

Metaverse-based objective structured clinical examinations: an exploratory approach to advancing clinical competency assessment

Yeon-Ju Huh^{1,2}, Joon Sung Shin^{1,3}, Narae Yoon^{1,3}, Ju Whi Kim^{1,4}, Do Hoon Kim⁵, Chanwoong Kim⁶, Seoi Jeong^{7,8,9}, Yejin Yoon⁷, Soyeon Shin⁷, Hyoun-Joong Kong^{7,8,10}, Sun Jung Myung^{1,11}

¹Office of Medical Education, Seoul National University College of Medicine, Departments of ²Surgery, ³Psychiatry, and ⁴Pediatrics, Seoul National University Hospital, ⁵Department of Family Medicine, Korea University College of Medicine, ⁶Department of Emergency Medicine, Chung Ang University College of Medicine, ⁷Department of Transdisciplinary Medicine, Seoul National University Hospital, ⁸Department of Clinical Medical Sciences, Seoul National University, ⁹Institute of Convergence Medicine with Innovative Technology, Seoul National University Hospital, ¹⁰Department of Medicine, Seoul National University College of Medicine, and ¹¹Department of Internal Medicine, Seoul National University Hospital, Seoul, Korea

Received: November 22, 2025

Revised: February 10, 2026

Accepted: March 06, 2026

Corresponding Author:

Sun Jung Myung

(<https://orcid.org/0000-0001-7332-0126>)

Office of Medical Education, Seoul National University College of Medicine and Department of Internal Medicine, Seoul National University Hospital, 103 Daehak-ro, Jongno-gu, Seoul 03080, Korea

Tel: +82-2-740-8177

Fax: +82-2-740-8072

E-mail: issac73@snu.ac.kr

[†]This work was presented at the 2025 Korea Health Personnel Licensing Examination Institute Academic Seminar (May 13, 2025, Hanyang University Seoul Campus, HIT Grand Conference Hall).

Purpose: This developmental study explored the conceptual feasibility and applicability of a metaverse-based clinical assessment platform as a complementary tool to conventional objective structured clinical examinations in undergraduate medical education.

Methods: A targeted literature review and expert consensus process were conducted to identify domains of clinical competence in which metaverse technologies could provide added value. Based on these findings, prototype virtual patient simulations were developed within a metaverse environment. Large language models (LLMs) were integrated to support dynamic, interactive history-taking simulations, and pilot modules for physical examination were also created.

Results: Integration of LLMs into virtual patient scenarios enabled realistic, context-sensitive medical interviews, facilitating interactive dialogue between examinees and simulated patients. In contrast, physical examination modules faced technical limitations, particularly in replicating procedures requiring tactile or haptic feedback, such as palpation and percussion. Nevertheless, the metaverse environment enabled delivery of consistent and reproducible scenarios, supporting objective assessment of communication and diagnostic reasoning skills.

Conclusion: Metaverse-based simulations augmented by LLMs offer a promising approach to scalable and standardized clinical assessment, particularly within cognitive and interpersonal competency domains. Although current technological constraints limit the fidelity of physical examination simulations, rapid advancements in immersive and haptic technologies may help overcome these barriers in the near future. Further research is needed to evaluate the educational efficacy, validity, and feasibility of deploying such platforms in summative, high-stakes assessment contexts.

Key Words: Clinical competence, Educational measurement, Large language models, Medical education, Metaverse

Introduction

1. Background/rationale

The rapid advancement of digital technologies has profoundly influenced medical education and professional training, transitioning traditional pedagogical models into more immersive and interactive learning environments [1]. Among these emerging technologies, the metaverse—defined as a three-dimensional, interconnected virtual space enabling real-time, user-driven interaction—has gained attention for its potential to enhance simulation-based education and assessment in healthcare [2-4].

Rigorous assessment of clinical competence remains a cornerstone of medical education, ensuring that examinees are adequately prepared to provide safe, effective, and professional care [5]. Within high-stakes licensing examinations, performance-based evaluations play a pivotal role in verifying the acquisition of clinical skills, diagnostic reasoning, and communication competencies [6]. The objective structured clinical examination (OSCE) has been widely adopted for this purpose because of its structured and systematic design, which allows standardized observation of clinical tasks across multiple domains [7,8]. Despite their widespread use, OSCEs have several well-documented limitations, including examiner variability, substantial resource requirements, logistical constraints, and a limited capacity to simulate diverse or dynamic clinical scenarios [8-10]. These challenges may compromise the reliability, validity, and scalability of clinical performance assessments, particularly in large-scale or geographically distributed educational settings [9,10].

Metaverse technologies offer a novel approach to addressing these constraints by enabling the creation of standardized, reproducible, and context-rich virtual clinical environments. Such platforms can simulate complex patient interactions and procedural tasks with greater consistency while also supporting remote access and asynchronous participation [11,12]. When integrated with artificial intelligence (AI) tools, such as large language models (LLMs), these environments can further support dynamic patient interactions and responsive history taking, thereby expanding the range of competencies that can be assessed beyond static, checklist-driven formats [13].

2. Objectives

The present study does not aim to establish the effectiveness of metaverse- or AI-enabled approaches for assessing medical students' clinical competence, nor to determine whether such technologies can replace conventional OSCE formats. Because the capabil-

ities and implementation conditions of these technologies are rapidly evolving, early empirical evaluations of specific platforms may have limited generalizability across contexts and time. Accordingly, this study adopts an exploratory orientation to (1) delineate key limitations of current OSCE practice as articulated by experts with sustained experience in OSCE station development and high-stakes examination administration, and (2) identify domains in which the affordances of metaverse environments and AI may, in principle, contribute to addressing these limitations. Through a literature review, expert consultation, and thematic analysis, the study seeks to define core competency areas that may benefit from currently available AI-driven, metaverse-based simulation technologies. We contend that such problem specification and conceptual mapping are essential precursors to subsequent purpose-aligned design and rigorous validation of technology-enabled assessment approaches.

Methods

1. Study design

This study employed a developmental research design to explore the feasibility and applicability of a metaverse-based clinical competency assessment system. The study was conducted in three sequential phases: (1) identification of core competencies and scenarios suitable for metaverse adaptation through expert consensus, (2) development of a virtual simulation platform incorporating LLMs, and (3) comparative analysis of the educational and technical characteristics of the developed system relative to conventional OSCEs.

2. Consensus workshop and metaverse-compatible competency identification

In the first phase, consensus-building expert workshops were conducted to identify clinical competencies and scenarios most suitable for metaverse-based assessment. A multidisciplinary panel of 10 experts included medical educators and specialists in internal medicine, surgery, emergency medicine, psychiatry, pediatrics, family medicine, as well as system developers. The workshops began with structured brainstorming sessions examining the limitations of traditional clinical examinations and assessment items, followed by discussions on the potential advantages and constraints of metaverse environments in clinical assessment.

Based on a comprehensive literature review and converging expert opinions, the panel identified clinical competencies particularly suited for metaverse-based assessment. These findings guided the selection of appropriate clinical scenarios, including jaundice

with ascites and abdominal pain, a child presenting with shortness of breath, and a case involving neurological abnormalities. The panel prioritized development of a scenario involving jaundice with ascites and abdominal pain, as this presentation includes physical findings that are difficult to reproduce consistently using standardized patients (SPs). The scenario was initially drafted by an internal medicine specialist and subsequently refined through detailed review by six experts in medical education, internal medicine, and surgery to ensure clinical accuracy and educational alignment.

3. Development of a metaverse-based simulation platform

In the second phase, a metaverse-based case assessment system was developed to simulate the selected clinical scenarios within a virtual environment (Fig. 1). The platform was specifically designed for this study as a clinical assessment tool, rather than as an independent generative AI application or a direct adaptation of a commercial chatbot. Although an LLM was incorporated to facilitate natural language interaction, its role was limited to enabling dialogue, while clinical accuracy and scenario consistency were ensured through predefined, expert-reviewed case data. The system was structured to closely replicate actual clinical workflows through two core interactive modules.

The first module focused on history taking. Within this module, examinees could ask clinical questions of avatar-based virtual patients, who provided contextually appropriate responses in real time. These responses were generated by an LLM-based conversa-

tional AI system designed to interpret examinee questions, deliver scenario-specific information, convey appropriate emotions, and maintain contextual continuity across repeated interactions. To ensure clinical accuracy and scenario fidelity, a retrieval-augmented generation (RAG) framework was implemented using a vector database architecture [14,15]. When examinees posed questions, the system performed semantic similarity searches to retrieve the most relevant case information, which was then dynamically incorporated into the LLM prompt. This granular vectorization strategy allowed precise retrieval of pertinent information while preserving scenario security and response consistency.

The prompting framework incorporated mechanisms to manage queries that fell outside the documented case scope. When similarity scores indicated that a question pertained to undocumented information, the system guided the LLM to respond with clinically realistic expressions of uncertainty (e.g., “I’m not sure about that,” “I don’t recall specifically”), thereby preserving scenario authenticity and simulating natural patient communication. To maintain assessment validity and prevent academic misconduct, multi-layered safeguards were implemented within the RAG-prompt pipeline [16]. These included explicit restrictions against revealing complete case scenarios or diagnostic conclusions. Query pattern detection flagged suspicious requests (e.g., “tell me everything about your condition”), prompting responses that redirected examinees toward systematic clinical inquiry. Retrieval filtering further excluded case record sections containing diagnostic conclusions or examiner-only notes, ensuring that only patient-facing information informed LLM responses. All flagged query patterns were logged for post-assessment review.

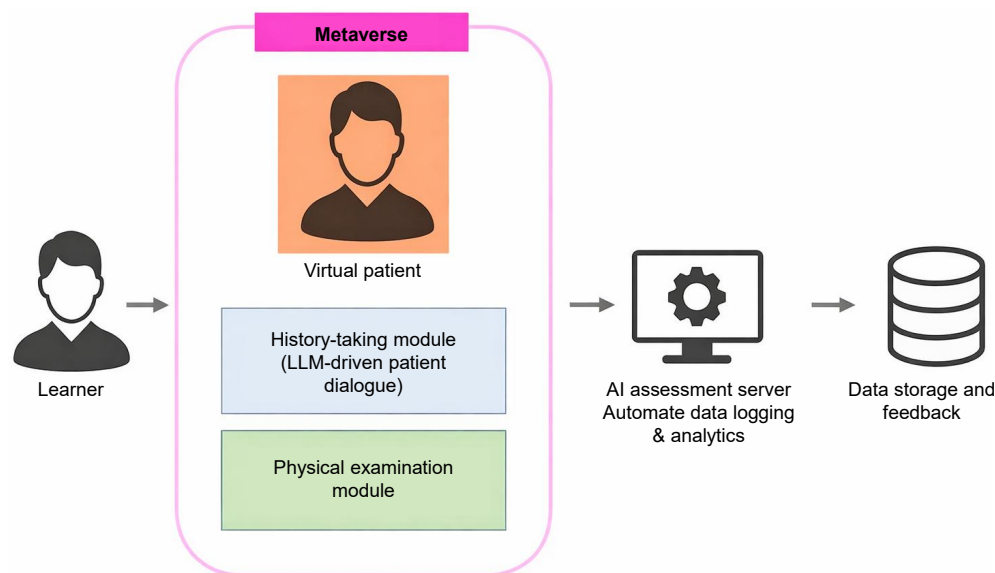


Fig. 1. Structure of a metaverse-based clinical competency assessment system. LLM: Large language models.

For example, when an examinee asked about the location, character, and radiation of pain in a patient presenting with abdominal discomfort, the RAG system retrieved relevant vectorized entries regarding pain characteristics. The virtual patient then provided clinically accurate responses using refined language, offering examinees a near-authentic history-taking experience while strictly adhering to the predefined clinical narrative.

The second module facilitated physical examination within the virtual environment. To enhance immersion and better replicate natural clinical workflows, the system utilized hand-tracking gesture recognition instead of conventional controller inputs. When examinees' hands interacted with examination targets, such as the patient's eyes or abdomen, visual feedback indicated successful contact and guided proper technique [17,18]. The virtual patient was designed to display characteristic signs of chronic liver disease—including icteric sclera and skin, sarcopenic extremities, and abdominal distension—supporting diagnostic reasoning that is difficult to reproduce with SPs. The module accommodated inspection, auscultation, percussion, and palpation, producing appropriate audiovisual responses when examinees interacted with relevant areas. During abdominal auscultation, spatially localized bowel sounds were generated when the stethoscope diaphragm contacted the abdomen. The percussion feature produced different tones depending on the targeted region, accurately reflecting the presence of ascites. During abdominal palpation, visual feedback indicated compression at the contact point, with image-based cues of abdominal tension or depression serving as partial substitutes for tactile information in the absence of haptic feedback. Although these functions cannot fully replicate a physical examination, they provide essential integrated information to effectively engage examinees' clinical reasoning and judgment within the virtual environment. The system was also engineered to record and analyze examinee interactions and performance data in real time, creating an objective and standardized assessment environment.

4. Comparative analysis of the advantages and limitations of the metaverse-based assessment system

Following prototype development, a comparative analysis was conducted to evaluate the advantages and limitations of the metaverse-based assessment system relative to conventional OSCE formats. This analysis, carried out through iterative expert review and structured comparison, focused on features related to assessment implementation rather than overall educational outcomes or psychometric performance. Specifically, experts compared the two formats in terms of (1) the feasibility and realism of patient–doctor interaction scenarios, (2) the degree of standardization and repro-

ducibility of examinee–patient interactions compared with SP portrayals, (3) the scalability of assessment delivery, and (4) the capacity for systematic and objective data capture during clinical encounters. At the same time, implementation-related limitations—such as restricted tactile feedback, limited emotional expressiveness of virtual avatars, and technical infrastructure requirements—were identified through expert evaluation informed by prototype use. This phase focused on exploring conceptual feasibility and developing a prototype through expert consensus. Formal usability testing involving learners, examinees, and faculty was intentionally deferred and will be conducted in subsequent validation studies.

Results

1. Consensus about the limitations and challenges of the current clinical skills examination

Clinical skills examinations, such as the OSCE, are widely used to assess competencies necessary for clinical practice. However, concerns persist regarding whether these exams authentically capture the complexity of real-world clinical environments. The consensus workshop identified several systemic issues, major limitations of the current OSCE, and potential directions for reform (Table 1). Key issues highlighted by the workshop included the following:

1) Restricted range of clinical presentation

Because current OSCEs rely on SPs, the range of clinical scenarios that can be simulated is inherently limited.

2) Test reliability

The use of SPs often necessitates checklist-based scoring rather than rating scales. Despite structured SP training and scoring rubrics, concerns about assessment reliability have been consistently raised by examinees.

3) Preparation burden

Students face substantial financial and time demands when preparing for OSCEs, often at the expense of authentic clinical learning. Institutions also encounter significant challenges in administering the examinations, with associated costs remaining high.

2. Selection of topics for the study

Although the OSCE is valuable for standardizing assessment, it has limitations in reproducing the complexity of real-world clinical scenarios. To address these gaps, we explored the potential of

Table 1. Comparison between Conventional OSCE and Metaverse-Based Assessment

Domain	Conventional OSCE	Metaverse-based assessment
Standardization & reproducibility	Variability in SP performance and examiner scoring	Identical scenarios, automated data logging
Reliability	Checklist-based, prone to examiner bias	Objective through log data, supports global rating scales
Educational impact	One-shot exam, limited formative feedback	Repetitive practice, personalized feedback, self-directed learning
Operational efficiency	High cost (human, facilities, time), concentrated scheduling	Continuous delivery, reduced infrastructure needs
Applicable cases	Common symptoms only; limited in pediatrics, rare, or risky cases	Expanded to rare diseases, pediatrics, ethical/risky cases
Main constraints	High authenticity but costly and resource-intensive	No haptic feedback, limited emotional realism, high-tech requirements

OSCE: Objective structured clinical examination, SP: Standardized patient.

Table 2. Examples of Clinical Scenarios Suitable for Metaverse-Based Assessment

Selection criteria	Clinical scenario examples	Educational implications
High incidence and educational value	Chest pain, abdominal pain, dyspnea	Frequent in practice, allows diagnostic reasoning training
Requires abnormal physical findings difficult for SPs	Abdominal pain with jaundice and ascites	Critical exam findings not reproducible with SPs
High cognitive complexity	Neurological deficits (hemiparesis, sensory loss)	Training in lesion localization and differential diagnosis
SP restrictions	Pediatric cases (dyspneic child, febrile seizure)	Overcomes ethical barriers of pediatric SP use
Ethical or high-risk cases	Violent patient, patients at risk of suicide, bad news delivery	Important but difficult scenarios for real OSCE

SP: Standardized patient, OSCE: Objective structured clinical examination.

metaverse technology to supplement traditional OSCEs. Through a structured consensus workshop, key clinical scenarios suitable for virtual simulation were identified, and a metaverse-based platform was piloted to enhance the authenticity and scalability of clinical skills evaluation. The workshop selected metaverse scenarios based on three criteria: (1) common presenting symptoms in clinical practice, such as chest pain, abdominal pain, and dyspnea; (2) feasibility of replicating these scenarios within a virtual simulation environment, particularly for assessing clinical reasoning and communication skills; and (3) cognitive complexity, emphasizing cases that require advanced diagnostic differentiation and higher-order clinical thinking.

Four topics were selected (Table 2): (1) abdominal pain with abnormal physical examination findings, where these findings are essential for differential diagnosis but cannot be realistically simulated using SPs; (2) dyspnea with multiple possible etiologies, a clinical presentation that physicians must recognize and manage with appropriate treatment planning; (3) cases involving various neurological abnormalities, in which accurate depiction of abnormal neurological examination findings is crucial for lesion localization and diagnosis; and (4) pediatric cases, which are challenging to implement due to the limitations of using pediatric SPs.

These topics were chosen because they represent clinical situations that are difficult to reproduce in traditional OSCEs due to resource constraints and procedural complexity.

3. The strengths and limitations of the metaverse-based platform used in this study

The implemented system exhibited several educational and operational strengths (Figs. 2, 3). Most notably, it enabled immersive and interactive clinical encounters that closely mirrored real-world decision-making. Integration of LLMs allowed dynamic patient communication, while the automated performance-logging system enhanced the reproducibility and objectivity of assessments. Furthermore, the virtual format supported simultaneous participation by multiple learners or examinees without requiring physical infrastructure or SP staffing, thereby improving scalability and cost efficiency. However, several limitations were noted. The lack of haptic feedback remained a major constraint, preventing full replication of tactile aspects of physical examination, such as assessing tenderness, rigidity, or rebound tenderness. Emotional expressiveness of avatar patients was also limited, reducing the realism of scenarios that require empathy, affective communication, or rap-

port-building. Additionally, the platform depended on high-performance computing equipment and stable internet connections, which may pose barriers to broader adoption. Some learners and faculty who were unfamiliar with virtual interfaces experienced a learning curve, highlighting the need for user-centered design refinements and training prior to implementation. Therefore, these results should be interpreted as observations related to implementation rather than as evidence of superiority in educational impact or assessment outcomes.

Discussion

This study explores the potential feasibility and conceptual applicability of a metaverse-based clinical assessment platform enhanced by LLMs, grounded in expert analysis of structural limitations in current OSCE practice. This study was not designed to determine whether metaverse- or AI-based systems can substitute for conventional OSCE formats. Rather, it synthesizes expert perspectives to characterize structural and operational constraints inherent in contemporary OSCE delivery and to examine, at a conceptual level, the extent to which particular technological affordances may offer

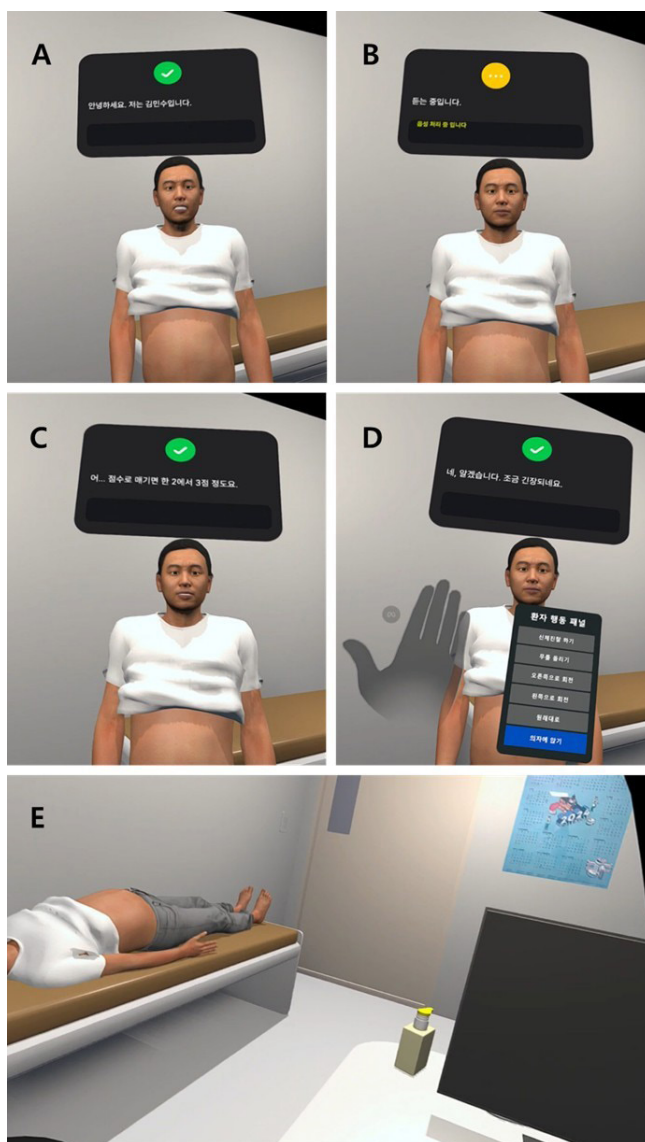


Fig. 2. Screenshot demonstrating the clinical flow in the metaverse history-taking module: (A) history-taking (self-introduction), (B) speech recognition, (C) history-taking (pain intensity), (D) patient activity panel, and (E) preparation for physical examination.

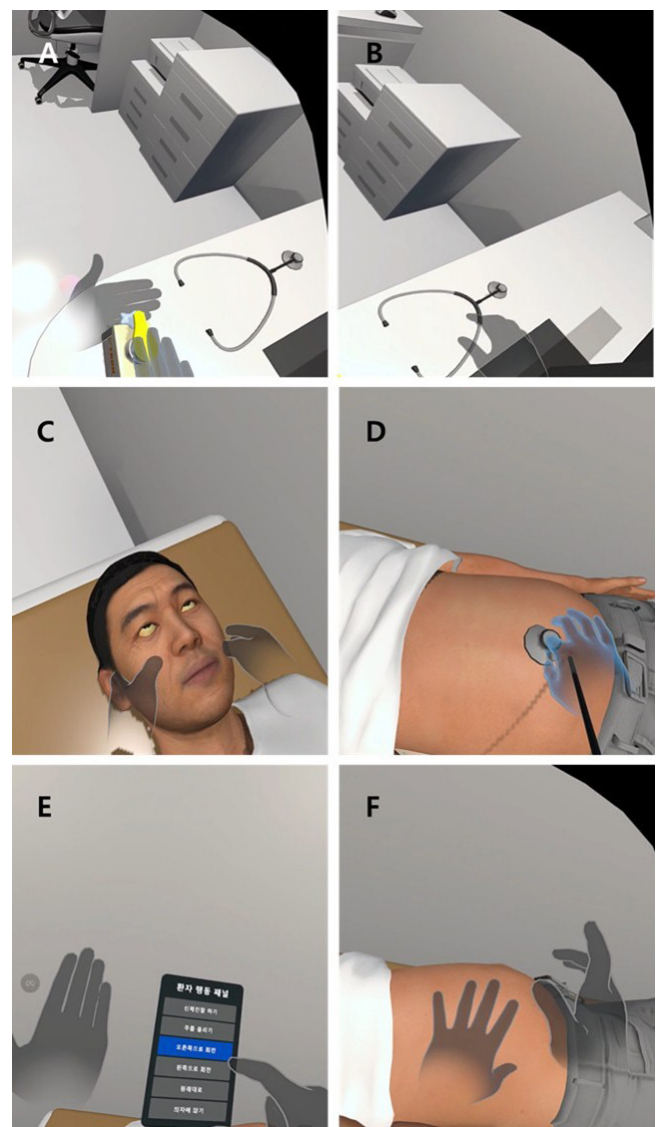


Fig. 3. Screenshot of key features in the metaverse physical examination module system: (A) hand disinfection, (B) stethoscope application, (C) scleral examination, (D) abdominal auscultation, (E) patient activity panel (position change), and (F) abdominal percussion (shifting dullness).

plausible strategies for mitigation. In this way, the study foregrounds assessment purpose and validity considerations, and treats technological implementation as a contingent means to address explicitly defined assessment challenges, rather than as an end in itself. In the context of increasing demands for standardization, scalability, and reproducibility in competency-based assessments [19,20], our findings suggest that metaverse simulations may complement traditional OSCEs by addressing selected long-standing structural limitations. Importantly, these findings reflect differences in assessment implementation and scenario delivery between the two formats, rather than comparisons of educational effectiveness, scoring processes, or passing score determinations.

The developed platform allowed examinees to engage in dynamic history taking with AI-driven virtual patients and to perform basic virtual physical examinations within immersive three-dimensional environments. Scenarios such as abdominal pain accompanied by jaundice and ascites—difficult to replicate using human SPs—were effectively simulated. This flexibility substantially expands the range of assessable conditions, particularly for rare, complex, or logistically challenging cases in conventional OSCE settings. A key advantage of the platform is its enhanced standardization. Unlike traditional OSCEs, which can be affected by variability in SP behavior and examiner judgment [9,10,21], virtual scenarios can be delivered consistently, reducing measurement error and supporting fairness. Furthermore, automated capture of examinee interactions enables objective evaluation and provides detailed assessment analytics for personalized formative feedback, which is often limited in conventional examinations [22,23].

Beyond checklist-based evaluation, the platform supports expert-led global rating scales that more effectively capture the depth of clinical reasoning and decision-making [24]. From an assessment perspective, metaverse-based clinical encounters are particularly well suited for rating scale-based evaluation in cognitive and communicative domains. Unlike checklist-driven assessments that focus on discrete task completion, immersive virtual encounters allow continuous observation of how examinees gather information, synthesize clinical data, adapt their reasoning, and communicate decisions throughout the encounter. These longitudinal, context-rich interactions enable examiners to apply global rating scales that reflect the quality, coherence, and adaptability of clinical reasoning and communication processes. This approach aligns with contemporary principles in competency-based medical education, supporting assessments that evaluate not only what learners do but also how they think. Prior research has similarly indicated that global rating scales are better suited for evaluating complex constructs such as clinical reasoning and communication

when sufficient clinical context is provided [25,26].

In addition to these assessment-specific advantages, the metaverse environment offers broader benefits related to assessment delivery and patient safety. Its immersive and interactive nature supports richer assessment interactions, enabling examinees' responses to unfold in ways that more closely resemble real-world clinical encounters [27]. Real-time AI patient responses and visually consistent environments facilitate observation of communication and active problem-solving processes [28], while remote accessibility and the ability to run parallel sessions reduce logistical constraints and support scalable assessment delivery [11]. Importantly, the platform allows performance in high-risk scenarios to be assessed without placing real patients at risk. Examinees' handling of complex situations, including the recognition of clinical errors, can be systematically captured and reviewed, aligning with the growing emphasis on incorporating patient safety competencies into simulation-based assessments [29,30].

Although the metaverse-based platform allows parallel assessment sessions without relying on physical examination rooms or SPs, this study did not include a formal cost-effectiveness analysis. Initial platform development, acquisition of virtual reality (VR) hardware, system maintenance, and ongoing costs related to LLM-based data processing represent important economic considerations for large-scale implementation. Additionally, while the system architecture was designed to accommodate multiple examinees simultaneously, the maximum number of concurrent users that can be supported without performance degradation has not yet been empirically determined. Therefore, statements regarding cost-effectiveness and scalability should be viewed as potential structural advantages rather than confirmed outcomes. Future research will be needed to conduct systematic economic evaluations and technical load testing to assess the feasibility of large-scale deployment in high-stakes assessment settings.

From a broader assessment perspective, high-stakes clinical performance examinations can be considered a composite system composed of multiple interrelated processes, including item development, clinical case and scenario creation, SP training, test administration, rating and scoring, and passing score determination. Within this framework, the present developmental study primarily contributes to early and mid-stage assessment processes, particularly case scenario creation, standardized scenario delivery, and interactive data capture during examinee–patient encounters. By leveraging a metaverse environment and LLMs, the platform supports reproducible and scalable implementation of complex clinical scenarios that are challenging to operationalize using conventional OSCE formats alone. This study does not address all components required for high-stakes implementation. Elements such

as examiner training, scoring standardization, passing score determination, and psychometric evaluation remain outside the scope of this phase and warrant further investigation to assess reliability, validity, educational impact, and acceptability within a comprehensive assessment system.

In addition, important limitations related to physical examination fidelity should be acknowledged. A metaverse-based platform can complement conventional OSCE stations that rely on written “result sheets” for findings that cannot be portrayed, by providing visual representations of clinical signs (e.g., jaundice and ascites) and a standardized, interactive workflow consistent with real clinical encounters. However, it cannot replace assessment components that require tactile or haptic feedback (e.g., palpation or percussion), limiting its suitability for certain physical examination competencies. In this prototype developmental phase, structured feedback from learners or examinees was not collected; therefore, the platform’s usability and acceptability remain to be established. Future studies will include formal evaluations of usability and acceptability, as well as direct comparisons with SP-based OSCE formats, particularly for clinical scenarios in which written result sheets are currently used to substitute for physical findings that cannot be directly demonstrated.

Despite these advantages, several limitations require careful consideration. The absence of haptic feedback prevents assessment of tactile components, such as palpation or percussion. Emotional realism is also constrained, as virtual avatars lack the nuanced affective expression of human actors, reducing authenticity in empathy-focused encounters. Technical challenges remain significant: high-quality VR hardware, stable internet connectivity, and institutional information technology infrastructure are essential for reliable implementation, and any technical issues may impact assessment performance. To ensure fairness in future high-stakes deployment, a standardized onboarding and training protocol will be necessary to minimize the learning curve associated with metaverse navigation and device operation. This would include a brief orientation to the interface and controls, practice with a short non-scored sample case, and verification of technical readiness (e.g., device fit, audio connectivity) prior to the assessment. Standardizing this process may reduce construct-irrelevant variance related to prior familiarity with the metaverse and enhance comparability among examinees. Furthermore, the integration of generative AI requires robust content validation to ensure clinical accuracy and prevent dissemination of misinformation.

Metaverse-based assessment has the potential to function as a complementary tool for national licensing and certification examinations. However, its integration into high-stakes contexts should be preceded by multi-institutional validation studies to establish

construct and content validity, correlation with outcomes, and operational feasibility. Collaboration with regulatory bodies will be essential to define quality assurance protocols, security standards, and guidelines for virtual case design. In the near term, a hybrid model appears most practical—employing virtual stations to assess cognitive, diagnostic, and communication competencies, while retaining in-person OSCEs for procedural and affective skills. This approach bridges the gap between assessment and authentic clinical performance by incorporating dynamic clinical encounters while preserving fidelity where direct patient interaction is critical. Simultaneously, it leverages the scalability and standardization advantages of metaverse-based assessment, supporting the development of adaptable and clinically competent physicians. Beyond summative assessment, the platform may also facilitate formative evaluation by enabling repeated observation of performance patterns and structured feedback aligned with assessment criteria [23]. When integrated into longitudinal assessment frameworks, such systems may help identify performance gaps and support targeted remediation [11,31]. Future developments could include assessment of interprofessional competencies. By simulating team-based care scenarios, the platform could serve as a tool to evaluate teamwork, communication, and collaborative decision-making across healthcare roles [27], aligning with global trends in competency-based assessment.

To fully realize this potential, continued innovation is needed. Advances in haptic technology, avatar emotional expressiveness, and multimodal AI could help close the fidelity gap in physical examinations and empathetic communication. AI-driven scoring models based on structured interaction logs may eventually enable automated, real-time assessment with reliability comparable to expert evaluators. Additionally, shared virtual case libraries and institutional consortia could further enhance scalability and reduce development costs.

In conclusion, the metaverse-based simulation platform represents a forward-looking and versatile approach that addresses many limitations of current clinical assessments. With strategic integration, ongoing technical refinement, and alignment with policy frameworks it has the potential to become a foundational component of medical education—enhancing fairness, broadening access, and supporting safer and more effective clinical training for the next generation of physicians.

ORCID

Yeon-Ju Huh: <https://orcid.org/0000-0001-6745-7097>

Joon Sung Shin: <https://orcid.org/0000-0002-2425-4406>

Narae Yoon: <https://orcid.org/0000-0002-5415-4617>

Ju Whi Kim: <https://orcid.org/0000-0001-8011-3547>

Do Hoon Kim: <https://orcid.org/0000-0001-7421-4501>
 Chanwoong Kim: <https://orcid.org/0000-0001-7821-8980>
 Seo Jeong: <https://orcid.org/0009-0009-3047-2889>
 Yejin Yoon: <https://orcid.org/0009-0000-2422-7371>
 Soyeon Shin: <https://orcid.org/0009-0000-2861-3202>
 Hyoun-Joong Kong: <https://orcid.org/0000-0001-5456-4862>
 Sun Jung Myung: <https://orcid.org/0000-0001-7332-0126>

Data sharing statement

Please contact the corresponding author for data availability.

Acknowledgements

None.

Funding

This work was supported by the Korea Health Personnel Licensing Examination Institute of Korea in 2024 (No. RE12-2505-01).

Conflicts of interest

No potential conflict of interest relevant to this article was reported.

Author contributions

Conceptualization: YJH, HJK, SJM. Data curation: YJH, JSS, NY, SJ, YY, SS, HJK, SJM. Formal analysis: YJH, JSS, NY, JWK, DHK, CK, SJ, YY, SS, HJK, SJM. Interpretation: YJH, JSS, NY, JWK, DHK, CK, SJ, YY, SS, HJK, SJM. Funding acquisition: SJM. Writing—original draft: YJH, SJM. Writing—review & editing: YJH, JSS, NY, JWK, DHK, CK, SJ, YY, SS, HJK, SJM. Final approval of the manuscript: all authors.

References

- Prosen M, Ličen S. Evaluating the digital transformation in health sciences education: a thematic analysis of higher education teachers' perspectives. *BMC Med Educ.* 2025;25(1):820. <https://doi.org/10.1186/s12909-025-07420-3>
- Zhang X, Chen Y, Hu L, Wang Y. The metaverse in education: definition, framework, features, potential applications, challenges, and future research topics. *Front Psychol.* 2022;13:1016300. <https://doi.org/10.3389/fpsyg.2022.1016300>
- Wang Y, Zhu M, Chen X, et al. The application of metaverse in healthcare. *Front Public Health.* 2024;12:1420367. <https://doi.org/10.3389/fpubh.2024.1420367>
- Perez-Baena AV, Rudolphi-Solero T, Lorenzo-Alvarez R, Dominguez-Pinos D, Ruiz-Gomez MJ, Sendra-Portero F. Evaluation of a virtual objective structured clinical examination in the metaverse (Second Life) to assess the clinical skills in emergency radiology of medical students in Spain: a cross-sectional study. *J Educ Eval Health Prof.* 2025;22:12. <https://doi.org/10.3352/jeehp.2025.22.12>
- Epstein RM. Assessment in medical education. *N Engl J Med.* 2007;356(4):387-396. <https://doi.org/10.1056/NEJMra054784>
- Bhanji F, Naik V, Skoll A, et al. Competence by design: the role of high-stakes examinations in a competence based medical education system. *Perspect Med Educ.* 2024;13(1):68-74. <https://doi.org/10.5334/pme.965>
- Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ.* 1979;13(1):41-54. <https://doi.org/10.1111/j.1365-2923.1979.tb00918.x>
- Khan KZ, Ramachandran S, Gaunt K, Pushkar P. The objective structured clinical examination (OSCE): AMEE Guide No. 81. Part I: an historical and theoretical perspective. *Med Teach.* 2013;35(9):e1437-e1446. <https://doi.org/10.3109/0142159X.2013.818634>
- Brannick MT, Erol-Korkmaz HT, Prewett M. A systematic review of the reliability of objective structured clinical examination scores. *Med Educ.* 2011;45(12):1181-1189. <https://doi.org/10.1111/j.1365-2923.2011.04075.x>
- Patrício MF, Julião M, Fareleira F, Carneiro AV. Is the OSCE a feasible tool to assess competencies in undergraduate medical education? *Med Teach.* 2013;35(6):503-514. <https://doi.org/10.3109/0142159X.2013.774330>
- Popov V, Mateju N, Jeske C, Lewis KO. Metaverse-based simulation: a scoping review of charting medical education over the last two decades in the lens of the 'marvelous medical education machine'. *Ann Med.* 2024;56(1):2424450. <https://doi.org/10.1080/07853890.2024.2424450>
- Burlacu A, Brinza C, Horia NN. How the Metaverse is shaping the future of healthcare communication: a tool for enhancement or a barrier to effective interaction? *Cureus.* 2025;17(3):e80742. <https://doi.org/10.7759/cureus.80742>
- Cook DA. Creating virtual patients using large language models: scalable, global, and low cost. *Med Teach.* 2025;47(1):40-42. <https://doi.org/10.1080/0142159X.2024.2376879>
- Kim J, Kim SJ, Ahn J, Lee S. LLM-based response generation for Korean adolescents: a study using the NAVER Knowledge iN Q&A Dataset with RAG. *Healthc Inform Res.* 2025;31(2):136-145. <https://doi.org/10.4258/hir.2025.31.2.136>
- Son N, Kang I, Kim I, Lee K, Nam S, Lee D. Development and evaluation of a retrieval-augmented generation-based electronic medical record chatbot system. *Healthc Inform Res.* 2025;31(3):218-225. <https://doi.org/10.4258/hir.2025.31.3.218>
- Seo J, Park S, Byun S, Choi J, Choi J, Shin H. Advancing Korean medical large language models: automated pipeline for Korean medical preference dataset construction. *Healthc Inform Res.* 2025;31(2):166-174. <https://doi.org/10.4258/hir.2025.31.2.166>

17. Buckingham G. Hand tracking for immersive virtual reality: opportunities and challenges. *Front Virtual Real.* 2021;2:728461. <https://doi.org/10.3389/frvir.2021.728461>
18. Saran M. Comparing hand-based and controller-based interactions in virtual reality learning: effects on presence and interaction performance. *PeerJ Comput Sci.* 2025;11:e3168. <https://doi.org/10.7717/peerj-cs.3168>
19. Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. *Med Teach.* 2010;32(8):676-682. <https://doi.org/10.3109/0142159X.2010.500704>
20. Norcini J, Anderson MB, Bollela V, et al. 2018 Consensus framework for good assessment. *Med Teach.* 2018;40(11):1102-1109. <https://doi.org/10.1080/0142159X.2018.1500016>
21. Pell G, Fuller R, Homer M, Roberts T. How to measure the quality of the OSCE: a review of metrics: AMEE Guide No. 49. *Med Teach.* 2010;32(10):802-811. <https://doi.org/10.3109/0142159X.2010.507716>
22. Ellaway RH, Topps D, Pusic M. Data, big and small: emerging challenges to medical education scholarship. *Acad Med.* 2019;94(1):31-36. <https://doi.org/10.1097/ACM.0000000000002465>
23. Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R. Mastery learning for health professionals using technology-enhanced simulation: a systematic review and meta-analysis. *Acad Med.* 2013;88(8):1178-1186. <https://doi.org/10.1097/ACM.0b013e31829a365d>
24. Huang TY, Hsieh PH, Chang YC. Performance comparison of junior residents and ChatGPT in the objective structured clinical examination (OSCE) for medical history taking and documentation of medical records: development and usability study. *JMIR Med Educ.* 2024;10:e59902. <https://doi.org/10.2196/59902>
25. Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. OSCE checklists do not capture increasing levels of expertise. *Acad Med.* 1999;74(10):1129-1134. <https://doi.org/10.1097/00001888-199910000-00017>
26. Ten Cate O, Regehr G. The power of subjectivity in the assessment of medical trainees. *Acad Med.* 2019;94(3):333-337. <https://doi.org/10.1097/ACM.0000000000002495>
27. Li Q, Duan H, Zhou X, Sun X, Tao L, Lu X. The use of metaverse in medical education: a systematic review. *Clin Med (Lond).* 2025;25(3):100315. <https://doi.org/10.1016/j.clinme.2025.100315>
28. Chance EA. The combined impact of AI and VR on interdisciplinary learning and patient safety in healthcare education: a narrative review. *BMC Med Educ.* 2025;25(1):1039. <https://doi.org/10.1186/s12909-025-07589-7>
29. Elendu C, Amaechi DC, Okatta AU, et al. The impact of simulation-based training in medical education: a review. *Medicine (Baltimore).* 2024;103(27):e38813. <https://doi.org/10.1097/MD.00000000000038813>
30. Sim JJM, Rusli KDB, Seah B, Levett-Jones T, Lau Y, Liaw SY. Virtual simulation to enhance clinical reasoning in nursing: a systematic review and meta-analysis. *Clin Simul Nurs.* 2022;69:26-39. <https://doi.org/10.1016/j.ecns.2022.05.006>
31. Wang Y, Li Y, Chen C, et al. Research on virtual reality-based assessment framework and application path in medical education. *PLoS One.* 2024;19(11):e0310782. <https://doi.org/10.1371/journal.pone.0310782>