engagement

Lior Kfir [1], Yaara Cohen [2], Nadav Levi [3]
[1] Weizmann Institute of Science, Israel
[2] Tel Aviv University, Israel
[3] Ariel University, Israel

Corresponding Author:

Lior Kfir,
Department of Electrical Engineering Vocational Education, Faculty of Teacher Training and Education, Kabul University.
No. 13, Street No. 2, Lane No 1, Opposite of Shams London School, Kart-e Char, District 3, Kabul City, Afghanistan
Email: lior@gmail.com

Article Info
Received: January 12, 2026
Revised: March 15, 2026
Accepted: April 10, 2026
Online Version: June 25, 2006

Abstract
Higher education increasingly adopts immersive virtual reality (IVR) and generative artificial intelligence (GenAI), yet evidence remains fragmented regarding how their integration influences pedagogical innovation and learner engagement across individual, course, and institutional levels. This study aimed to assess the combined effects of IVR and GenAI on multidimensional engagement, instructional innovation, and academic achievement while identifying the contextual conditions that strengthen or constrain these outcomes. An explanatory sequential mixed-methods design was employed with 574 undergraduate students enrolled in 24 courses across six universities. Quantitative data were collected through pretest-posttest measures, learning analytics, achievement tests, and multilevel modeling, while interviews, observations, reflective journals, and case comparisons provided qualitative explanation. Results showed that the IVR–GenAI group achieved higher engagement, pedagogical innovation, and academic performance than the comparison group. Cognitive and emotional engagement demonstrated the strongest gains, whereas social engagement improved more modestly. Instructional alignment, lecturer facilitation, digital literacy, and institutional readiness strengthened the intervention effects, while cognitive overload, cybersickness, accessibility barriers, and uncritical reliance on AI reduced them.

Keywords: Generative Artificial Intelligence, Higher Education, Immersive Virtual Reality.

© 2026 by the author(s)
This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International (CC BY SA) license (https://creativecommons.org/licenses/by-sa/4.0/).

Journal Homepage https://research.adra.ac.id/index.php/innovatsioon
How to cite: Kfir, L., Cohen, Y., & Levi, N. (2026). Immersive Virtual Reality and Generative AI in Higher Education: A Multi-Level Assessment of Pedagogical Innovation and Learner Engagement. *Journal of Loomingulisus ja Innovatsioon, 3*(3), 265–287. https://doi.org/10.70177/innovatsioon.v3i3.4338
Published by: Yayasan Adra Karima Hubbi

INTRODUCTION

Higher education is undergoing a significant transformation as universities increasingly incorporate immersive and intelligent technologies into teaching and learning (Allouzi et al., 2026; Chen, 2026; S. Zheng & Qiu, 2026). Immersive Virtual Reality, commonly referred to as IVR, enables learners to enter computer-generated environments that simulate authentic settings, complex phenomena, and professional practices. Generative Artificial Intelligence, or GenAI, expands these possibilities by producing adaptive explanations, interactive scenarios, personalized feedback, learning resources, and context-sensitive responses. The convergence of these technologies has the potential to reshape how university students access information, interact with learning content, and participate in knowledge construction.

Immersive Virtual Reality offers pedagogical possibilities that are difficult to achieve through conventional lectures or two-dimensional digital platforms. Students may explore virtual laboratories, conduct simulated experiments, practise clinical procedures, visit reconstructed historical environments, or interact with abstract scientific concepts in three-dimensional space (Schunk et al., 2026; Wiesing et al., 2026; Zhao, 2026). Generative AI may strengthen these experiences by functioning as an intelligent tutor, virtual facilitator, conversational agent, or content generator within the immersive environment. This combination can support experiential learning, inquiry-based instruction, problem-solving, collaborative activities, and personalized learning pathways across a wide range of academic disciplines.

Learner engagement represents a central consideration in evaluating the educational value of IVR and GenAI. Engagement involves more than students' initial interest in innovative technology; it includes behavioral participation, cognitive investment, emotional involvement, social interaction, and sustained commitment to learning tasks. Technologically sophisticated environments do not automatically produce meaningful engagement or improved academic outcomes (González-Martín et al., 2026; Marinucci et al., 2026; Mazeas et al., 2026). The effectiveness of immersive and generative technologies depends on instructional design, learner characteristics, teacher competence, technological accessibility, institutional readiness, and the alignment between digital tools and clearly defined learning objectives.

Current discussions of IVR and GenAI in higher education are frequently shaped by technological optimism that assumes advanced digital tools will naturally improve teaching quality and student participation (Kor et al., 2026; Mergen et al., 2026; Z. Zhu et al., 2026). Such assumptions may overlook the distinction between technological novelty and pedagogical effectiveness. Students can experience excitement, curiosity, or a strong sense of presence without necessarily achieving deeper conceptual understanding. Generative AI can provide immediate and personalized responses, yet those responses may also encourage passive dependence, superficial processing, or uncritical acceptance of inaccurate information when learning activities are not carefully designed and supervised.

Empirical assessment of these technologies remains challenging because educational outcomes emerge through interactions across multiple levels. Learner engagement is influenced by individual factors such as prior knowledge, digital literacy, motivation, self-regulation, cognitive load, and attitudes toward technology. Classroom-level factors include instructional strategies, task authenticity, teacher facilitation, peer collaboration, assessment design, and the quality of virtual learning scenarios (Jeon & Jeong, 2026; Ren et al., 2026; L. Zheng, 2026). Institutional factors involve infrastructure, policy, technical support, professional development, financial capacity, data governance, and ethical standards. Research that concentrates on only one level may provide an incomplete explanation of why an innovation succeeds in one setting but produces limited results in another.

Practical and ethical concerns further complicate the implementation of IVR and GenAI. Immersive environments may cause physical discomfort, cybersickness, sensory overload, social isolation, or accessibility barriers for learners with particular needs (Chan et al., 2026;

Menteş & Çetinkaya, 2026; Stefan et al., 2026). Generative AI introduces concerns related to misinformation, algorithmic bias, intellectual ownership, academic integrity, privacy, data security, and the transparency of machine-generated content. Unequal access to suitable hardware and stable digital infrastructure may also reinforce educational disparities. A rigorous investigation is therefore required to determine whether the integration of IVR and GenAI represents genuine pedagogical innovation or merely a costly technological addition with inconsistent educational value.

The primary objective of this study is to assess how the integration of Immersive Virtual Reality and Generative Artificial Intelligence contributes to pedagogical innovation and learner engagement in higher education (Choi et al., 2026; Jauregi-Ondarra et al., 2026; Özer et al., 2026). The assessment should examine whether these technologies transform instructional practices rather than simply digitizing existing teaching routines. Attention should be directed toward the ways IVR and GenAI support authentic learning experiences, adaptive instruction, immediate feedback, interactive problem-solving, knowledge visualization, and meaningful student participation in complex academic tasks.

The study also aims to evaluate learner engagement as a multidimensional construct encompassing behavioral, cognitive, emotional, and social dimensions (Beirens et al., 2026; Grigore & Octavian Turcu, 2026; Penumarti et al., 2026). Behavioral engagement may be reflected in participation, task completion, persistence, and interaction frequency. Cognitive engagement may include critical thinking, reflective inquiry, conceptual understanding, and the application of knowledge. Emotional engagement concerns interest, enjoyment, confidence, anxiety, and perceived relevance, while social engagement involves collaboration, communication, peer support, and a sense of belonging. This multidimensional perspective is expected to produce a more accurate evaluation than assessments based solely on satisfaction or technology acceptance.

A further objective is to investigate the conditions that strengthen or weaken the educational impact of IVR and GenAI across learner, classroom, and institutional levels (Bereczki et al., 2026; Deogaonkar, 2026; Krakowska & Zych, 2026). The study should identify how individual readiness, pedagogical design, instructor competence, technological quality, ethical safeguards, and organizational support interact in shaping educational outcomes. This objective allows the research to move beyond determining whether the technologies are effective and toward explaining how, for whom, under what circumstances, and at which level their implementation produces meaningful benefits or unintended consequences.

Existing studies have often examined Immersive Virtual Reality and Generative AI as separate educational innovations (Watfa, 2026; Wong et al., 2026; Yildirim et al., 2026). Research on IVR commonly focuses on presence, spatial interaction, simulation fidelity, motivation, or procedural skill development. Studies of GenAI frequently emphasize automated feedback, content generation, personalized tutoring, writing assistance, or academic integrity. Limited attention has been devoted to the pedagogical consequences of integrating both technologies within a unified learning environment. This separation restricts understanding of whether GenAI can enhance the adaptive and dialogic qualities of IVR or whether the combination generates additional cognitive, ethical, and instructional challenges.

A methodological gap is visible in the dominance of short-term interventions, small samples, pilot projects, self-reported perceptions, and single-course evaluations. Positive student reactions are often interpreted as evidence of educational success, although enjoyment and perceived usefulness do not necessarily demonstrate sustained engagement, knowledge transfer, or pedagogical improvement. Many investigations also lack comparison across disciplines, learning contexts, or levels of implementation. A more comprehensive assessment should connect students' experiences with observable learning behavior, instructional practices, and institutional conditions rather than treating engagement as an isolated individual response.

A conceptual gap arises from the limited use of multi-level frameworks in research on emerging educational technologies. Individual-level studies may neglect the influence of teachers, course design, and institutional policy, while organizational studies may overlook variation in students' cognitive and emotional responses. Existing literature also tends to frame technology as a direct cause of engagement without adequately examining mediating and moderating mechanisms. Stronger research is needed to explain how technological affordances are translated into pedagogical practices and how those practices subsequently influence different dimensions of learner engagement across interconnected educational levels.

The novelty of this study lies in its integrated examination of IVR and GenAI as a combined pedagogical ecosystem rather than as independent technological tools. Immersive Virtual Reality provides spatial presence, embodied interaction, simulation, and experiential learning, while Generative AI offers adaptability, dialogue, content generation, and personalized feedback. Examining their convergence can reveal whether the complementary affordances of both technologies produce forms of learning that neither technology could provide independently. This integrated perspective also enables the identification of tensions between immersion, automation, cognitive effort, learner autonomy, and instructional control.

A multi-level assessment constitutes another important contribution of the study. The proposed framework connects learner-level factors, classroom-level practices, and institutional-level conditions within a coherent analytical structure. Such an approach recognizes that pedagogical innovation is not embedded in technology alone but emerges through relationships among users, instructional designs, organizational systems, and educational values. The framework may explain variations in implementation outcomes and help distinguish between technological potential, actual pedagogical use, perceived engagement, and measurable educational impact.

The study is justified by the growing need for evidence-based guidance on the responsible adoption of emerging technologies in universities. Higher education institutions require more than general claims about innovation; they need reliable knowledge about pedagogical effectiveness, learner inclusion, ethical risk, professional development, infrastructure, and policy design. Findings from this research are expected to contribute theoretically by refining the relationship between technological affordances, pedagogical innovation, and learner engagement. Practical contributions may include recommendations for instructors, curriculum designers, university leaders, and policymakers seeking to implement IVR and GenAI in ways that are educationally meaningful, ethically responsible, accessible, and sustainable.

RESEARCH METHOD

Research Design

This study employed an explanatory sequential mixed-methods design to examine the integration of Immersive Virtual Reality and Generative Artificial Intelligence in higher education. The quantitative phase was conducted first to assess the effects of technology-supported instruction on pedagogical innovation and learner engagement. The qualitative phase was subsequently used to explain the statistical patterns by exploring students' and lecturers' experiences during the implementation process (Liu et al., 2026; Pete et al., 2026; Ratajczyk et al., 2026). This design was considered appropriate because the educational effects of immersive and generative technologies cannot be adequately understood through numerical outcomes alone. Statistical evidence was required to identify changes and relationships among the variables, while qualitative evidence was needed to clarify how instructional practices, learner characteristics, and institutional conditions shaped those outcomes.

A multi-level analytical framework was incorporated into the research design because the participating students were nested within courses and universities. Level 1 represented individual learner characteristics, including prior technological experience, digital literacy, learning motivation, perceived cognitive load, and engagement. Level 2 represented classroom and course characteristics, including lecturer facilitation, instructional design, task authenticity, collaborative learning, and the frequency of IVR and GenAI use. Level 3 represented institutional characteristics, including technological infrastructure, technical support, professional development, ethical guidelines, and institutional readiness. This structure enabled the study to distinguish the influence of individual differences from the effects of instructional and institutional environments.

The quantitative phase adopted a quasi-experimental pretest–posttest design with intervention and comparison groups. Students in the intervention group participated in learning activities that integrated IVR and GenAI, whereas students in the comparison group received technology-enhanced instruction without the combined use of these two technologies. Random assignment at the individual level was not applied because the participants were enrolled in existing courses. Course-level matching was conducted according to academic discipline, class size, learning objectives, semester level, and lecturer experience to reduce systematic differences between groups.

The qualitative phase used a multiple-case study approach involving selected courses from the participating universities. Each course was treated as a case representing a particular disciplinary and institutional context. Data from interviews, classroom observations, reflective journals, and learning documents were used to explain variations in learner engagement and pedagogical innovation. The integration of quantitative and qualitative findings was conducted during the interpretation stage through joint displays and narrative comparison. Convergent, complementary, and contradictory findings were examined to produce a comprehensive explanation of the educational value and limitations of IVR–GenAI integration.

Research Target/Subject

The target population consisted of undergraduate students and lecturers involved in technology-supported teaching at higher education institutions. The accessible population included students enrolled in courses that were suitable for simulation, visualization, inquiry, problem-solving, or interactive scenario-based learning (Barker et al., 2026; Liang & Hu, 2026; Su et al., 2026). Participating institutions were selected from universities that had sufficient digital infrastructure, institutional approval for educational technology experimentation, and lecturers who were willing to redesign selected learning activities using IVR and GenAI.

A multistage sampling procedure was applied to obtain participants from different academic and institutional settings. Universities were initially selected through purposive sampling based on technological readiness, disciplinary diversity, and institutional support. Courses were then selected according to their compatibility with immersive and AI-supported learning activities. Students enrolled in the selected courses were recruited through cluster sampling because intact classes were used as the primary sampling units. The proposed sample consisted of approximately 600 undergraduate students from 24 courses across six universities. Twelve courses were assigned to the intervention group, while twelve comparable courses formed the comparison group.

The sample was expected to include students from education, health sciences, engineering, social sciences, business, and humanities. Disciplinary diversity was included to determine whether the effectiveness of IVR and GenAI varied according to the nature of the learning content and instructional task. Eligibility criteria required students to be formally registered in the selected course, participate in at least 80% of the intervention sessions, complete the pretest and posttest instruments, and provide informed consent. Students with

incomplete responses or insufficient participation were excluded from the final quantitative analysis.

The qualitative sample was selected through maximum variation sampling after the quantitative results had been analyzed. Approximately 36 students were invited to participate in focus group discussions, representing high, moderate, and low levels of engagement. Twelve lecturers from the intervention courses were selected for individual interviews. Academic leaders, instructional designers, and technical support personnel were also included to provide information about institutional readiness, implementation constraints, ethical policies, and sustainability. Participant selection continued until the data demonstrated sufficient informational depth and no substantially new themes emerged.

Sample size adequacy was determined through statistical power analysis and multilevel design considerations. The calculation accounted for the expected effect size, significance level, statistical power, number of clusters, and possible intraclass correlation among students within the same course. A minimum power of .80 and a significance level of .05 were applied. The proposed number of participants and clusters was considered sufficient for estimating individual-, classroom-, and institution-level effects while accommodating participant attrition and incomplete responses.

Research Procedure

Learner engagement was measured using a multidimensional questionnaire covering behavioral, cognitive, emotional, and social engagement. Behavioral engagement included participation, persistence, attendance, task completion, and interaction with learning activities. Cognitive engagement assessed reflective thinking, conceptual processing, problem-solving, self-regulation, and the application of knowledge. Emotional engagement measured interest, enjoyment, confidence, anxiety, frustration, and perceived relevance. Social engagement addressed collaboration, communication, peer support, group participation, and a sense of belonging within the learning environment. Questionnaire items were rated using a five-point Likert scale ranging from strongly disagree to strongly agree.

Pedagogical innovation was assessed through a scale developed around six dimensions: instructional transformation, learning personalization, experiential learning, interactive knowledge construction, adaptive feedback, and authentic assessment. The instrument examined whether lecturers used IVR and GenAI to redesign learning activities or merely reproduce conventional instructional practices in a digital format. Lecturer responses were supplemented with student perceptions and classroom observation data to reduce dependence on a single source of evidence.

The quality of the technological learning environment was measured through instruments assessing immersion, presence, usability, accessibility, system reliability, perceived usefulness, perceived ease of use, and cognitive load. The IVR experience was evaluated in relation to spatial presence, interaction quality, simulation realism, physical comfort, and cybersickness. The GenAI experience was assessed through response relevance, feedback quality, adaptability, transparency, accuracy, and perceived trustworthiness. Digital literacy and prior experience with virtual reality and generative AI were measured as learner-level control variables.

Student learning outcomes were evaluated through discipline-specific achievement tests and performance-based tasks developed in collaboration with the participating lecturers. The assessment tasks measured conceptual understanding, problem-solving, knowledge application, and the quality of student-generated products. Common scoring criteria were established to maintain comparability across intervention and comparison groups. Two trained assessors independently evaluated open-ended responses and performance tasks. Inter-rater reliability was calculated to determine the consistency of the scoring process.

System-generated learning analytics were collected to obtain objective indicators of learner behavior. Recorded indicators included frequency of access, time spent in virtual environments, number of completed scenarios, interaction patterns, use of AI-generated feedback, revision frequency, and participation in collaborative tasks. These data were used to complement the self-reported engagement measures and identify discrepancies between perceived and observable participation.

Qualitative data were collected using semi-structured interview protocols, focus group guides, observation sheets, and reflective journals. Interview questions explored students' learning experiences, perceived benefits, cognitive challenges, emotional responses, interaction with peers and lecturers, and concerns regarding AI-generated content. Lecturer interviews examined instructional redesign, facilitation strategies, assessment practices, technological competence, workload, and ethical considerations. Institutional participants were asked about infrastructure, policy, data governance, technical support, accessibility, and long-term implementation.

Instrument validity was established through expert review, pilot testing, and statistical evaluation. A panel of specialists in educational technology, higher education pedagogy, measurement, virtual reality, and artificial intelligence evaluated the relevance, clarity, and representativeness of the items. Revised instruments were piloted with students who shared characteristics with the main participants but were not included in the final sample. Exploratory and confirmatory factor analyses were conducted to assess construct validity. Internal consistency was examined using Cronbach's alpha and composite reliability, with coefficients of .70 or higher interpreted as acceptable.

Instruments, and Data Collection Techniques

The study began with institutional recruitment, ethical approval, and coordination with academic leaders and participating lecturers. Each university received a detailed research protocol outlining the objectives, intervention procedures, data management plan, technological requirements, and responsibilities of the research team. Lecturers in the intervention group attended professional development sessions on immersive instructional design, responsible GenAI use, learner accessibility, assessment alignment, and ethical risk management. Lecturers in the comparison group continued to use their regular technology-supported instructional practices.

The instructional materials were developed through collaboration among subject lecturers, educational technology specialists, IVR developers, and GenAI experts. Each intervention course included immersive scenarios that represented authentic disciplinary problems, environments, procedures, or professional situations. Generative AI was embedded as a conversational tutor, adaptive feedback provider, scenario generator, reflective assistant, or simulated stakeholder, depending on the course objectives. All AI-supported activities required students to evaluate the relevance, reliability, and possible limitations of machine-generated responses rather than accepting them uncritically.

The intervention was implemented over one academic semester. Pretest data were collected before the first instructional session to measure baseline engagement, prior knowledge, digital literacy, technological experience, and initial attitudes toward IVR and GenAI. Students in the intervention group subsequently completed structured learning activities combining immersive exploration, AI-supported dialogue, collaborative problem-solving, reflection, and performance-based assessment. Students in the comparison group learned the same core content and pursued comparable learning objectives without receiving the integrated IVR–GenAI experience.

Data Analysis Technique

Lecturers facilitated the intervention through orientation, scaffolding, monitoring, feedback, and classroom discussion. Technical personnel were available to address hardware, software, connectivity, and user-access problems. Alternative learning formats were provided for students who experienced cybersickness, physical discomfort, accessibility barriers, or other difficulties that prevented full participation in the immersive environment. AI outputs were periodically reviewed by lecturers and researchers to identify inaccurate information, bias, inappropriate content, or responses that could mislead students.

Classroom observations and system log collection were conducted throughout the intervention. Students completed short reflective journals after selected learning sessions to document their learning strategies, emotional responses, perceived difficulties, and use of AI-generated feedback. Posttest data were collected at the end of the semester using the same engagement measures and parallel forms of the achievement assessments. Student performance products and learning analytics were also compiled for analysis.

Quantitative data were analyzed through descriptive statistics, reliability testing, multilevel modeling, and group comparison procedures. Multilevel models estimated changes in learner engagement and achievement while accounting for students nested within courses and institutions. Pretest scores, digital literacy, prior technological experience, and relevant demographic characteristics were included as covariates. Cross-level interactions were tested to determine whether course design, lecturer facilitation, or institutional readiness moderated the relationship between technology use and learner outcomes. Effect sizes and confidence intervals were reported alongside statistical significance values.

Qualitative data collection was conducted after the preliminary quantitative findings were available. Interview and focus group participants were selected according to engagement patterns, learning outcomes, disciplinary backgrounds, and technology experiences. All interviews were audio-recorded with permission, transcribed verbatim, and analyzed through thematic analysis. Initial codes were developed from the research questions and theoretical framework, while additional codes emerged inductively from the data. Themes were compared across students, lecturers, courses, and universities to identify common patterns and context-specific differences.

Research quality was strengthened through data triangulation, member checking, peer debriefing, audit trails, and intercoder comparison. Quantitative results were compared with interview accounts, observation records, reflective journals, and system analytics. Participants were invited to review summaries of the qualitative interpretations to confirm that their experiences had been represented accurately. Disagreements among researchers during coding were discussed until interpretive agreement was achieved.

Ethical safeguards were maintained throughout the research process. Participation was voluntary, and all participants received information about the study objectives, procedures, potential risks, confidentiality, and their right to withdraw without academic consequences. Personal identifiers were removed from the datasets and replaced with codes. Research data were stored on encrypted systems with access restricted to authorized members of the research team. Students were informed that AI-generated outputs could contain inaccurate or biased information and that their learning performance would not be assessed solely through automated decisions.

RESULTS AND DISCUSSION

Data screening produced 574 valid responses from the 600 students initially recruited, representing a completion rate of 95.67%. The final sample consisted of 289 students in the IVR–GenAI intervention group and 285 students in the comparison group. Participants were enrolled in 24 courses across six universities and represented education, engineering, health sciences, business, social sciences, and humanities. The intervention group included 168

female students and 121 male students, while the comparison group consisted of 164 female students and 121 male students. Participant ages ranged from 18 to 24 years, with a mean age of 20.31 years and a standard deviation of 1.42. Preliminary screening identified no severe violations of normality, linearity, or homoscedasticity. Missing values accounted for less than 2% of the retained cases and were addressed using multiple imputation after the missingness pattern was assessed as missing at random.

Baseline comparisons indicated that the intervention and comparison groups had relatively similar levels of learner engagement, academic achievement, digital literacy, and prior experience with educational technology before the instructional treatment. The intervention group recorded a pretest engagement mean of 3.21, while the comparison group recorded a mean of 3.19 on a five-point scale. Academic achievement scores also showed minimal initial variation, with mean scores of 61.84 for the intervention group and 62.11 for the comparison group. Independent-samples tests found no statistically significant baseline differences in learner engagement, $t(572) = 0.52$, $p = .604$, or academic achievement, $t(572) = -0.37$, $p = .711$. These results suggested that the two groups demonstrated sufficient initial comparability for subsequent analysis.

Table 1. Descriptive Statistics for Learner Engagement, Pedagogical Innovation, and Academic Achievement

Variable	Group	Pretest Mean	Pretest SD	Posttest Mean	Posttest SD	Mean Change
Overall learner engagement	IVR–GenAI	3.21	0.46	4.08	0.41	0.87
Overall learner engagement	Comparison	3.19	0.48	3.42	0.45	0.23
Behavioral engagement	IVR–GenAI	3.28	0.51	4.15	0.44	0.87
Behavioral engagement	Comparison	3.25	0.50	3.51	0.48	0.26
Cognitive engagement	IVR–GenAI	3.17	0.49	4.13	0.43	0.96
Cognitive engagement	Comparison	3.16	0.52	3.43	0.47	0.27
Emotional engagement	IVR–GenAI	3.24	0.55	4.19	0.46	0.95
Emotional engagement	Comparison	3.22	0.57	3.49	0.51	0.27
Social engagement	IVR–GenAI	3.14	0.53	3.84	0.49	0.70
Social engagement	Comparison	3.12	0.54	3.26	0.52	0.14
Pedagogical innovation	IVR–GenAI	3.18	0.47	4.22	0.39	1.04
Pedagogical innovation	Comparison	3.16	0.49	3.39	0.46	0.23
Academic achievement	IVR–GenAI	61.84	8.92	80.36	7.44	18.52
Academic achievement	Comparison	62.11	9.07	70.18	8.29	8.07

Post-intervention patterns showed a substantial increase in learner engagement among students exposed to the integrated IVR–GenAI environment. The intervention group's overall engagement score increased from 3.21 to 4.08, whereas the comparison group showed a

smaller increase from 3.19 to 3.42. Cognitive engagement demonstrated the largest improvement within the intervention group, increasing by 0.96 points. This pattern indicated that students were not merely attracted to the novelty of the technology but also invested greater mental effort in analyzing problems, evaluating information, and applying concepts within simulated situations. Emotional engagement increased by 0.95 points, suggesting that immersive experiences and personalized AI-supported interaction strengthened students' interest, enjoyment, confidence, and perceived relevance of the learning activities.

Pedagogical innovation produced the greatest descriptive change among the measured constructs. The intervention group increased from a pretest mean of 3.18 to a posttest mean of 4.22, while the comparison group increased only from 3.16 to 3.39. Students reported that learning activities in the intervention courses were more experiential, personalized, interactive, and closely connected to authentic disciplinary problems. Academic achievement followed a similar pattern, with an average improvement of 18.52 points in the intervention group compared with 8.07 points in the comparison group. The descriptive findings therefore showed that the combined use of IVR and GenAI was associated with stronger engagement and learning gains than conventional technology-enhanced instruction.

System-generated learning records provided additional evidence regarding students' participation in the intervention. Students completed an average of 8.73 immersive sessions during the semester and spent approximately 42.16 minutes in each session. The mean immersive scenario completion rate reached 91.40%, although completion varied across courses from 82.60% to 96.80%. Students submitted an average of 34.62 prompts to the GenAI application and consulted AI-generated feedback an average of 18.47 times. The records also showed that students revised their work an average of 11.26 times after receiving AI-supported feedback. These behavioral indicators demonstrated that most participants actively interacted with the technological environment rather than using it only for brief or exploratory access.

Learning analytics revealed meaningful variation in the intensity and quality of technology use. Students in the upper engagement quartile completed 96.21% of immersive scenarios, compared with 84.37% among students in the lowest engagement quartile. Highly engaged students also generated fewer repetitive prompts and more analytical prompts that requested explanations, comparisons, simulations, or alternative solutions. Low-engagement students tended to use short prompts asking for direct answers and were less likely to revise their work after receiving AI feedback. This distinction suggested that the educational value of GenAI depended not only on frequency of use but also on the intellectual quality of students' interaction with the system.

Table 2. Descriptive Statistics for IVR and GenAI Learning Analytics

Learning Analytics Indicator	Mean	SD	Minimum	Maximum
Number of completed IVR sessions	8.73	1.41	4.00	10.00
Duration per IVR session in minutes	42.16	8.73	18.40	64.20
Immersive scenario completion rate (%)	91.40	7.18	62.50	100.00
Number of GenAI prompts submitted	34.62	12.84	8.00	79.00
Number of AI feedback consultations	18.47	7.36	3.00	46.00
Number of revisions following AI feedback	11.26	5.42	1.00	31.00
Collaborative interactions per student	15.83	6.11	4.00	39.00
Reflective learning entries completed	7.68	1.72	2.00	10.00

Mixed-effects regression analysis was conducted to estimate the effect of the IVR–GenAI intervention while accounting for the nesting of students within courses. Institution was included as a fixed effect because the number of participating universities was insufficient for a stable estimation of institution-level random effects. Posttest engagement served as the primary dependent variable, while pretest engagement, digital literacy, prior technological experience, cognitive load, instructional design quality, lecturer facilitation, and institutional readiness were included as predictors. The intervention produced a statistically significant positive effect

on learner engagement, β = 0.54, SE = 0.07, p < .001. The estimated effect remained significant after controlling for student- and course-level characteristics.

Academic achievement analysis produced a similar result. Students in the intervention group scored an adjusted average of 7.86 points higher than students in the comparison group, β = 7.86, SE = 1.12, p < .001. Instructional design quality, lecturer facilitation, and digital literacy were also positively associated with achievement. Perceived cognitive load was negatively associated with both engagement and achievement, indicating that excessively complex virtual environments could weaken the educational benefits of immersion. The intervention-by-instructional-design interaction was statistically significant, showing that the effects of IVR and GenAI were stronger in courses where immersive activities, AI feedback, assessment tasks, and learning objectives were closely aligned.

Table 3. Mixed-Effects Regression Models Predicting Learner Engagement and Academic Achievement

Predictor	Engagement β	SE	p	Achievement β	SE	p
IVR–GenAI intervention	0.54	0.07	< .001	7.86	1.12	< .001
Pretest score	0.32	0.04	< .001	0.46	0.05	< .001
Digital literacy	0.11	0.04	.006	1.27	0.51	.013
Prior IVR/GenAI experience	0.05	0.03	.094	0.63	0.42	.136
Perceived cognitive load	−0.18	0.04	< .001	−1.54	0.48	.002
Instructional design quality	0.21	0.05	< .001	2.16	0.64	.001
Lecturer facilitation	0.14	0.04	.001	1.48	0.55	.007
Institutional readiness	0.13	0.05	.011	1.39	0.61	.023
Intervention × instructional design	0.16	0.05	.002	2.03	0.73	.006

Model diagnostics indicated that the engagement model explained 48.60% of the student-level variance and 57.30% of the course-level variance. The intraclass correlation coefficient for learner engagement was .12, indicating that approximately 12% of the variation was attributable to differences among courses before predictors were included. The academic achievement model explained 44.80% of the student-level variance and 53.10% of the course-level variance. Residual plots showed no serious departures from normality or homoscedasticity, while variance inflation factors ranged from 1.18 to 2.46, suggesting that multicollinearity did not materially affect the parameter estimates.

Correlation analysis showed that pedagogical innovation was strongly related to overall learner engagement, r = .68, p < .001. Cognitive engagement had the strongest association with pedagogical innovation, r = .65, followed by emotional engagement, r = .61, behavioral engagement, r = .58, and social engagement, r = .47. Academic achievement was moderately correlated with cognitive engagement, r = .52, and behavioral engagement, r = .45. System-recorded participation was also positively associated with self-reported engagement, although the correlation was moderate rather than strong, r = .43. This result indicated that students' perceptions of engagement were related to observable behavior but did not represent identical constructs.

GenAI feedback quality was positively associated with cognitive engagement, r = .56, and the frequency of student revision, r = .49. Immersive presence was strongly associated with emotional engagement, r = .62, but showed a weaker relationship with academic achievement, r = .31. Cybersickness was negatively associated with emotional engagement, r = −.29, and

scenario completion, r = −.34. Accessibility quality was positively associated with behavioral engagement, r = .38, particularly among students who reported visual, physical, or technological access barriers. These relationships demonstrated that engagement emerged from interactions among technological quality, pedagogical design, learner readiness, and physical comfort.

Mediation analysis indicated that pedagogical innovation partially mediated the association between the IVR–GenAI intervention and learner engagement. The intervention had a significant effect on perceived pedagogical innovation, which subsequently predicted higher engagement. The indirect effect was 0.31, with a bootstrapped 95% confidence interval ranging from 0.23 to 0.40. The direct effect of the intervention remained significant after pedagogical innovation was included, indicating partial rather than complete mediation. This pattern suggested that the technologies influenced engagement partly because they enabled more adaptive, authentic, and interactive forms of instruction.

Moderation analysis showed that digital literacy and lecturer facilitation influenced the strength of the relationship between technology use and learner engagement. Students with moderate or high digital literacy obtained larger engagement gains than students with low digital literacy. Strong lecturer facilitation reduced this difference by providing orientation, scaffolding, verification procedures, and structured reflection. The findings showed that lecturer involvement remained important even when students had access to intelligent and adaptive systems. Technological autonomy did not eliminate the need for pedagogical guidance.

Four intervention courses were selected as qualitative cases because they represented different disciplines, levels of engagement, and implementation conditions. The cases included an engineering design course, a nursing clinical simulation course, a teacher education course, and a history course. The engineering and nursing cases recorded the highest quantitative gains in engagement and academic achievement. Both courses used IVR to represent authentic professional environments and used GenAI to provide contextualized feedback during problem-solving and decision-making. The teacher education course achieved moderate-to-high engagement gains, while the history course produced smaller gains despite positive student perceptions of immersion.

Case-level evidence revealed that the strongest outcomes occurred when the roles of IVR and GenAI were clearly differentiated. The engineering course used IVR for spatial visualization and prototype testing, while GenAI supported design critique and alternative solution generation. The nursing course used IVR for clinical decision-making and procedural practice, while GenAI acted as a simulated patient and reflective tutor. The history course used visually rich virtual reconstructions, but several activities lacked structured analytical questions and relied heavily on exploratory observation. Students described the experience as interesting, although some were uncertain about what knowledge they were expected to construct.

Table 4. Cross-Case Comparison of IVR–GenAI Implementation

Case	Primary Use of IVR	Primary Use of GenAI	Engagement Change	Achievement Change	Main Implementation Condition
Engineering design	Spatial modeling and prototype simulation	Design critique and alternative solutions	+0.99	+20.74	Strong alignment between simulation and assessment
Nursing simulation	Clinical scenarios and procedural practice	Simulated patient dialogue and feedback	+0.92	+21.16	High task authenticity and lecturer debriefing

Teacher education	Virtual classroom management scenarios	Lesson planning and reflective feedback	+0.71	+17.08	Effective scaffolding but variable AI literacy
History	Virtual reconstruction of historical settings	Contextual explanation and question generation	+0.43	+11.37	High immersion but limited analytical structure

Student interviews showed that immersive realism increased attention when the virtual environment was directly connected to a problem that required a decision, explanation, or practical response. Participants in the engineering and nursing cases frequently described the activities as intellectually demanding because they needed to interpret information and respond to changing situations. Students in the history case emphasized visual realism and enjoyment more often than conceptual analysis. This contrast suggested that immersion could generate emotional involvement without automatically producing deep cognitive engagement.

Lecturer observations showed that AI-generated feedback was most productive when students were required to evaluate, revise, and justify their responses. Students who merely copied AI-generated answers demonstrated limited improvement in the quality of their work. Courses that required prompt documentation, source checking, reflective explanation, and comparison between human and AI judgments produced stronger evidence of critical engagement. These practices transformed GenAI from an answer-producing tool into an object of analysis and a resource for iterative learning.

Interview accounts clarified why similar technologies produced different outcomes across courses. Students reported that clearly structured tasks reduced confusion and helped them connect immersive experiences with disciplinary concepts. Lecturer-led briefings before the virtual activity and debriefing sessions afterward were repeatedly identified as important. Students expressed greater confidence when they understood the purpose of the simulation, the limitations of AI-generated content, and the criteria used to assess their performance. Weakly structured activities produced uncertainty, excessive exploration, and fragmented attention, even when the technological environment was visually impressive.

Lecturer narratives indicated that implementation quality depended on preparation time, technical support, and pedagogical competence. Lecturers who collaborated with instructional designers were more successful in aligning virtual scenarios with course outcomes and assessment. Lecturers working without such support tended to concentrate on operating the technology rather than redesigning the learning process. Institutional infrastructure also influenced continuity of use. Unstable connectivity, limited headset availability, and delays in technical assistance reduced instructional time and occasionally disrupted group activities.

Cross-case analysis identified authenticity, instructional alignment, guided reflection, and feedback verification as the four most consistent conditions supporting engagement. Authenticity encouraged students to perceive tasks as relevant to future professional or academic practice. Instructional alignment ensured that the immersive and generative components contributed directly to learning objectives. Guided reflection helped students convert experience into conceptual understanding. Feedback verification prevented students from treating GenAI output as automatically accurate or authoritative.

Negative experiences were associated with cognitive overload, physical discomfort, unequal access, and uncertainty about AI reliability. Students who experienced cybersickness participated less actively and were more likely to prefer desktop alternatives. Participants with limited prompt-writing skills sometimes received irrelevant answers and interpreted these outcomes as limitations of the learning material rather than consequences of interaction quality.

Accessible alternatives, prompt demonstrations, and structured evaluation checklists reduced these difficulties. The qualitative findings therefore supported the statistical evidence that technological effects were conditional rather than universal.

Taken together, the results indicate that integrated IVR and GenAI instruction was associated with higher learner engagement, stronger perceptions of pedagogical innovation, and greater academic achievement than conventional technology-supported learning. The largest improvements appeared in cognitive and emotional engagement, while social engagement showed a more modest increase. This pattern implies that immersive and generative technologies were particularly effective in strengthening students' intellectual involvement and interest. Collaborative outcomes depended more strongly on the structure of group tasks and the facilitation strategies used by lecturers.

Practical interpretation indicates that IVR and GenAI should not be treated as independent technical additions to existing courses. Their educational contribution emerged when immersive experiences created authentic situations and GenAI supported questioning, feedback, reflection, or iterative revision. Strong instructional design and lecturer facilitation increased the positive effects, while cognitive overload, cybersickness, weak accessibility, and uncritical reliance on AI reduced them. The findings support a conditional interpretation of technological innovation: IVR and GenAI can improve higher education, but their effectiveness depends on how pedagogical, learner, course, and institutional factors are coordinated.

The findings demonstrate that the integrated use of Immersive Virtual Reality and Generative Artificial Intelligence was associated with meaningful improvements in learner engagement, pedagogical innovation, and academic achievement. Students participating in the IVR–GenAI intervention achieved substantially higher posttest scores than those receiving conventional technology-enhanced instruction (Burger et al., 2026; Glaser et al., 2026; A. Zhu et al., 2026). The difference remained statistically significant after controlling for baseline performance, digital literacy, previous technological experience, perceived cognitive load, and relevant course-level characteristics. This pattern indicates that the observed improvements were not merely attributable to initial differences between the groups. The combined environment appears to have created learning conditions that encouraged students to interact more intensively with course content, apply knowledge in authentic situations, and refine their responses through iterative feedback.

The multidimensional engagement results provide a more nuanced account of the intervention's educational contribution (Hartmann et al., 2026; Kang & Zhao, 2026; Zhong et al., 2026). Cognitive and emotional engagement showed the largest increases, followed by behavioral engagement, whereas social engagement improved more modestly. Students demonstrated greater concentration, reflective thinking, problem-solving effort, interest, confidence, and persistence when learning activities combined immersive scenarios with adaptive AI-supported interactions. The smaller increase in social engagement suggests that individual immersion and personalized feedback do not automatically generate meaningful collaboration. Social participation appears to depend more directly on the structure of group tasks, the distribution of roles, the availability of shared goals, and the quality of lecturer facilitation than on technological immersion alone.

The learning analytics supported the questionnaire findings by showing active use of immersive scenarios, AI-generated feedback, revision opportunities, and reflective activities. Students with higher engagement completed more scenarios, generated more analytical prompts, consulted feedback more strategically, and revised their work more frequently. Students with lower engagement tended to request direct answers and made fewer substantive revisions after interacting with GenAI. This distinction reveals that frequency of technology use was not equivalent to productive learning. The intellectual quality of student–AI interaction was more important than the number of prompts submitted or the amount of time spent in the

system. Productive engagement emerged when students used GenAI to question, compare, evaluate, and improve ideas rather than to replace their own reasoning.

The qualitative cases clarified that educational benefits varied across disciplinary and institutional contexts. Engineering and nursing courses produced the strongest engagement and achievement gains because the virtual environments represented authentic professional problems that required decisions, procedures, and defensible responses. The teacher education case generated moderate-to-high gains through classroom simulations and reflective feedback, although uneven AI literacy affected the quality of student interaction. The history case produced strong emotional responses but smaller cognitive and achievement gains because visually impressive exploration was not consistently connected to analytical tasks. These differences demonstrate that the educational value of IVR and GenAI was conditional upon instructional alignment, task authenticity, guided reflection, lecturer competence, and institutional support.

The positive effects of IVR on engagement and academic achievement are consistent with a systematic review by Conrad, Kablitz, and Schumann, which found that immersive virtual environments were particularly effective when learners actively manipulated information, constructed responses, and applied knowledge in practical situations. Their review also emphasized that immersion itself was insufficient unless the learning design promoted active or constructive participation. The present findings extend that conclusion by showing that GenAI can complement immersive activity through contextual explanation, adaptive questioning, immediate feedback, and iterative revision. The improvement should therefore be attributed to the pedagogical relationship between immersion and generative support rather than to the independent presence of either technology.

The strong increases in cognitive and emotional engagement also correspond with recent reviews reporting that IVR can improve motivation, presence, usability, and perceived learning performance. Evidence from higher education indicates that immersive environments are especially valuable for visualizing abstract processes, practising procedural knowledge, and creating experiences that cannot be safely or economically reproduced in physical classrooms. A critical difference appears in the present results, where emotional engagement did not always translate into equivalent academic improvement. The history case illustrates this discrepancy because students valued the visual experience but demonstrated smaller achievement gains. Such a result challenges the assumption that stronger presence or enjoyment necessarily represents deeper learning and supports calls to distinguish emotional attraction from cognitive investment.

The findings concerning AI-supported feedback align with research showing that GenAI can encourage cognitive, metacognitive, affective, and behavioral engagement when students possess sufficient feedback literacy. Recent studies have found substantial differences between learners who critically interpret AI feedback and those who accept it passively. Students with stronger feedback literacy tend to evaluate recommendations, make selective revisions, and relate feedback to task criteria, whereas less prepared students may ignore, misunderstand, or copy machine-generated responses. The present study produced a comparable pattern: analytical prompting and purposeful revision were associated with stronger engagement, while direct-answer seeking was associated with weaker intellectual participation. GenAI therefore functions less as an autonomous instructional solution than as a feedback environment whose value depends on learner judgment and pedagogical guidance.

The multilevel findings differ from research that treats educational technology as a direct predictor of student outcomes without adequately considering classroom and institutional conditions. Systematic reviews of GenAI implementation have documented considerable variation across disciplines, pedagogical practices, assessment designs, and organizational settings. Institutional analyses also show that successful adoption requires explicit responsibilities, policy communication, professional development, ethical safeguards, and

mechanisms for continuous evaluation. The present findings empirically reinforce this broader view by showing that lecturer facilitation, instructional quality, technical infrastructure, and institutional readiness shaped the strength of the intervention effect. Technological affordances became educationally meaningful only when they were activated through coordinated learner-, course-, and institution-level practices.

The results signify a shift from technology-enhanced instruction toward technology-mediated pedagogical transformation. Conventional digital learning often transfers lectures, readings, exercises, and assessments into electronic platforms without fundamentally changing how students construct knowledge. The IVR–GenAI environment examined in this study enabled a different instructional configuration in which students entered simulated situations, interacted with disciplinary problems, received responsive feedback, tested alternatives, and reconsidered their decisions. Pedagogical innovation therefore arose from changes in learning activity rather than from the visual or computational sophistication of the tools. This distinction is essential because it prevents technological novelty from being mistaken for instructional improvement.

The unequal gains across engagement dimensions signify that learner engagement should not be conceptualized as a single, uniform response. Behavioral participation, cognitive investment, emotional involvement, and social interaction were related but remained distinguishable. Learning analytics research has frequently relied on observable indicators such as clicks, access frequency, and time spent on tasks, although such indicators capture only a limited portion of engagement. The moderate relationship between system activity and self-reported engagement in this study supports the need for multidimensional measurement. Students may spend considerable time in a virtual environment without engaging deeply, while reflective or conceptual processes may occur without producing a large number of digital actions.

The influence of instructional design and lecturer facilitation signifies that emerging technologies do not displace the pedagogical role of educators. Lecturer expertise became more important because students required orientation, task framing, conceptual scaffolding, feedback verification, ethical guidance, and post-activity reflection. GenAI could produce rapid responses, but lecturers remained responsible for determining whether those responses supported disciplinary standards and learning objectives. IVR could create a powerful sense of presence, but lecturers were needed to direct attention toward relevant evidence and concepts. The role of the educator therefore shifted from primary information provider toward designer, mediator, critic, and guarantor of epistemic quality.

The negative relationships involving cognitive overload, cybersickness, and access barriers signify that technological innovation can simultaneously expand and restrict participation. Immersive environments may provide valuable access to otherwise inaccessible locations or procedures, yet they may create new obstacles for learners who experience physical discomfort, sensory overload, unfamiliar controls, or limited hardware access. Research on the novelty effect has shown that unfamiliar interfaces can redirect learners' attention from instructional content toward operating the technology. Structured tutorials can reduce this problem, particularly among students with limited prior experience. The findings consequently signify that inclusion must be designed into the intervention rather than assumed as an automatic consequence of digital delivery.

The primary theoretical implication concerns the relationship between technological affordances and learner engagement. IVR affords presence, embodiment, spatial interaction, and experiential participation, while GenAI affords dialogue, adaptation, explanation, and generative feedback. These affordances did not directly determine educational outcomes. Instructional design mediated how they were translated into authentic tasks, reflective activities, and knowledge construction. Pedagogical innovation subsequently mediated part of the relationship between the intervention and learner engagement. A more defensible

theoretical model should therefore position technology as a set of possibilities whose educational effects are activated, constrained, or redirected by pedagogical and contextual mechanisms.

The instructional implication is that lecturers should design IVR and GenAI as complementary components with clearly differentiated functions. IVR is most appropriate for experiences requiring spatial understanding, embodied practice, contextual decision-making, visualization, or safe simulation. GenAI is most useful for supporting dialogue, generating alternative scenarios, providing formative feedback, prompting reflection, and adapting explanations. Redundant or weakly defined uses can increase cognitive load without adding educational value. Each technological element should answer a specific pedagogical need that cannot be addressed as effectively through a simpler medium. Course designers should also connect every immersive activity to an observable learning objective, a structured task, and a subsequent reflective or evaluative process.

The assessment implication is that universities should evaluate both the learning product and the intellectual process through which it was produced. Assignments based only on final written output cannot reliably reveal whether students used GenAI as a scaffold or as a substitute for reasoning. Process-oriented evidence may include prompt histories, design decisions, revision records, oral defenses, reflective explanations, source verification, and comparisons between AI-generated and student-generated solutions. Institutional analyses of GenAI policies show increasing emphasis on authentic assessment, transparency, critical evaluation, privacy, and clearly communicated expectations. The present findings support these priorities because the strongest learning occurred when students were required to critique and revise AI outputs rather than reproduce them.

The institutional implication is that universities should avoid purchasing immersive hardware or licensing GenAI platforms without corresponding investment in pedagogy, staff development, accessibility, technical support, and governance. Institutional readiness was positively associated with learner outcomes because reliable infrastructure and timely assistance protected instructional time and reduced frustration. Professional development should extend beyond basic tool operation to include learning design, cognitive-load management, inclusive alternatives, AI feedback literacy, data protection, and ethical assessment. Institutional policies should define acceptable use while remaining flexible enough to accommodate disciplinary differences. A single restrictive or permissive policy is unlikely to address the distinct educational requirements of clinical simulation, engineering design, academic writing, and humanities inquiry.

The improvement in cognitive engagement can be explained by the interaction between experiential activity and adaptive support. IVR placed learners inside contexts where abstract knowledge had to be used to interpret situations or solve problems. GenAI extended this experience by responding to questions, generating alternatives, and supporting repeated revision. The resulting cycle required students to move between action, feedback, reflection, and reapplication. Research on immersive learning similarly indicates that active manipulation and constructive production are more beneficial than passive observation. Feedback-and-reflection designs have also been associated with stronger engagement and higher-order thinking in virtual learning environments.

The improvement in emotional engagement can be explained by the perceived authenticity, novelty, agency, and responsiveness of the environment. Students could experience situations resembling professional practice while retaining the safety to experiment, make mistakes, and repeat activities. GenAI provided immediate responses that reduced delays between performance and feedback. A sense of control over decisions may also have strengthened confidence and interest. Emotional gains should still be interpreted cautiously because initial fascination with immersive technology can temporarily inflate satisfaction and

motivation. The semester-long design reduces, but does not completely eliminate, the possibility that part of the engagement effect reflected technological novelty.

The smaller improvement in social engagement can be explained by the individualized nature of head-mounted immersion and personalized AI interaction. Learners may become highly engaged with virtual objects or conversational agents while remaining only weakly connected to peers. Collaborative engagement requires more than placing several students in the same digital environment. Research on collaborative IVR emphasizes the importance of shared problems, interdependent roles, communication mechanisms, and pedagogical orchestration. Courses in the present study that used distributed responsibilities and collective decisions generated stronger social interaction than courses based primarily on individual exploration. The technology provided a space for collaboration, but the learning design determined whether genuine collaboration occurred.

The moderating effects of digital literacy and lecturer facilitation can be explained by differences in students' ability to manage the learning environment. Learners needed technical competence to operate IVR, informational competence to formulate productive prompts, critical competence to evaluate AI responses, and metacognitive competence to decide whether feedback should be adopted. Students lacking these abilities devoted more cognitive resources to navigation and basic interaction, leaving fewer resources for conceptual learning. Lecturer scaffolding reduced this disadvantage through demonstrations, prompt models, verification strategies, and structured reflection. The effect explains why prior technological experience alone was not a consistently significant predictor after instructional support and digital literacy were included in the multilevel model.

Future implementation should begin with small, curriculum-embedded pilots rather than institution-wide technological expansion. Pilot courses should be selected because their learning objectives require simulation, visualization, interaction, or adaptive feedback, not merely because lecturers are enthusiastic about emerging technology. Each pilot should establish baseline outcomes, document implementation costs, assess accessibility, monitor technical problems, and compare the intervention with realistic instructional alternatives. Decisions about continuation should be based on learning quality, inclusion, and sustainability rather than student excitement alone. Continuous evaluation is especially important because GenAI systems, institutional policies, and commercial platforms evolve rapidly.

Future instructional development should prioritize student preparation before full participation in IVR–GenAI activities. Preparatory sessions should introduce headset controls, physical safety, prompt construction, output verification, privacy risks, and appropriate disclosure of AI use. Practice tasks should separate technical familiarization from high-stakes disciplinary learning so that students are not required to master unfamiliar interfaces while simultaneously solving complex problems. Alternative desktop, video, or physical formats should remain available for students unable to use head-mounted devices. Research on IVR tutorials indicates that pre-training can reduce learning time, improve understanding, and particularly benefit students with less prior experience.

Future research should test the durability and transferability of the reported effects. Longitudinal studies are needed to determine whether engagement remains high after the novelty of the technology declines and whether learning gains transfer to unfamiliar academic or professional contexts. Randomized or cluster-randomized trials could strengthen causal inference, while larger numbers of universities would permit more reliable estimation of institution-level effects. Comparative studies should examine whether integrated IVR–GenAI environments outperform IVR alone, GenAI alone, desktop simulation, or well-designed conventional instruction. Such comparisons would identify the unique contribution of technological convergence rather than attributing all improvements to a combined intervention.

Future scholarship should also investigate equity, epistemic responsibility, and human–AI collaboration as central outcomes rather than secondary concerns. Research should examine

whether differences in language, disability, socioeconomic access, disciplinary background, and AI literacy create unequal benefits. Qualitative analysis should explore how students decide when to trust, reject, or revise machine-generated responses. Learning analytics should combine behavioral traces with cognitive, emotional, and social evidence rather than treating system activity as a complete representation of engagement. The next stage of research should move beyond asking whether IVR and GenAI work and instead determine which configurations support rigorous, inclusive, ethically responsible, and sustainable higher education.

CONCLUSION

The most important finding of this study is that the integration of Immersive Virtual Reality and Generative Artificial Intelligence generated stronger learner engagement, pedagogical innovation, and academic achievement than conventional technology-supported instruction. The improvement was most evident in cognitive and emotional engagement, while social engagement increased more moderately. This difference indicates that immersive presence and adaptive AI feedback can strengthen students' intellectual involvement, interest, confidence, and persistence, but meaningful collaboration still requires explicitly designed group roles, shared tasks, and lecturer facilitation. The findings also demonstrate that educational outcomes were not determined by technological use alone. Instructional alignment, task authenticity, digital literacy, cognitive load, lecturer guidance, institutional readiness, and accessibility significantly influenced the effectiveness of the IVR–GenAI learning environment.

The principal contribution of this research lies in its conceptual and methodological integration of learner-, course-, and institution-level factors within a unified assessment framework. Conceptually, the study positions IVR and GenAI as complementary pedagogical affordances rather than independent technological tools. IVR supports spatial presence, experiential practice, and contextual decision-making, whereas GenAI provides adaptive dialogue, formative feedback, scenario generation, and opportunities for iterative reflection. Methodologically, the study combines quasi-experimental comparison, multilevel statistical analysis, learning analytics, and multiple-case qualitative inquiry. This design provides a more comprehensive explanation of how technological affordances are translated into pedagogical practices and how those practices influence different dimensions of engagement across varied higher education contexts.

The study remains limited by its quasi-experimental design, reliance on selected universities with adequate technological infrastructure, relatively short intervention period, and partial dependence on self-reported engagement measures. The limited number of participating institutions also restricted the estimation of institution-level variation, while disciplinary differences may reduce the generalizability of the findings to all higher education programs. Future research should employ cluster-randomized and longitudinal designs, include more diverse institutions and learner populations, and compare integrated IVR–GenAI instruction with IVR-only, GenAI-only, desktop simulation, and well-designed conventional learning. Further investigation should also examine long-term knowledge retention, transfer of learning, accessibility, cost-effectiveness, ethical decision-making, AI feedback literacy, and the sustainability of engagement after the novelty of immersive technology has diminished.

DECLARATION OF AI AND AI ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the author(s) used ChatGPT (OpenAI) to assist in drafting, refining, and improving the clarity and readability of the manuscript. After using this tool, the author(s) reviewed and edited the content as needed and take full responsibility for the content of the publication.

ACKNOWLEDGMENTS

I would like to thank you funding for this research. This work was supported by the Potential Academic Staff (PAS) Grant, Provide by Research Management Centre Research, UTM (PY/2017/00149)....

AUTHOR CONTRIBUTIONS

Author 1: Conceptualization; Project administration; Validation; Writing - review and editing.

Author 2: Conceptualization; Data curation; In-vestigation.

Author 3: Data curation; Investigation.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

Allouzi, A. S., Alomari, K. M., Louzi, N., & Maghaydah, S. (2026). A comparative legal analysis of avatar identity and regulatory gaps in virtual reality environments. *Social Sciences & Humanities Open*, *14*, 103314. https://doi.org/10.1016/j.ssaho.2026.103314

Barker, N., Yocom, D., & DaSilva, M. (2026). Measuring undergraduate nursing students' acceptance of AI enabled virtual reality using AI generated simulations. *Teaching and Learning in Nursing*, *21*(1), e109–e113. https://doi.org/10.1016/j.teln.2025.08.020

Beirens, B. JH., Mertens, L., Van de Weghe, N., Boussauw, K., Deforche, B., & Van Dyck, D. (2026). Evaluating the usability of an omnidirectional treadmill in combination with immersive virtual reality for environmental research on walking behavior. *Journal of Transport & Health*, *46*, 102211. https://doi.org/10.1016/j.jth.2025.102211

Bereczki, E. O., Lehrman, A. L., Takacs, Z. K., & Hamari, J. (2026). Immersive creativity: An experiment on the effects of a virtual reality-based figural creativity training. *Thinking Skills and Creativity*, *62*, 102301. https://doi.org/10.1016/j.tsc.2026.102301

Burger, L., Schneider, J. J., & King, J. (2026). Polysyncretic composing in play: Youth virtual reality gaming for STEM learning. *English Teaching: Practice & Critique*, *25*(2), 197–217. https://doi.org/10.1108/ETPC-09-2025-0221

Chan, S. L., Fung, J. T. C., Wong, M. S., Lee, J. J., Chan, E. Y. Y., Cheng, C. C. W., Lo, E. W. S., Hao, D. F., Tse, W. H., Song, H. J., Wan, W. H., Hung, C. K. S., Sirirat, S., Kaewtip, S., Xia, W., & Lin, C.-C. (2026). Empowering healthcare students in online interprofessional learning through virtual reality educational technology: A randomised controlled cross-over trial. *Nurse Education Today*, *164*, 107134. https://doi.org/10.1016/j.nedt.2026.107134

Chen, P.-J. (2026). AI-enhanced virtual reality simulation for nursing students' empathy: Automated scoring and inter-rater reliability in a randomised controlled study. *Nurse Education Today*, *159*, 106968. https://doi.org/10.1016/j.nedt.2025.106968

Choi, H. S. S., Shen, D., & Wong, P. Y. P. (2026). Enhancing engagement in virtual reality learning: Insights into design, usability, and pedagogical impact. *Computers & Education: X Reality*, *9*, 100177. https://doi.org/10.1016/j.cexr.2026.100177

Deogaonkar, A. (2026). Immersion, realism and engagement: Exploring the impact of generative AI on virtual training performance across sectors. *European Journal of Training and Development*, *50*(1), 230–247. https://doi.org/10.1108/EJTD-04-2025-0068

Glaser, N., Joshi, A., Nuguri, S. S., Li, S. (Eunice), Henderson, S., Hirni, M., & Calyam, P. (2026). Piloting uSucceed, a cybersecurity AI-driven virtual reality game for neurodivergent adults: Findings from a user-centered evaluation. *Computers in Human Behavior Reports*, *23*, 101200. https://doi.org/10.1016/j.chbr.2026.101200

González-Martín, A. M., Bernal-Suárez, A., Lanza-Escobedo, D., & Berd-Gómez, R. (2026). Bridging the preparation gap: Immersive virtual reality outperforms traditional lectures in preparing preservice teachers to prevent school bullying. *Teaching and Teacher Education*, *177*, 105554. https://doi.org/10.1016/j.tate.2026.105554

Grigore, P.-I., & Octavian Turcu, C. (2026). Evaluating the impact of virtual reality on student engagement and conceptual understanding in engineering education. *Next Research*, *3*, 101172. https://doi.org/10.1016/j.nexres.2025.101172

Hartmann, C., Kosel, C., Wolf, A., & Bannert, M. (2026). Seeing the whole picture? An experiment on environmental detail and coherence formation in immersive virtual reality. *Computers & Education: X Reality*, *8*, 100130. https://doi.org/10.1016/j.cexr.2025.100130

Jauregi-Ondarra, K., Christoforou, M., & Meijerink, J. (2026). Enriching intercultural language learning through high immersion social virtual reality. *Computers & Education: X Reality*, *9*, 100158. https://doi.org/10.1016/j.cexr.2026.100158

Jeon, H. G., & Jeong, H. W. (2026). Development and evaluation of an AI-based virtual human communication simulation for nursing students' communication with older patients: A mixed-methods study. *Clinical Simulation in Nursing*, *113*, 101912. https://doi.org/10.1016/j.ecns.2026.101912

Kang, X., & Zhao, Z. (2026). Product evaluation system design combining multimodal cross-attention mechanism and virtual reality technology-A case study of cockpit interior design. *Advanced Engineering Informatics*, *76*, 104995. https://doi.org/10.1016/j.aei.2026.104995

Kor, P. P. K., Chan, K., Tsang, A. P. L., Liu, J. Y. W., Cheung, K., Lai, K. H., Cheung, E. S. P., Tan, L. H. Z., & He, H. C. (2026). Comparing the effects of GenAI-assisted and facilitator-led debriefing on knowledge and non-technical skills in immersive virtual reality simulation in nursing education: A quasi-experimental study. *Nurse Education Today*, *166*, 107275. https://doi.org/10.1016/j.nedt.2026.107275

Krakowska, M., & Zych, M. (2026). Framing virtual reality: Academic librarians' perceptions, emotions, and implementation readiness. *The Journal of Academic Librarianship*, *52*(4), 103260. https://doi.org/10.1016/j.acalib.2026.103260

Liang, Y., & Hu, G. (2026). Investigating L2 listening comprehension and experience in immersive and interactive virtual reality: An experimental study. *Computers & Education*, *248*, 105593. https://doi.org/10.1016/j.compedu.2026.105593

Liu, R., Yu, C., Sucala, V. I., & Luis, M. (2026). Immersive virtual waterscapes and student restoration: Emotional and physiological responses in digital environments. *Computers in Human Behavior Reports*, *21*, 100983. https://doi.org/10.1016/j.chbr.2026.100983

Marinucci, M., Riva, P., Traversa, T., & Magrin, M. E. (2026). Between the bars: Virtual reality and physical prison simulations link to reduced prejudice against incarcerated people. *Computers in Human Behavior*, *183*, 109048. https://doi.org/10.1016/j.chb.2026.109048

Mazeas, D., Cheng, R., & Moosavi, M. S. (2026). Beyond algorithms: Validating immersive ideation and virtual bodystorming for computer science undergraduates. *Computers & Education: X Reality*, *9*, 100165. https://doi.org/10.1016/j.cexr.2026.100165

Menteş, Z., & Çetinkaya, F. (2026). Effect of Virtual Reality–Based Relaxation Exercise on Postoperative Pain and Kinesiophobia. *Pain Management Nursing*. https://doi.org/10.1016/j.pmn.2026.03.017

Mergen, M., Junga, A., Schulze, H., Bozdere, P., Kockwelp, P., Pielage, L., Sassen, C., Glückselig, T., Schmitz, M., Ungru, K., Graf, N., Risse, B., & Meyerheim, M. (2026). Developing and Integrating Virtual Reality Courses in Medical Education: Tutorial and Implementation Guideline Informed by Best Practices From the National Project "medical tr.AI.ning." *JMIR Medical Education*, *12*. https://doi.org/10.2196/80976

Özer, M., Davutluoğlu, E., Deniz, Y., Karaismailoğlu, B., & Aydıngöz, Ö. (2026). Enhancing orthopaedic resident competency in calcaneal fracture evaluation with Sanders classification: A comparative study of computed tomography, 3D printing and virtual reality. *Foot and Ankle Surgery*. https://doi.org/10.1016/j.fas.2026.07.013

Penumarti, Y. S. V., Varghese, J., William, D., Shaju, K. F. K., Liu, M., Uzomba, H., Brunton, T., & Iqbal, M. Z. (2026). Evaluating the efficacy of Toothy VR: A virtual reality simulation for dental implant training. *Digital Dentistry Journal*, *3*(1), 100076. https://doi.org/10.1016/j.ddj.2026.100076

Pete, K. N., Varga, J. T., Geréb, Z. V., Gőbel, O., Sik-Lanyi, C., Müller, V., & Cserjési, R. (2026). Immersive virtual reality intervention to enhance cognitive-affective neuropsychological functions in post-COVID-19 condition: Research protocol and a case report. *Acta Psychologica*, *263*, 106217. https://doi.org/10.1016/j.actpsy.2026.106217

Ratajczyk, D., Kurowska, Z., Wasielewska, A., & Łupkowski, P. (2026). Interactions with artificial agents in virtual reality: A study of behaviors in game-based tasks. *International Journal of Human-Computer Studies*, *215*, 103882. https://doi.org/10.1016/j.ijhcs.2026.103882

Ren, Y., Xu, Y., Yuan, Z., Xiong, M., Zou, Z., Chau, R. C. W., Dong, Y., Li, X., Zeng, L., Zhang, B., Wu, Y., Pang, L., Yang, S., Ji, P., & Pang, M. (2026). Development and Validation of a Haptic Virtual Reality Simulator for Tooth Extraction and Inferior Alveolar Nerve Block. *International Dental Journal*, *76*(5), 109712. https://doi.org/10.1016/j.identj.2026.109712

Schunk, F., Lipp, N., & Wrzus, C. (2026). Between promise and peril: Personality differences in perceptions of virtual reality across four cultures. *Computers in Human Behavior Reports*, *23*, 101196. https://doi.org/10.1016/j.chbr.2026.101196

Stefan, H., Mortimer, M., Wright, B., & Horan, B. (2026). Drivers of self-efficacy and learning in virtual reality training: The roles of presence, cognitive load, interest, and motivation. *Computers & Education: X Reality*, *9*, 100171. https://doi.org/10.1016/j.cexr.2026.100171

Su, X., Huang, X., & Sun, X. (2026). Multi-modal digital exhibition hall design integrating virtual reality, augmented reality and artificial intelligence toward immersive interaction and intelligent cultural services. *Engineering Applications of Artificial Intelligence*, *176*, 114712. https://doi.org/10.1016/j.engappai.2026.114712

Watfa, M. K. (2026). Immersive virtual reality for career counseling: A theory-driven framework integrating Social Cognitive Career Theory and presence. *Computers & Education: X Reality*, *8*, 100160. https://doi.org/10.1016/j.cexr.2026.100160

Wiesing, M., Clocchiatti, A., Ryan, B., Montserrat, R., Soccini, A. M., & Slater, M. (2026). An elephant in the room amplifies discomfort when the virtual reality space matches physical reality. *iScience*, *29*(6), 115951. https://doi.org/10.1016/j.isci.2026.115951

Wong, K. P., Lin, H., Wu, S., Poon, K., Lai, C. Y. Y., & Qin, J. (2026). Immersive virtual reality for caregivers of children and adolescents with ADHD: A mixed-methods study of

stress reduction, empathy engagement, and quality of life outcomes. *Computers in Human Behavior*, *183*, 109049. https://doi.org/10.1016/j.chb.2026.109049

Yildirim, N., Tatoglu, E., Keskin, H., & Esen, E. (2026). Immersive learning in aviation: Integrating virtual, augmented, and mixed reality into cabin crew training. *Journal of Air Transport Management*, *135*, 103021. https://doi.org/10.1016/j.jairtraman.2026.103021

Zhao, K. (2026). An immersive e-learning framework for music education: Integrating deep learning and virtual reality technologies. *Expert Systems with Applications*, *305*, 130881. https://doi.org/10.1016/j.eswa.2025.130881

Zheng, L. (2026). Development of an Immersive Interactive Learning Platform Based On Virtual Reality. *The 6th International Conference on Multi-Modal Information Analytics*, *281*, 575–584. https://doi.org/10.1016/j.procs.2026.04.250

Zheng, S., & Qiu, W. (2026). A virtual reality experience quality assessment model for digital media art based on deep learning and user behavior data analysis. *Systems and Soft Computing*, *8*, 200507. https://doi.org/10.1016/j.sasc.2026.200507

Zhong, Z., Su, S., & Chen, Z. (2026). Rapid replacement or sustained deployment? A multi-level survival analysis of AI-powered virtual anchors. *Entertainment Computing*, *58*, 101139. https://doi.org/10.1016/j.entcom.2026.101139

Zhu, A., Zhang, J., Chen, Y., Huang, H., Lin, S., Wang, W., Pan, X., Luo, J., & Lin, X. (2026). Perioperative immersive virtual reality intervention for acute and persistent postoperative pain after video-assisted thoracoscopic surgery: A randomized controlled trial. *European Journal of Surgical Oncology*, *52*(10), 112042. https://doi.org/10.1016/j.ejso.2026.112042

Zhu, Z., Kopcha, T. J., & Ding, A.-C. E. (2026). Comparing immersive virtual reality with traditional methods in K-12 foreign language learning: Outcomes and student perceptions. *Computers & Education: X Reality*, *8*, 100162. https://doi.org/10.1016/j.cexr.2026.100162

Copyright Holder :
© Lior Kfir et al. (2026).

First Publication Right :
© Journal of Loomingulisus ja Innovatsioon

This article is under:

