Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Level Up: Gamification and Virtual Reality in Medical Education-A BERTopic Mapping of Research Frontiers

Junhua Xian
Yangtze Normal University
Hongqun Li
lihongqun2001@126.com

Yangtze Normal University

Research Article

Keywords: Gamification, virtual reality, medical education, BERTopic modeling

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10949361/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Level Up: Gamification and Virtual Reality in Medical Education-A BERTopic Mapping of Research Frontiers

Junhua Xian[1], Hongqun Li[2*]

1. Faculty of Teacher Education, Yangtze Normal University, Chongqing, China

2. School of Modern Agriculture and Bioengineering, Yangtze Normal University, Chongqing, China

Corresponding author: Hongqun Li, Email: lihongqun2001@126.com

Abstract: Numerous reviews have mapped gamification and virtual reality (VR) research in medical education, yet none has systematically examined the semantic interrelations among themes or uncovered the structural imbalances in the knowledge base, as prior work treats topics as isolated and descriptive. To address this gap, this study applied bidirectional encoder representations from transformers (BERTopic) to 5,640 Web of Science articles, identifying 16 Topics across four categories: clinical skills, pedagogy, rehabilitation, and foundational competencies. Similarity analysis exposed a pronounced asymmetry: technology-driven skill acquisition dominates, while cognitive foundations and meaning-construction pedagogies are neglected. Based on these empirical patterns, this research propose a three-tier Integrated Educational Functions Framework comprising Cognitive Foundation, Skill Acquisition, and Meaning Construction. This framework guides future research and practice, urging balanced investment across all tiers to foster critical thinking and equitable access in technology-enhanced medical education.

Keywords: Gamification, virtual reality, medical education, BERTopic modeling

1.Introduction

Access to quality healthcare remains a fundamental human right, yet the world faces a persistent shortage

of well-trained medical professionals, a challenge magnified by aging populations and emerging diseases [1]. Traditional training models, heavily reliant on patient contact and cadaveric dissections, are increasingly constrained by ethical considerations, resource limitations, and patient safety imperatives [2]. Gamification and virtual reality (VR) have thus emerged as promising pedagogical solutions, offering scalable, risk free environments for clinical skill acquisition and democratizing access to high-quality training [3]. Unsurprisingly, scholarly output in this interdisciplinary domain has surged exponentially. However, this rapid proliferation has produced a fragmented and unwieldy knowledge base. Existing literature reviews, which predominantly rely on manual labelling or elementary counting, are labor intensive, prone to subjective bias, and incapable of detecting nuanced thematic patterns from massive textual corpora [4]. More importantly, they typically dissect topics in isolation, failing to reveal how combinations of research topics configure the field's hotspots and developmental trajectories. Furthermore, prior reviews neglect temporal dynamics and the semantic interrelationships and structural configurations of thematic bundles within the knowledge landscape, leaving stakeholders without strategic guidance [5]. To surmount these obstacles, this research introduces a dual method framework that integrates bidirectional encoder representations from transformers (BERTopic), a transformer based neural model preserving contextual embeddings. This unsupervised approach captures nascent or subtle topics without predefined categories. This study pursues three objectives: to identify core research topics, to analyze their semantic interrelations, and to derive actionable implications for future research and practice [6]. By moving beyond descriptive bibliometrics, our integrative strategy provides a holistic understanding of the field's intellectual structure, guiding educators, policymakers, and funding bodies to strategically advance medical education in service of global health.

2.Literature review

Research on gamification and VR in medical education concentrates on three primary thematic streams: procedural skill acquisition through surgical simulation, learner engagement via gamified motivational design, and therapeutic rehabilitation for aging and neurodevelopmental populations [7]. Despite promising findings within each stream, these domains have largely evolved in parallel with limited cross fertilization.

Methodologically, early systematic reviews relied on manual coding, which restricted scalability and introduced subjective bias [8]. Subsequent bibliometric studies employing co-word or co-citation analysis partially automated the synthesis process, yet they depend on predefined taxonomies and treat keywords as independent units [9], thereby losing contextual semantics embedded in abstracts. More critically, these approaches remain descriptive, visualizing clusters without explaining why certain thematic bundles attract scholarly traction.

Beyond these methodological constraints, prior reviews exhibit three substantive deficiencies. Thematic identification remains incomplete, often overlooking nascent or peripheral topics such as computational thinking and digital language tools [10]. Temporal dynamics and relational structures among topics have received scant attention, offering only static snapshots [11]. Most critically, no prior investigation has systematically mapped the semantic association structures and topic clustering patterns among thematic combinations, relying instead on isolated net effects which contradict the interdependent nature of scientific knowledge production [12].

To surmount these obstacles, we introduce a dual method framework integrating BERTopic, a transformer based neural model that preserves contextual embeddings without predefined categories [6]. This unsupervised approach captures emergent topics and maps their semantic interconnections. By moving beyond descriptive bibliometrics, our strategy explains why particular thematic bundles drive

field evolution, offering a holistic understanding that prior reviews have failed to provide. Guided by these identified voids, this study addresses the following three interrelated research questions:

RQ1: What are the core research topics in gamification and VR for medical education?

RQ2: How are these topics semantically interrelated?

RQ3: What implications does the thematic structure hold for future research and practice?

3.Methods

3.1 Data Collection and Preprocessing

This study systematically searched the Web of Science Core Collection on June 30, 2026, for peer-reviewed literature on gamification and VR in medical education, covering publications from January 2020 to June 2026. The reason for choosing this time frame is that the field evolves rapidly, and recent studies are more representative of current trends and practices. The search string combined three concept groups using Boolean operators: (gamification OR game-based learning OR serious games OR gaming) AND (virtual reality OR VR OR simulation OR immersive technology) AND (medical education OR health professions education OR clinical training OR nursing education OR surgical training OR anatomy education). Only English-language articles with available titles and abstracts were considered.

Figure 1 shows the data collection procedure. The initial search yielded 41,392 records. After limiting to the 2020-2026 period, 26,750 remained. Restriction to document type “Article” reduced the count to 17,049. Subsequent filtering by research areas (Education Educational Research, Health Care Sciences Services, Medicine Research Experimental, Nursing, and Medicine General Internal) resulted in 5,990 records. Finally, retaining only English-language publications left 5,640 articles. Two independent reviewers then screened titles and abstracts against inclusion criteria (inter-rater agreement $\kappa = 0.89$),

excluding editorials, conference abstracts, book reviews, and studies not focused on educational applications. This screening process yielded the final corpus of 5,640 articles for BERTopic analysis.

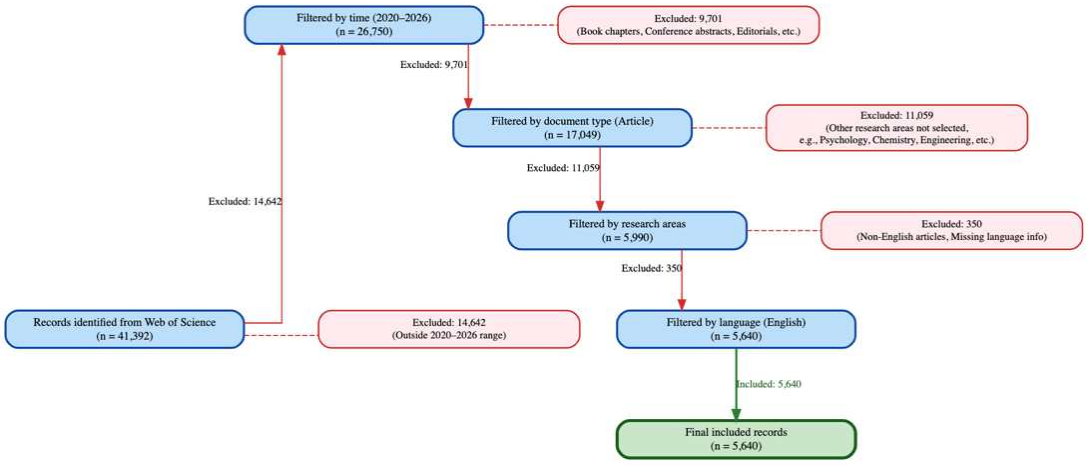


Figure 1. The Data Collection Procedure

The textual corpus comprised concatenated titles and abstracts. Preprocessing included tokenization, stop-word removal with a customized list, lowercasing, lemmatization using WordNet [13], and extraction of bigrams and trigrams such as “VR simulation” prior to stop-word removal.

3.2 BERTopic Modeling

This research employed BERTopic, a neural topic modeling technique leveraging transformer-based language models to preserve contextualized word embeddings [14]. The pipeline comprised five sequential phases.

In the embedding phase, each document was transformed into a 384-dimensional dense vector using all MiniLM L6 v2 sentence transformer [15]. This model was selected for its demonstrated performance on academic English corpora and its computational efficiency with medium sized datasets [16]. While domain specific models like BioBERT exist, our pilot testing showed MiniLM produced more coherent topic clusters for this interdisciplinary corpus without requiring fine tuning, consistent with recent benchmarking studies [17].

For dimensionality reduction, we applied Uniform Manifold Approximation and Projection with optimized parameters: n_neighbors = 15 and min_dist = 0.0. These values were selected through iterative grid search over the ranges n_neighbors = 5–50 and min_dist = 0.0–0.5, balancing cluster compactness with separation.

Clustering employed Hierarchical Density Based Spatial Clustering of Applications with Noise with min_cluster_size = 25 and min_samples = 5. These parameters were determined through systematic testing against topic coherence, with min_cluster_size adjusted between 10 and 60 to avoid overly granular or excessively broad topics. Documents not assigned to any cluster were designated as noise and excluded from subsequent topic level analyses.

Topic representation used class-based term frequency inverse document frequency, generating ranked term lists for each cluster. Topics were labelled by examining the top 10 representative terms and manually reviewing sample documents. The optimal topic number was determined through a combined quantitative and qualitative approach. Coherence scores (c_v) averaged 0.62 across all topics, ranging from 0.55 to 0.71, with topic diversity exceeding 0.90, indicating minimal redundancy. After iterative experimental runs with varying hyperparameters, we retained 16 stable and interpretable topics.

4.Findings

4.1 Topic Identification and Descriptions

The 16 topics identified through BERTopic are described below (see **Table 1** for basic information and **Figure 2** for term scores).

Topic 0 (T0) Foundational Technology in Health Professions Education

AI and data-driven technologies transform clinical teaching and medical education. AI-based systems enhance simulation training, assess competence via analytics, and foster interprofessional development

[18]. Learning analytics enable continuous evaluation and evidence-based strategies [19]. Data-driven quality assurance addresses workforce challenges and improves educational outcomes [20].

Topic 1 (T1) Gamification for Learner Engagement in Medical Education

Gamification uses points, leaderboards, badges, challenges, and role-playing to boost engagement and motivation [21]. It increases participation, learning outcomes, satisfaction, and reduces anxiety [22]. Effective design aligns with pedagogical goals and learner profiles [23].

Topic 2 (T2) Escape Room Pedagogy in Nursing and Clinical Education

Escape rooms improve nursing students' clinical decision-making, critical thinking, and problem-solving [24]. Combined with VR/AR, they enrich interactive learning and offer safe, engaging hands-on training [25].

Topic 3 (T3) Mobile Health Interventions for Patient and Provider Education

Mobile health apps support chronic disease self-management and patient education through personalized content and real-time feedback [26]. Future research should integrate mHealth into clinical training and evaluate multi-perspective effectiveness.

Topic 4 (T4) VR Clinical Simulation for Procedural Training

VR simulation provides immersive environments for procedural skills training, enhancing engagement, confidence, and spatial visualization [27,28]. It offers standardization, objective tracking, and immediate feedback, making it suitable for OSCE preparation and surgical training [29].

Topic 5 (T5) Digital Language Tools for Medical Communication Training

Digital tools like Duolingo and Quizlet aid medical terminology acquisition [30]. VR with NLP-driven dialogue and GenAI enables realistic patient interactions and personalized communication training [31,32].

Topic 6 (T6) Exergaming and Cognitive Rehabilitation in Aging Populations

Exergaming combines physical exercise with cognitive tasks, improving cognitive function in older adults with dementia and mild cognitive impairment [33]. It promotes neuroplasticity and cognitive abilities [34].

Topic 7 (T7) VR Simulation in Surgical Skills Training

VR offers controlled, repeatable environments for surgical skill practice, significantly outperforming traditional methods [35], especially in laparoscopic and endoscopic training with objective assessment.

Topic 8 (T8) VR for Motor Rehabilitation and Pain Management

Immersive VR improves strength, flexibility, coordination, and neuroplasticity, while providing analgesic effects [36]. It is more enjoyable than conventional therapy and accelerates motor recovery in neurological and musculoskeletal conditions.

Topic 9 (T9) Online Platforms for Health Education and Infectious Disease Training

Platforms like Open WHO provide scalable, free courses covering outbreak response and infection prevention, with over 315 courses in 75 languages and 9.2 million registrations [37]. MOOCs are essential during public health crises.

Topic 10 (T10) VR for Neurodevelopmental Disorder Interventions

VR improves attention in ADHD and enhances social skills and emotional regulation in ASD through personalized scenarios [38]. It serves as both assessment and training tool for cognitive and social impairments.

Topic 11 (T11) iVR Environments for Medical Learning

Head-mounted display-based iVR creates immersive learning environments, effective in anatomy, nursing, and skills training, with better knowledge retention than traditional methods [39].

Topic 12 (T12) Generative AI and Robotics in Medical Pedagogy

Generative AI and robotics are applied in surgical video analysis, robotic assistance, clinical reasoning, virtual patient interactions, and personalized learning [40]. They increase engagement and simulation realism.

Topic 13 (T13) Computational Thinking in Biomedical and Health Informatics Education

Computational thinking covers data handling, programming, analysis, interpretation, communication, and ethical awareness [41]. It should be cultivated through interdisciplinary, inquiry-based learning for data-intensive clinical environments [42].

Topic 14 (T14) Mathematics and Quantitative Skills in Undergraduate Medical Curricula

Mathematics supports biophysics, biochemistry, and informatics. Teaching hours for advanced mathematics are generally reasonable [43]. Recommendations include biomathematics in year one and modelling/statistics in year two [44].

Topic 15 (T15) AR for Anatomy and Clinical Skills Training

AR improves knowledge retention up to 75%, optimizes anatomy understanding, and facilitates safe clinical procedure teaching [45]. It provides cost-effective, accessible teaching methods [46,47].

Table 1. 16 Topics Basic Information and Characteristics

Topics	Documents Distribution	Topic Theme	Terms
0	1268	Foundational Technology in Health Professions Education	student, education, gamification, training, teach, technology, objective, virtual, digital, online
1	411	Gamification for Learner	education, student, digital, teacher, teach,

		Engagement in Medical Education	technology, knowledge, environment, school, development
2	379	Escape Room Pedagogy in Nursing and Clinical Education	gamification, student, gamified, education, motivation, elements, engagement, gamification education, effects, teacher
3	346	Mobile Health Interventions for Patient and Provider Education	nursing, escape, student, education, medical, nursing student, clinical, nurse, care, professional
4	258	VR Clinical Simulation for Procedural Training	health, interventions, physical, children, mental, mobile, apps, objective, m-health, mental health
5	248	Digital Language Tools for Medical Communication Training	virtual, reality, VR, training, medical, simulation, immersive, clinical, student, education
6	131	Exergaming and Cognitive Rehabilitation in Aging Populations	language, English, vocabulary, reading, learners, digital, student, English language, skills, experimental
7	112	VR Simulation in Surgical Skills Training	cognitive, adults, exergame, dementia, physical, aging, exercise, objective, health, training
8	100	VR for Motor Rehabilitation	surgical, simulation, training, medical,

		and Pain Management	surgery, reality, virtual, skills, VR, laparoscopic
9	98	Online Platforms for Health Education and Infectious Disease Training	rehabilitation, pain, stroke, patients, exercise, training, objective, motor, virtual, VR
10	89	VR for Neurodevelopmental Disorder Interventions	health, medical, education, infection, training, care, online, objective, vaccine, student
11	87	iVR Environments for Medical Learning	children, attention-deficit hyperactivity disorder (ADHD), disorder, autism, spectrum, hyperactivity, deficit, spectrum disorder, autism spectrum, disorder ADHD
12	68	Generative AI and Robotics in Medical Pedagogy	virtual, reality, VR, immersive, environment, education, technology, student, experience, laboratory
13	59	Computational Thinking in Biomedical and Health Informatics Education	artificial, education, genai, generative, robotics, robot, robots, generative artificial, student, technology
14	51	Mathematics and Quantitative Skills in Undergraduate Medical Curricula	computational, computational thinking, thinking, programming, student, computer, skills, computer science, teach,

			science
15	50	AR for Anatomy and Clinical Skills Training	education, student, mathematics, learners, teach, undergraduate, motivational, teacher, knowledge, designed

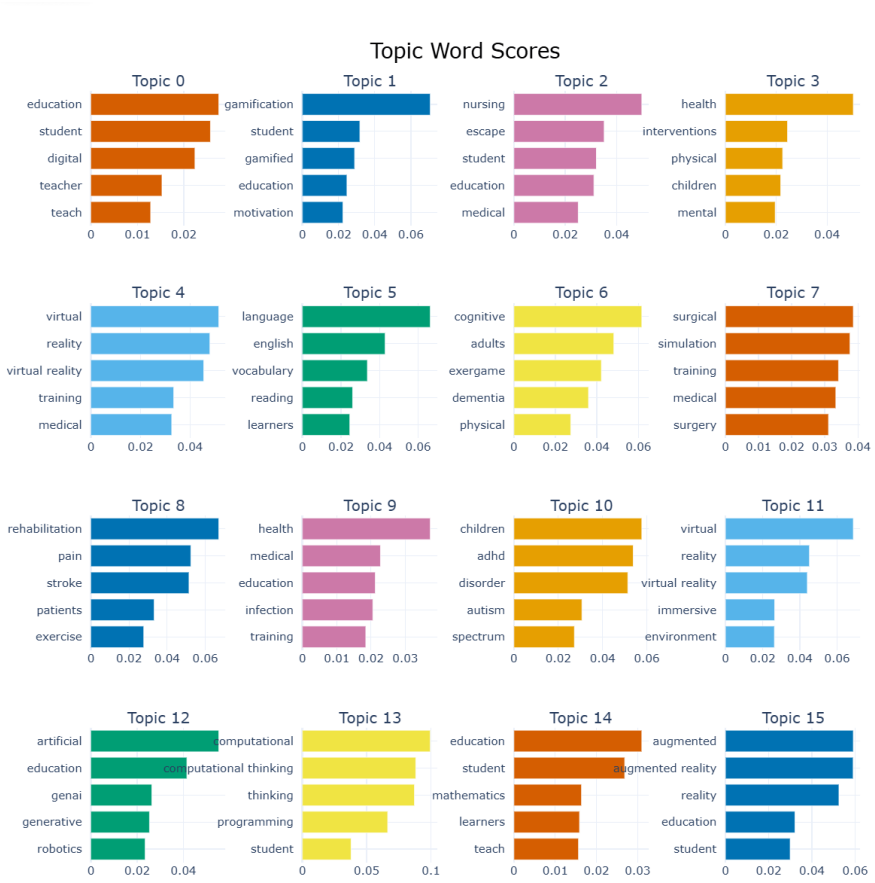


Figure 2. The Scores of Topic Words

4.2 Semantic Interrelations Among Topics

The similarity matrix (**Figure 3**) reveals a clear hierarchical structure among the 16 topics, with several highly cohesive clusters and a few isolated outliers. The strongest association is observed between Topic 4 and Topic 11 (0.91), indicating substantial conceptual overlap, likely driven by shared methodological or contextual features. Similarly, Topics 13, 14, and 15 form a tight triad, with pairwise correlations ranging from 0.81 to 0.85, suggesting that these three topics represent a coherent subtopic domain. Topic

0 also shows consistently high similarity with Topic 5 (0.86), Topic 14 (0.85), and Topic 15 (0.82), positioning it as a central node that bridges multiple clusters.

At the opposite extreme, Topic 1 exhibits the lowest overall connectivity, with its weakest link to Topic 4 (0.46) and generally modest correlations with most others, rarely exceeding 0.71. This pattern implies that Topic 1 addresses a relatively distinct or peripheral research angle. A more subtle but noteworthy finding is the asymmetric relationship between Topic 2 and Topic 7 (0.81) versus Topic 2 and Topic 4 (0.75), while Topic 4 shows a much stronger tie to Topic 11 (0.91) than to Topic 2, underscoring that high similarity with one topic does not necessarily generalize to its close neighbors.

Another unexpected insight is the moderate similarity between Topic 8 and Topic 6 (0.79) and between Topic 8 and Topic 3 (0.74), despite Topic 8 having low correlations with Topics 12 and 13 (0.48–0.50). This divergence suggests that Topic 8 shares specific subthemes with certain topics but remains conceptually separated from others in the same broad domain. The matrix also highlights that Topic 3 and Topic 6 are highly related (0.82), while Topic 9 serves as a moderate connector, with values around 0.60–0.75 across most pairs, yet never reaching the extreme highs seen in the core clusters. Collectively, these patterns indicate that the topic space is not uniformly dense, but rather organized into a few densely interconnected cores, several peripheral singletons, and a small number of bridging topics that facilitate cross-cluster connections.

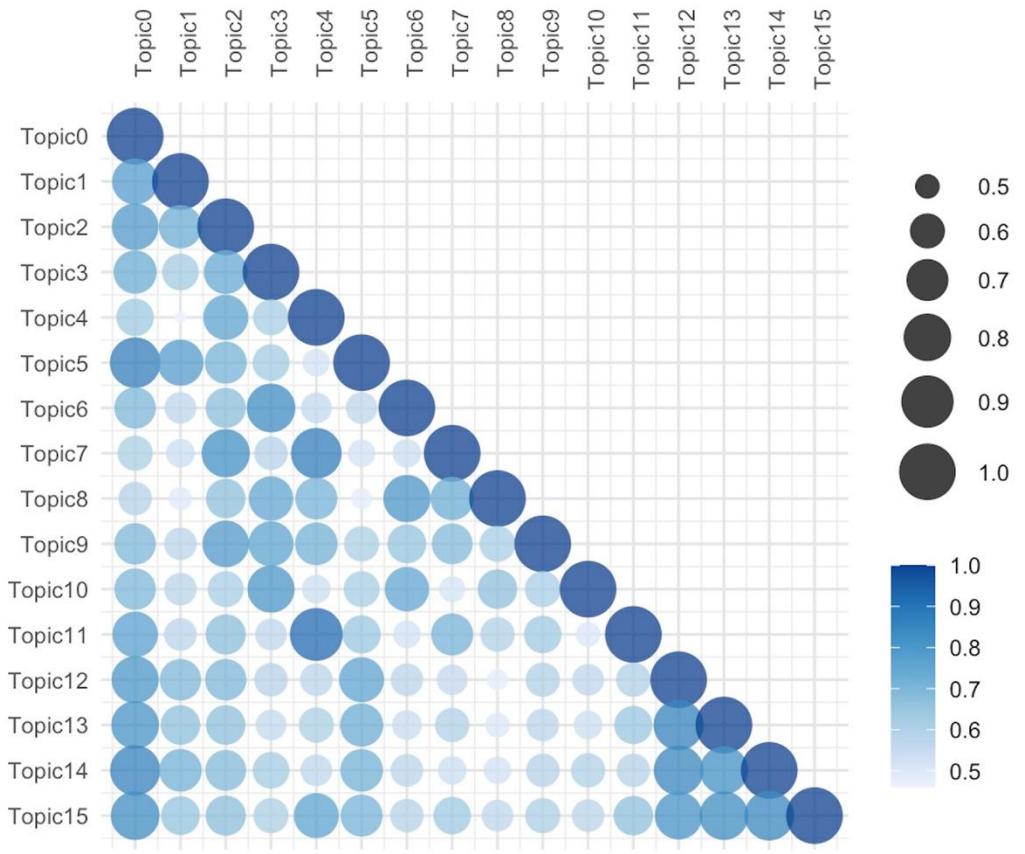


Figure 3. A Clear Hierarchical Structure Among the 16 Topics

4.3 Aggregating Topics into Four Pedagogical Categories

The 16 topics are aggregated into four overarching categories (**Figure 4**) based on their educational functions and the relational structures in the semantic similarity matrix. This classification exposes a pronounced structural asymmetry: technological applications attract disproportionate attention relative to foundational cognitive prerequisites—an imbalance that will be addressed in the Integrated Educational Functions Framework (Section 5.3).

Category 1: Clinical Skills and Health Technology Applications (Topics 3, 4, 7, 9, 15) – focuses on direct clinical skill acquisition via mobile health interventions, VR/AR simulation for procedural and surgical training, and online health education platforms, all sharing the goal of translating technological innovations into measurable clinical competencies and patient outcomes.

Category 2: Pedagogy and Learner Engagement (Topics 1, 2, 12) – covers gamification strategies, escape

room pedagogies, and generative AI/robotics integration, united by their emphasis on enhancing learner participation, interaction, and cognitive engagement through innovative instructional designs.

Category 3: Rehabilitation and Special Population Interventions (Topics 6, 8, 10, 11) – groups VR/exergaming for cognitive ageing, motor rehabilitation and pain management, neurodevelopmental disorder interventions, and immersive VR environments for medical learning, all using immersive or gamified approaches for restorative, therapeutic, or supportive purposes in vulnerable or clinical populations.

Category 4: Foundational Competencies and General Technologies (Topics 0, 5, 13, 14) – addresses foundational technology in health professions education, digital language tools for medical communication, computational thinking in biomedical informatics, and mathematics in undergraduate curricula, collectively cultivating the essential cognitive, analytical, and technical competencies that underpin more specialized clinical and pedagogical applications.

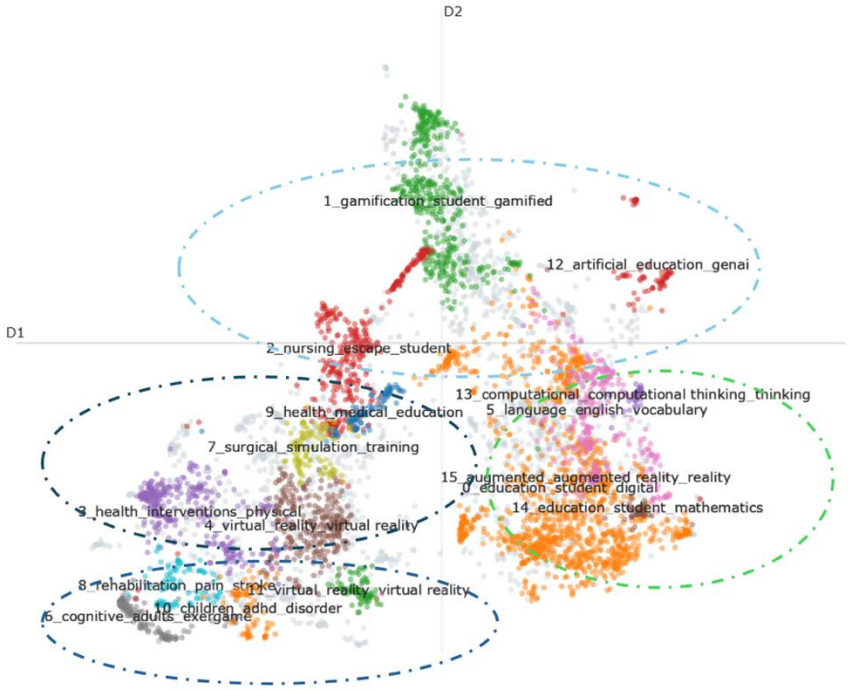


Figure 4. Four Category of These Topics

5.Discussion

5.1 For RQ1: What are the core research topics in gamification and VR for medical education?

Prior reviews have tended to portray the development of gamification and VR in medical education as a linear narrative of technological progress driving educational transformation. However, the thematic distribution in this study reveals a considerably more complex picture, with the most salient feature being the pronounced disparity in publication volume between technology application topics (e.g., T3, T4, T7, T9, T15) and educational foundation topics (T0, T5, T13, T14).

A noteworthy observation is that topics concerning mathematics, quantitative skills, and computational thinking (T13 and T14) account for a minimal proportion of the total literature, whereas foundational technology (T0), encompassing AI and learning analytics, alone contributes a substantial share of all publications. This distribution pattern reveals a fundamental tension: substantial research efforts have been directed toward developing AI diagnostic tools [18], VR simulators, and learning analytics systems, while considerably less attention has been paid to whether medical students possess the cognitive prerequisites for understanding the underlying logic of these tools [20]. Viewed through the lens of Bloom's taxonomy, foundational quantitative reasoning and computational thinking belong to the higher-order cognitive categories of analysis and evaluation, whereas technical operations training resides primarily at the application level [48]. The pronounced skew in current research priorities suggests a potential inversion of the cognitive gradient, wherein application-oriented training has been emphasized at the expense of foundational comprehension. When learners become proficient in operating VR simulators yet lack the capacity to interpret the algorithmic logic underlying the feedback mechanisms, the educational outcome may be characterized as operational compliance with a technological black box rather than genuine enhancement of clinical judgment [27]. This concern does

not constitute a rejection of technological applications; rather, it calls for a deeper conceptualization of technological literacy in medical education, one that encompasses critical understanding of technological principles rather than being limited to operational proficiency.

Another noteworthy anomalous pattern concerns T15. Although systematic reviews have demonstrated that AR can substantially improve knowledge retention and provide a viable alternative in the context of increasingly scarce cadaveric resources, the publication volume for this theme ranks at the bottom of all topics [45]. This paradox of high evidentiary support coupled with low research investment cannot be attributed solely to technological immaturity, as the technical feasibility of AR in medical education has been well established [47]. A more plausible explanation involves the incentive structures of the academic publishing market, in which researchers and journal editors tend to priorities topics carrying the label of novelty, such as iVR (T11) and generative AI (T12), while AR, having been commercially available for a longer period, struggles to attract the premium of novelty in scholarly dissemination [36, 46]. This bias warrants attention not only because it distorts the efficient allocation of research resources, but also because it has tangible implications for equity in global medical education. When the majority of resource-constrained medical schools cannot afford head-mounted VR equipment [27], the low-cost and highly accessible technological pathway represented by AR should constitute the most realistic option for advancing digital transformation [47], yet its educational potential remains underrealized owing to insufficient scholarly attention. From the perspective of global health, this research gap constitutes a knowledge production imbalance that merits active correction.

Furthermore, the rehabilitation and special population intervention topics (T6, T8, T10, and T11) exhibit overall publication volumes that stand in stark contrast to the escalating demands posed by global population aging and the rising prevalence of neurodevelopmental disorders. The critical issue here is

not a deficiency of evidence regarding therapeutic efficacy, as the clinical effects of VR and exergaming in cognitive rehabilitation have received ample empirical support [25]. Rather, it lies in the fact that medical education systems have not yet integrated digital rehabilitation technologies into their core curricula [20]. In the majority of medical schools, gamified cognitive training and VR-assisted rehabilitation remain elective or even absent modules. This disconnect between educational provision, and clinical need reflects a systemic inertia in curriculum renewal mechanisms, which lag behind the pace of technological iteration [19]. Researchers should extend their inquiry beyond whether VR can be used for rehabilitation to address how future clinicians can be trained to master and effectively integrate these emerging therapies into clinical pathways. The latter question, which is central to medical education's responsibility to meet societal health needs, remains conspicuously underrepresented in the existing literature.

5.2 For RQ2: How are these topics semantically interrelated?

The similarity matrix reveals deep structural features of the field's knowledge architecture. Its significance extends beyond the statistical description of inter-topic associations to expose epistemological challenges that medical education, as an interdisciplinary field, faces in knowledge integration.

One structural feature warranting close examination is the overall low connectivity of T1. The similarity scores between T1 and the majority of other topics remain consistently low, and its similarity with T4 represents one of the lowest values across all topic pairs. This indicates that, within the semantic space, motivating learners and training clinical skills through simulation are positioned almost orthogonally to each other. The educational implications of this isolation merit careful interpretation. Gamification research tends to concentrate on the internal mechanisms of motivational psychology [23], including

self-determination theory, flow experience, and achievement goal orientation [27], while paying relatively less attention to the specific learning outcomes that these motivational states ultimately serve [28]. Concurrently, VR simulation research emphasizes fidelity, skill transfer effects, and objective performance assessment, yet rarely incorporates learning motivation and instructional design theory into intervention frameworks in a systematic manner. When what is learned and why one is willing to learn remain the territories of two scholarly communities that lack effective communication, technologically driven instructional innovations in medical education will struggle to achieve the deep integration that learning science demands. The most educationally effective interventions would ideally integrate high-fidelity simulation with carefully structured motivational design, yet the current knowledge structure pushes these two components toward opposite directions.

However, the low connectivity of T1 may also be partially attributable to terminological differentiation in academic publishing within this field. A substantial body of research adopts gamification design elements in practice, including immediate feedback, progress tracking, and graded task challenges, but frames these in keywords and abstracts using terms such as simulation training, interactive learning, or technology-enhanced instruction [49], which carry greater perceived academic legitimacy in the medical education context. This strategic terminological choice may stem from the cultural resistance that the concept of gamification faces in the tradition of medical education, where associating clinical training with games might be perceived by some educators and reviewers as diminishing the seriousness of the research. If this interpretation holds, then the isolation of T1 is not merely a statistical artifact of semantic clustering, but also a reflection of a conceptual avoidance tendency within the field's scholarly culture that warrants critical reflection. This observation suggests that advancing the deep integration of gamification and VR research requires not only methodological bridge building but also progress in

overcoming a priori biases against the concept of gaming in academic discourse.

Another structural finding deserving attention is the hub position of T0. T0 exhibits high similarity with T5, T14, and T15, indicating that the theme of AI and learning analytics serves as a bridging node for cross-category knowledge integration. Yet the positive implications of this finding require cautious assessment. The high centrality of T0 also implies that the field's knowledge network lacks sufficient robustness, as the flow of knowledge relies heavily on a single hub node. Should the research agenda of T0 shift, for instance from educational applications toward clinical decision support, multiple peripheral topics could face the risk of losing their knowledge grounding. A mature academic field should possess a polycentric knowledge circulation network; the current monocentric star topology suggests that knowledge integration in this field remains at an early stage of development.

The most theoretically illuminating structural feature is the high-correlation triad formed by T13, T14, and T15. The theoretical significance of this cluster transcends the methodological level of semantic similarity. It suggests that the field is spontaneously forming a cognitive competence cluster centered on spatial-quantitative reasoning, where logical deduction in mathematical modelling, abstract layering and algorithmic expression in computational thinking, and three-dimensional spatial localization and mental rotation in AR anatomy all share a common cognitive foundation [50]. From the perspective of educational psychology, the emergence of this cluster is not coincidental; it corroborates theoretical predictions from cognitive science regarding the deep connection between spatial ability and quantitative reasoning [51]. The critical implication of this finding, however, lies in the observation that all three topics rank at the bottom of the publication distribution. This indicates that medical education researchers have not yet fully recognized the importance of cultivating this transdisciplinary cognitive competence. In current curricula, anatomy is taught by anatomists, mathematics by mathematicians, and programming

by computer scientists, with little meaningful cognitive integration across the three courses [47]. The structural relationships revealed in this study suggest that these domains share intrinsic cognitive affinities, and medical education should leverage these affinities to design cross-curricular learning experiences, rather than allowing disciplinary boundaries to further fragment an already scarce cognitive training resource [47]. This finding provides a specific cognitive rationale for integrated curriculum reform, moving beyond advocacy that remains solely at the level of educational philosophy.

5.3 For RQ3: What implications does the thematic structure hold for future research and practice?

Building upon the empirical findings of the four higher-order categories, this study proposes an Integrated Educational Functions Framework. It is important to note that this framework is advanced as a theoretically grounded integrative proposal, which synthesizes the empirical clustering results with established educational psychology theories (e.g., hierarchical models of cognitive processing and instructional design), rather than as a pattern directly or mechanically revealed by the similarity data. This framework reorganizes the four descriptive categories into a three-tier structure with intrinsic hierarchical relationships and functional complementarity. The construction of this framework is motivated by a core theoretical judgment: while the four-category classification has descriptive utility, it does not reveal the hierarchical differences and complementary relationships among the categories in terms of educational function. The following reorientation of the 16 topics, therefore, represents a theory-driven regrouping based on these two pedagogical dimensions, aimed at offering a coherent response to the observed imbalances.

The first dimension is the cognitive processing hierarchy, ranging from basic symbolic representation to clinical action implementation. This dimension draws upon hierarchical theories from cognitive psychology to distinguish a continuum from abstract symbol manipulation to concrete action execution.

At one end lie basic cognitive abilities, including formal operational reasoning in mathematics (T14), algorithmic construction in computational thinking (T13), and symbolic processing in language comprehension (T5). At the other end lie clinical action abilities, including fine motor control in surgical procedures (T7), procedural execution in rehabilitation interventions (T8), and decision implementation in clinical reasoning (T4). Foundational competencies (Category 4) and clinical skills (Category 1) are situated at the respective ends of this dimension. The empirical findings reveal a critical structural deficiency along this dimension: the thematic connections between the two ends are extremely weak, indicating that cognitive preparation and action application lack effective articulation in research practice. From a pedagogical standpoint, this discontinuity implies that learners transitioning from foundational cognitive training to clinical action training do so without structurally designed cognitive scaffolding to support the progression.

The second dimension is the instructional function orientation, which distinguishes between technology provision and pedagogical strategy design. This dimension differentiates two fundamentally distinct research orientations. At one end are technology-driven topics, including VR simulation systems (T4, T7), AI assessment tools (T0, T12), and AR visualization platforms (T15), where technology development and usability testing constitute the core research paradigm. At the other end are strategy-driven topics, including gamification design (T1), escape room pedagogy (T2), and online platform operations (T9), where instructional design theory and learning experience optimization are the primary concerns. Clinical skills topics (Category 1) are predominantly oriented toward technology tools, while pedagogy topics (Category 2) are oriented toward strategy design. An unexpected finding of this study is that the relationship between these two orientations is not a simple dichotomy. T0 (foundational technology) spans both ends, indicating that general technology can serve as both a tool and a strategic

vehicle, whereas T1 (gamification) is almost entirely fixed at the strategy end and rarely connects effectively with technology topics. This suggests that, in the knowledge production of this field, which technology to adopt and which pedagogy to design are still treated as relatively separate research questions, rather than as integrated components of a unified design [52]. From the perspective of instructional intervention design, however, the integration of technological tools and pedagogical strategies should not be an add-on step after research is completed but should be incorporated into a unified theoretical framework from the initial stage of research design.

Based on these two dimensions, this study proposes a three-tier Integrated Educational Functions Framework (**Figure 5**).

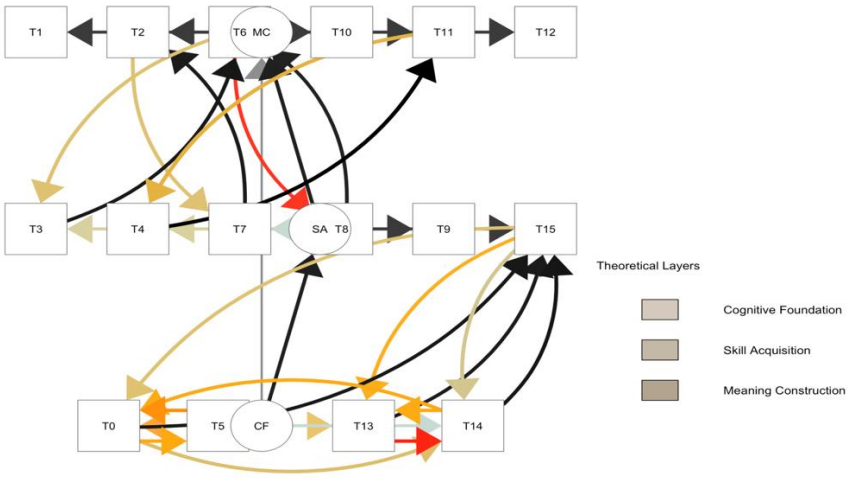


Figure 5. A Three-tier Integrated Educational Functions Framework

The first tier is Cognitive Foundation, corresponding to Category 4 (T0, T5, T13, T14). Its educational function is to provide medical students with the foundational cognitive tools necessary for effective interaction with technological environments, including quantitative reasoning ability, computational thinking, data literacy, and academic language proficiency. This tier constitutes the cognitive prerequisite for all technology-enhanced learning. Without adequate development of this tier, subsequent technical

training faces the cognitive risk of operational competence without conceptual understanding [53]. Currently, this tier receives severely insufficient scholarly attention in the field, constituting the weakest foundational layer of the framework.

The second tier is Skill Acquisition, corresponding to the procedural training topics within Category 1 and Category 3 (T3, T4, T7, T8, T9, T15). Its educational function is to translate cognitive abilities into operable clinical skills, facilitating the transformation from declarative knowledge to procedural knowledge in simulated environments. This tier represents the research center of the field, with the highest concentration of publication volume, yet it urgently needs to establish systematic connections with the first and third tiers [54].

The third tier is Meaning Construction, corresponding to Category 2 and the affectively and cognitively integrative topics within Category 3 (T1, T2, T6, T10, T11, T12). It should be clarified that T8, though also in Category 3, is assigned to Skill Acquisition rather than Meaning Construction. This is because T8 primarily focuses on procedural motor skill training and physical conditioning, whereas T6, T10, and T11 are placed in Meaning Construction for their emphasis on intrinsic motivation, social scenario provision, emotional regulation, and cognitive restructuring. These latter functions involve deeper personal and contextual meaning-making beyond procedural automation. Its educational function is not to directly transmit measurable skills, but to provide the affective foundation and contextual support for the effective functioning of the first two tiers by activating intrinsic motivation, offering embodied situational experiences, and constructing personal and social meaning in learning. This tier is currently the most dispersed in research focus and exhibits the weakest connections with the other two tiers. Yet from the perspective of longitudinal evidence in educational psychology, motivation and meaning construction are precisely the critical variables that determine whether technology-enhanced learning

produces lasting effects.

The core theoretical contribution of this framework lies in revealing the vertical integrative relationships that should exist among the three tiers, as well as the significant deviation of the current publication distribution from this ideal structure. Specifically, current research investment is heavily concentrated in the second tier (Skill Acquisition), while the first tier (Cognitive Foundation) is severely underdeveloped, and the third tier (Meaning Construction) operates in isolation. If technological applications in medical education remain confined to practicing skills with increasingly sophisticated tools, while neglecting adequate preparation of the learner's cognitive foundation and effective support for meaning construction, their educational effects will be limited to the shallow goal of procedural automation. Such an approach cannot cultivate clinicians who possess critical thinking, adaptive expertise, and intrinsic professional commitment. A genuinely learner-centered approach to research on technology in medical education should strive to establish pathways connecting the three tiers, and this is the fundamental directional guidance that the proposed framework offers for future curriculum design, research agendas, and funding strategies.

6.Conclusion

This BERTopic mapping of 16 research topics reveals that the field of gamification and VR in medical education is structurally imbalanced. Technology driven clinical skill acquisition dominates the literature, while foundational cognitive competencies and meaning construction pedagogies receive scant attention. The proposed three tier Integrated Educational Functions Framework, comprising Cognitive Foundation, Skill Acquisition, and Meaning Construction, highlights this disparity. Current research overemphasizes the middle tier, neglecting the cognitive prerequisites and motivational supports essential for deep learning. This imbalance limits the transformative potential of these technologies, calling for a more

holistic approach that integrates all three tiers to foster critical thinking and adaptive expertise in future clinicians.

Several limitations should be acknowledged. First, the dataset was restricted to Web of Science Core Collection, potentially excluding relevant studies from other databases or non-English publications, which may introduce selection bias. Second, BERTopic modeling involves subjective parameter choices, such as embedding model selection and clustering thresholds, which affect the resulting topic structure. Although topics were manually validated, labeling and categorization remain reliant on researcher interpretation. Third, this study provides a static snapshot up to 2025 and does not incorporate citation network or longitudinal dynamics, limiting insights into knowledge evolution and influence. Finally, treating topics as aggregate constructs may obscure intra topic heterogeneity and subtle distinctions within clusters, reducing analytical granularity.

The findings offer several directions for future research and practice. Researchers should prioritize investigations into cognitive foundational skills, including computational thinking and quantitative reasoning, to equip learners for advanced technology use. Bridging the current divide between technology oriented and pedagogy-oriented themes requires designing integrated interventions that combine skill training, cognitive scaffolding, and motivational design. Educators and curriculum developers are encouraged to build vertically integrated curricula that progressively link mathematical reasoning, computational thinking, and clinical simulation. Policymakers and funding bodies should reallocate resources to underfunded yet impactful areas, particularly accessible technologies like AR for low resource settings and the pedagogical integration of rehabilitation technologies. Future studies should employ longitudinal and comparative designs to examine how topic combinations influence research impact and to track thematic evolution, thereby informing adaptive and equitable strategies for

global health education.

Abbreviations

VR= virtual reality

AR= augmented reality

BERT= bidirectional encoder representations from transformers

ASD= autism spectrum disorder

AI= artificial intelligence

ADHD= attention-deficit hyperactivity disorder

iVR = immersive virtual reality

Acknowledgements

This achievement was a collective effort. The authors also acknowledge Clarivate Analytics for providing access to the Web of Science database, which was essential for this research.

Author contribution

Junhua Xian: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – Original Draft Preparation.

Hongqun Li: Conceptualization, Methodology, Project Administration, Resources, Validation, Writing – Review & Editing.

All authors reviewed and approved the final manuscript.

Funding

Not applicable.

Data availability

The data that support the findings of this study were derived from the Web of Science database, available

under license from Clarivate Analytics. Due to licensing restrictions, the raw dataset cannot be publicly redistributed. The data can be accessed and recreated by executing the reported search strategy within the Web of Science platform.

Declaration

Ethics approval and consent to participate

As this manuscript is a review of previously published scientific literature, it did not involve any new studies with human, or animal participants conducted by the authors. Therefore, declarations regarding ethics approval and consent to participate are not applicable to this article.

Consent for publication

Not applicable.

Competing interests

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

[1] Iyer, A. A., & Umadevi, K. S. (2025). Design and analysis of TwinCardio framework to detect and monitor cardiovascular diseases using digital twin and deep neural network. *Scientific Reports*, *15*(1), 24376.

[2] Xian, J., Wu, J. G., & Lee, S. M. (2025). Modeling the evolution of virtual reality in nursing education: a BERTopic-based analysis of research trends and future directions. *BMC nursing*, *24*(1), 1332.

[3] Xian, J., & Wu, J. G. (2026). Past, present, and future of virtual reality-based biological science education (2010–2025): a BERTopic-driven analysis. *BMC Medical Education*.

[4] Zhao, P., Zhang, X., Cao, J., Cheng, M. M., Yang, J., & Li, X. (2026). A literature review of literature reviews in pattern analysis and machine intelligence. *Science China Information Sciences*, *69*(5),

151101.

[5] Campbell, J. T., & Fiss, P. C. (2026). Tackling the complexity challenge: When and how to engage in configurational and hybrid theorizing. *Academy of Management Review*, (ja), amr-2024.

[6] Egger, R., & Yu, J. (2022). A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts. *Frontiers in sociology*, *7*, 886498.

[7] Rafi, A., Habibi, A., & Kalantarion, M. (2025). Traditional medical education under siege: the revolutionary integration of virtual reality and gamification in medical escape rooms. *BMC Medical Education*, *25*(1), 898.

[8] Li, L., Mathrani, A., & Susnjak, T. (2026). Transforming evidence synthesis: A systematic review of the evolution of automated meta-analysis in the age of AI. *Research Synthesis Methods*, 1-48.

[9] Carballo-Meilan, A., McDonald, L., Pragot, W., Starnawski, L. M., Saleemi, A. N., & Afzal, W. (2022). Development of a data-driven scientific methodology: From articles to chemometric data products. *Chemometrics and Intelligent Laboratory Systems*, *225*, 104555.

[10] Alasmari, T., & Al-Zahrani, A. (2026). From Crisis to Innovation: Forecasting Educational Technology Research Through Bibliometric Analysis of Pandemic-Era Publications (2020–2024). *Journal of Educational Computing Research*, *64*(3), 756-786.

[11] Hewett, R., & Meier-Barthold, M. (2026). Embracing complexity in HRM research: A call for system and process perspectives. *Human Resource Management*, *65*(4), 995-1013.

[12] Wu, Q., Wu, Z., Zhang, W., & Wang, W. (2026). Is Interdisciplinary Research More Conducive to Knowledge Innovation? Evidence From Swiss National Science Foundation Projects. *European Journal of Education*, *61*(2), e70530.

[13] Chai, C. P. (2023). Comparison of text preprocessing methods. *Natural language engineering*, *29*(3),

509-553.

[14] Yang, C., & Kim, Y. (2025). Enhancing topic coherence and diversity in document embeddings using LLMs: A focus on BERTopic. *Expert Systems with Applications*, *281*, 127517.

[15] Zafar, A., Wasim, M., Zulfiqar, S., Waheed, T., & Siddique, A. (2024). Transformer-based topic modeling for Urdu translations of the Holy Quran. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *23*(10), 1-21.

[16] Khodeir, N., & Elghannam, F. (2025). Efficient topic identification for urgent MOOC Forum posts using BERTopic and traditional topic modeling techniques. *Education and Information Technologies*, *30*(5), 5501-5527.

[17] Mersha, M. A., & Kalita, J. (2024). Semantic-driven topic modeling using transformer-based embeddings and clustering algorithms. *Procedia Computer Science*, *244*, 121-132.

[18] Khakpaki, A. (2025). Advancements in artificial intelligence transforming medical education: a comprehensive overview. *Medical education online*, *30*(1), 2542807.

[19] Izquierdo-Condoy, J. S., Arias-Intriago, M., Corrales, L. M., & Ortiz-Prado, E. (2026). Artificial intelligence in medical education: transformative potential, current applications, and future implications. *JMIR Medical Education*, *12*(1), e77127.

[20] Sapci, A. H., & Sapci, H. A. (2020). Artificial intelligence education and tools for medical and health informatics students: systematic review. *JMIR medical education*, *6*(1), e19285.

[21] Mouzarou, A., & Raffay, V. (2026). Systematic Review of Gamification in Medical Teaching and Learning. *Medical-Clinical-Research*, *2*(3), 53-58.

[22] Wang, Y. F., Hsu, Y. F., Fang, K. T., & Kuo, L. T. (2024). Gamification in medical education: identifying and prioritizing key elements through Delphi method. *Medical Education Online*, *29*(1),

2302231.

[23] Khazaei, M. R., Hosseini, M. A., & Ramezani, G. (2025). Gamification in medical sciences education: a scoping review. *Recenti Progressi in Medicina*, *116*(5), 291.

[24] Köse, A., & Özcan, S. (2025). Effectiveness of escape rooms in nursing education: A systematic review. *Nurse Education Today*, *151*, 106745.

[25] Xian, J., Wu, J. G., & Li, H. (2026). Effects of virtual reality-based instruction on secondary school learners' computational thinking, biological knowledge acquisition, and motivation during oocyte meiosis learning. *Computers & Education: X Reality*, *9*, 100167.

[26] Wang, Y., Min, J., Khuri, J., Xue, H., Xie, B., Kaminsky, L. A., & Cheskin, L. J. (2020). Effectiveness of mobile health interventions on diabetes and obesity treatment and management: systematic review of systematic reviews. *JMIR mHealth and uHealth*, *8*(4), e15400.

[27] Xian, J., Wu, J. G., & Wei, W. (2025, December). From static to fantastic: exploring the impact of virtual reality in biological science instruction-a scoping review. In *2025 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)* (pp. 1-8). IEEE.

[28] Chance, E. A. (2025). The combined impact of AI and VR on interdisciplinary learning and patient safety in healthcare education: a narrative review. *BMC medical education*, *25*(1), 1039.

[29] Savir, S., Khan, A. A., Yunus, R. A., Rehman, T. A., Saeed, S., Sohail, M., ... & Matyal, R. (2023). Virtual reality: the future of invasive procedure training?. *Journal of cardiothoracic and vascular anesthesia*, *37*(10), 2090-2097.

[30] Panfilova, V., Spichak, V., & Zhumakhanova, A. (2022). Educational mobile games as a tool for increasing vocabulary when learning a foreign language. *International Journal of Web-Based Learning and Teaching Technologies (IJWLTT)*, *17*(1), 1-27.

[31] Cook, D. A., Overgaard, J., Pankratz, V. S., Del Fiol, G., & Aakre, C. A. (2025). Virtual patients using large language models: scalable, contextualized simulation of clinician-patient dialogue with feedback. *Journal of Medical Internet Research*, *27*, e68486.

[32] Xiong, Z., Wu, J. G., Wu, S., & Xian, J. (2026). Enhancing HOTS through role-playing GenAI: a case study in classical literature education. *Innovation in Language Learning and Teaching*, 1-18.

[33] Liao, Y. Y., Chen, I. H., Hsu, W. C., Tseng, H. Y., & Wang, R. Y. (2021). Effect of exergaming versus combined exercise on cognitive function and brain activation in frail older adults: A randomised controlled trial. *Annals of Physical and Rehabilitation Medicine*, *64*(5), 101492.

[34] Zhao, Y., Feng, H., Wu, X., Du, Y., Yang, X., Hu, M., ... & Zhao, Y. (2020). Effectiveness of exergaming in improving cognitive and physical function in people with mild cognitive impairment or dementia: systematic review. *JMIR serious games*, *8*(2), e16841.

[35] Kaplan, A. D., Cruit, J., Endsley, M., Beers, S. M., Sawyer, B. D., & Hancock, P. A. (2021). The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: A meta-analysis. *Human factors*, *63*(4), 706-726.

[36] Freeman, D., Haselton, P., Freeman, J., Spanlang, B., Kishore, S., Albery, E., ... & Nickless, A. (2018). Automated psychological therapy using immersive virtual reality for treatment of fear of heights: a single-blind, parallel-group, randomised controlled trial. *The Lancet Psychiatry*, *5*(8), 625-632.

[37] Utunen, H., Piroux, C., Balaciano, G., Brennan, E., Attias, M., & Mattar, L. (2026). Provision of Massive Open Online Courses in Just-in-Time Modalities: Experience Report of the OpenWHO. org Platform's 7 Years of Learning Response to Health Emergencies. *Learning Health Systems*, *10*(1), e70058.

[38] Xian, J., Wu, J. G., & Zou, D. (2026). Effects of enhancing education for students with intellectual disability through game-based learning: A meta-analysis. *Educational Technology & Society*, *29*(4), 159-177.

[39] Zhang, W., Ding, Z., Bakaev, M., Razumnikova, O., Kludacz-Alessandri, M., & Wu, J. (2025). Immersive virtual reality based on head-mounted display in medical education: a systematic review. *BMC medical education*, *25*(1), 1593.

[40] He, H., Xu, X., Li, S., Bueno-Vesga, J. A., Duan, Y., & Gu, Y. (2026). Training nursing skills in a generative artificial intelligence-enhanced virtual reality patient encounter simulation: a qualitative study from a student perspective. *International Journal of Educational Technology in Higher Education*, *23*(1), 12.

[41] Akhtar, Z. B. (2023). Accelerated computing a biomedical engineering and medical science perspective. *Annals of the Academy of Romanian Scientists Series on Biological Sciences*, *12*(2), 138-164.

[42] Hanna, M. G., Pantanowitz, L., Dash, R., Harrison, J. H., Deebajah, M., Pantanowitz, J., & Rashidi, H. H. (2025). Future of artificial intelligence—machine learning trends in pathology and medicine. *Modern Pathology*, *38*(4), 100705.

[43] Sun, J., Liu, T., & Li, H. (2024). Study on the status and problems of teaching system of "medical advanced mathematics": data based on a research of 11 universities in China. *BMC Medical Education*, *24*(1), 36.

[44] Bhatti, M. A., & Pardaev, J. (2025). Integrating Statistical Software Usage into Biomathematics Courses: Impact on Student Research Skills. *Letters in Biomathematics*, *12*(1), 54-66.

[45] Williams, A., Sun, Z., & Vaccarezza, M. (2025). Comparison of augmented reality with other

teaching methods in learning anatomy: a systematic review. *Clinical Anatomy*, *38*(2), 168-185.

[46] Lastrucci, A., Wandael, Y., Barra, A., Ricci, R., Maccioni, G., Pirrera, A., & Giansanti, D. (2024). Exploring augmented reality integration in diagnostic imaging: Myth or reality?. *Diagnostics*, *14*(13), 1333.

[47] Patra, A., Asghar, A., Chaudhary, P., & Ravi, K. S. (2022). Integration of innovative educational technologies in anatomy teaching: new normal in anatomy education. *Surgical and radiologic anatomy*, *44*(1), 25-32.

[48] Zhao, N., & Fu, Q. (2025, April). Under the Framework of Deep Learning Cognitive Theory and Bloom's Taxonomy: Investigating the Role of Artificial Intelligence in Fostering Higher-order Thinking in Engineering Education. In *2025 7th International Conference on Computer Science and Technologies in Education (CSTE)* (pp. 892-896). IEEE.

[49] Kaldaras, L., Wang, K. D., Nardo, J. E., Price, A., Perkins, K., Wieman, C., & Salehi, S. (2024). Employing technology-enhanced feedback and scaffolding to support the development of deep science understanding using computer simulations. *International Journal of STEM Education*, *11*(1), 30.

[50] Piri, Z., & Cagiltay, K. (2024). Can 3-dimensional visualization enhance mental rotation (MR) ability?: A systematic review. *International Journal of Human–Computer Interaction*, *40*(14), 3683-3698.

[51] Zhang, Z., Ding, X., Liang, X., Zhou, Y., Qin, B., & Liu, T. (2025). Brain and cognitive science inspired deep learning: a comprehensive survey. *IEEE Transactions on Knowledge and Data Engineering*, *37*(4), 1650-1671.

[52] McLain, M. (2022). Towards a signature pedagogy for design and technology education: A literature

review. *International Journal of Technology and Design Education*, *32*(3), 1629-1648.

[53] Yao, D., & Lin, J. (2025). Cognitive enhancement through competency-based programming education: a 12-year longitudinal study. *Education and Information Technologies*, *30*(14), 20347-20383.

[54] Stock, W. G., Reichmann, G., Dorsch, I., & Schlögl, C. (2023). Counting research publications, citations, and topics: A critical assessment of the empirical basis of scientometrics and research evaluation. *Journal of Information Science Theory and Practice*, *11*(2), 37-66.