Research Article

Strengths and Limitations of Using ChatGPT in OSCE: A Preliminary Examination of Generative AI in Medical Education

Charlotte A. Taylor-Drigo*,1 and Anshul Kumar2

Abstract

Introduction: Objective Structured Clinical Examinations (OSCEs) are essential components of medical education, designed to assess clinical competence through structured tasks such as history-taking, physical examinations, and patient communication across multiple stations. Examiners utilize standardized rubrics to ensure fairness and objectivity in evaluation. The COVID-19 pandemic accelerated the use of technology in OSCEs, with virtual platforms introduced to maintain assessments while observing safety protocols. These changes highlighted the need for innovative, interactive, and realistic simulations. Artificial intelligence (AI) tools such as ChatGPT offer promising opportunities in this context. With advanced conversational abilities, ChatGPT can replicate patient interactions and provide immediate feedback, fostering active learning, cognitive engagement, and experiential skill development. Grounded in established educational frameworks, ChatGPT represents a novel strategy to augment OSCEs by strengthening history-taking training and enhancing the assessment of clinical competence.

Method: A pilot study was conducted with 20 faculty members responsible for designing and evaluating OSCE scenarios. Participants engaged with ChatGPT in three simulated cases structured to resemble traditional OSCE encounters. Following the sessions, participants completed a survey via Qualtrics to evaluate ChatGPT's usability and effectiveness in supporting history-taking exercises.

Results: Faculty valued ChatGPT's ability to serve as a consistent, responsive simulated patient, noting its role in improving clinical reasoning while minimizing intimidation. Limitations included the absence of non-verbal communication, limited empathy, and the inability to perform physical examinations. Technical inconsistencies also posed challenges. While 20% of participants expressed interest in future integration, most favored a hybrid model combining AI with standardized patients to balance realism with experiential learning.

Conclusion: Integrating ChatGPT into OSCEs provides an innovative approach to medical education, with the potential to enrich assessment accuracy and enhance student preparedness for real-world clinical practice.

Keywords: OSCE, Artificial Intelligence, Medical education, Simulation in medical training, Simulation in healthcare

Background

The Objective Structured Clinical Examination (OSCE) is a critical assessment tool in medical education, designed to evaluate the clinical competence of medical students [1]. Developed by Harden and Gleeson in

1979, OSCE requires students to rotate through structured stations to demonstrate skills such as history-taking, physical examination, and patient communication. Examiners use standardized checklists and rubrics to ensure consistent and reliable evaluation of student performance [1]. The COVID-19 pandemic significantly disrupted traditional educational and healthcare models, accelerating the adoption of technology as an essential means of maintaining continuity Virtual OSCEs emerged as a practical alternative to in-person assessments, enabling continued evaluation of students while adhering to social distancing guidelines [2]. These virtual assessments, typically conducted via video conferencing, highlighted the need for interactive and realistic simulations to maintain examination fidelity. Moreover, Generation Me students prefer structured yet interactive learning experiences, further emphasizing the necessity of technologically enhanced educational tools [3,4].

Artificial intelligence (AI) has gained prominence as a transformative tool in medical education, particularly in OSCEs [5]. AI-driven applications, such as ChatGPT by OpenAI, offer innovative solutions for clinical education by simulating realistic patient interactions, providing real-time feedback, and supporting the development of clinical reasoning skills [2,9,11]. The ability of ChatGPT to generate human-like responses makes it a suitable candidate for OSCE integration, enhancing both learning experiences and assessment accuracy. Research supports the benefits of AI-based OSCE simulations. Virtual standardized patients effectively assess students' information-gathering skills [12], and AI-driven OSCE simulations have received positive student feedback, particularly in improving clinical reasoning [2,11]. AI-based simulations also cater to the learning preferences of millennial and Gen Z students, who favor flexible, technology-based resources [3,5]. By providing on-demand, standardized patient interactions, AI enables students to practice independently without scheduling constraints. Furthermore, AI-based simulations reduce student anxiety, creating a safer environment for practicing clinical questioning and diagnosis [9,13].

Integrating ChatGPT into OSCEs aligns with established educational theories. Constructivist Learning Theory suggests that learners build knowledge through active engagement with content [6,7]. ChatGPT supports this by enabling interactive practice of clinical scenarios with AI-guided scaffolding. Experiential Learning Theory posits that knowledge emerges through hands-on experience [10,12], and ChatGPT allows students to engage in simulated clinical encounters, apply theoretical knowledge, and reflect on their performance. Cognitive Load Theory emphasizes minimizing extraneous cognitive load to optimize learning [6,8]; ChatGPT contributes by breaking complex cases into manageable segments and providing real-time clarifications, allowing students to focus on essential aspects of clinical reasoning. This study aims to explore the training of ChatGPT to simulate lifelike patient interactions within OSCEs and to assess its impact on medical education. It is hypothesized that AI integration will enhance student learning experiences and improve the accuracy of clinical competence evaluations. By combining technological advancements with established educational theories, AI-driven OSCE simulations have the potential to revolutionize medical education, making assessments more accessible, flexible, and effective.

Methods

This study employs a quantitative research design using a structured survey administered via "Qualtrics". The study aims to measure specific variables related to the research question, allowing for statistical analysis to identify patterns, correlations, or differences among the study population.

Ethics Approval

The Department Chair of the St. George's University IRB approved the study and confirmed that the study conforms to all applicable guidelines and that all ethical matters were dealt with accordingly. A consent form was provided to all participating faculty in the study, which included information about the project. All consent forms were signed prior to the start of participation.

Criteria for and Methods of Study Selection

Inclusion Criteria:

Methods of Study Selection:

All participants were adults aged 18 years or older and were proficient in English, as the survey was administered in that language. Internet access was required to complete the online survey, and all participants met this criterion. At the time of the study, participants were employed at St. George's University (SGU) and were actively engaged in administering OSCEs to students enrolled in the PCM 501.

Participants

The participants were faculty members currently employed at St. George's University who assess students at OSCEs. A total of 20 participants were recruited. We chose a small sample size in the initial phase (pilot study) due to resource limitations, as we were using one account in ChatGPT to collect the data, which would be time-consuming with a larger cohort.

OSCE Methodology

The Objective Structured Clinical Examination (OSCE) is a widely used method for assessing the clinical competence of healthcare professionals, encompassing history-taking,

physical examination, communication, clinical reasoning, and procedural skills. Typically, history-taking and physical examination are conducted with a standardized patient (SP)an individual trained to consistently portray the characteristics, symptoms, and emotional responses of a real patient. SPs provide a controlled yet realistic environment in which learners can practice and be evaluated on clinical and communication skills. In this study, ChatGPT was employed as an SP. The model was trained via a single account using the following steps (full documentation provided in Appendix 1): collection of diverse medical scenarios, including acute and chronic illnesses and mental health presentations; preprocessing and annotation of responses to indicate scenario type, patient condition, and emotional tone; role-play training with the prompt *"I would like you to behave as a standardized patient"*; continuous validation to ensure alignment with current medical knowledge and best practices; and customization of scenarios for participant use (scripts provided in Appendix 2), including abdominal, musculoskeletal, and psychiatry protocols. Each participant completed all three cases in a single 60-minute session, with each case lasting 17 minutes (15-minute encounter plus 2-minute transition). Following the OSCE, participants completed a survey reflecting on their experience interacting with ChatGPT as a standardized patient.

Materials and Procedures

In July 2024, 20 participants attended the Simulation Lab, located on the 3rd level of St. George's Hall at St. George's University. They performed the OSCE in the same sequence on a computer in room 21 of the lab. Each participant was allotted 15 minutes to interact with ChatGPT for each case, with a five-minute warning provided before the end of the session. Additionally, participants were given a two-minute transition between cases. This process was completed in one day. The survey was disseminated via email, and performance was ungraded.

Measurement Methods and Data Collection Techniques

Survey Instrument:

The survey was created using Qualtrics, a web-based platform that allows for the development of customized surveys with diverse question formats (e.g., multiple-choice, Likert scale, open-ended questions).

Data Collection Techniques:

The survey was distributed via email. Participants completed the survey anonymously to promote honest and unbiased responses. The survey remained accessible until the end of the study.

Data Analysis Procedures

Data Cleaning:

After data collection, responses were downloaded from Qualtrics in an Excel format.

Data was screened for incomplete responses or responses that did not meet the inclusion criteria, and those were removed.

Reporting:

Results are presented in tables for clear visualization of the data.

Results

In this study, participants rated their experience with AI as a standardized patient in various ways (table 1): 35% rated it as excellent, 35% as good, and 30% as average. The most frequently mentioned benefits of AI as a standardized patient were its responsiveness and clarity, with participants appreciating the rapid and clear responses provided (table 2). The AI was also noted for its potential to help develop clinical skills, particularly in direct questioning and clinical reasoning (table 2). Additionally, participants felt more comfortable interacting with AI as a standardized patient compared to real patients, as it reduced feelings of intimidation (table 2).

Regarding challenges, 85% of participants identified issues related to non-verbal communication, such as the inability to utilize non-verbal cues and limitations in performing physical examinations. Participants also highlighted difficulties in communication and rapport-building, citing a lack of empathy and proper clinical reasoning in AI as a standardized patient interaction. Technical and system issues were also

Table 1: Quantitative feedback results

Question	Answer choices and results				
How would you rate your overall experience with the AI standardized patient?	Excellent 35%	Good 35%	Average 30%	Poor 0%	Terrible 0%
Would you prefer to have more AI standardized patient interactions in future OSCEs?	Yes 20%	Maybe 50%	No 30%		
In your opinion, which type of standardized patient contributes more to a student's learning experience in an OSCE setting?	Both equally 20%	Real patient 65%	Artificial intelligence 15%		

noted, including system crashes and inconsistent responses. Participants expressed concerns about the realism and authenticity of AI as a standardized patient, stating that it was less realistic than interacting with a real patient, particularly in terms of communication and interpersonal skills. (table 3)

Despite these challenges, 20% (table 1) of participants indicated interest in incorporating more AI in future OSCEs, acknowledging its potential to enhance clinical skills development, reduce pressure, and provide usefulness in specific scenarios that do not require physical examination. However, they noted that AI as a standardized patient should not be used solely for practical assessments but rather as part of a hybrid model alongside real standardized patient (table 1). This approach would help address the limitations of AI as a standardized patient particularly in developing communication skills and empathy.

Table 2: What aspects of the AI standardized patient experience did you find most beneficial

Themes	Example excerpts
Development of Clinical Skills	Good for direct questioning. Can help to develop clinical reasoning especially information elicited during the history
	The information given was clear once the appropriate questions were asked
Comfort Level	Did not feel intimidated as I would with a patient
Responsiveness and Clarity	Rapid response
	Real time responses
	Quick responses but need to ask correctly to elicit an appropriate answer
	The responses were clear and concise

Table 3: Preference to have more AI standardized patient interactions in future OSCEs

Benefits for Clinical Skills Development	Advancing clinical skills especially clinical reasoning
	Beneficial for extra training, revisiting scenarios, and cold cases in preparation for standardized patient encounter
	AI responses are generated quickly. I can ask another question while the response to the previous question is being generated
Limitations in Practical Assessments	Cannot be used solely for a practical assessment that involves performing a physical examination
	Nearly impossible to address communication skills and empathy using AI
	It will take a student longer to get the information if they must type their questions and read responses versus asking and actively listening
Hybrid System Suggestion	Maybe consider a hybrid system where they would use a simulated mannequin to perform the exam
	The option to have a percentage of AI SP in OSPE would help in comparison and perfecting the tool
Reduced Pressure and Enhanced Focus	There is less pressure as compared with a real patient
	Responses are generated quickly, allowing for follow-up questions
Usefulness in Specific Scenarios	Small cases not requiring physical examination
	Beneficial for extra training and revisiting scenarios

Participants also emphasized the value of real standardized patients in contributing to a student's learning experience. They highlighted the importance of real experiences, observation, and non-verbal communication in clinical reasoning and practical skills development. They also mentioned the limitations of AI, such as its scripted responses and inability to replicate non-verbal cues, while recognizing its potential role in patient management and scenario revision.

In conclusion, while AI as a standardized patient offers several benefits, especially in enhancing clinical reasoning and reducing pressure, its limitations, particularly in non-verbal communication and practical skills assessment, suggest that it is best used as part of a hybrid model alongside real standardized patients to provide a more comprehensive learning experience for students.

Limitations of the study

Students were not included as participants in this study, as the focus was solely on faculty members involved in OSCE development and evaluation. This decision was made to gather expert insights on the use of ChatGPT in clinical simulations before involving students in subsequent stages. Additionally, transcripts of the simulated interactions were not analyzed in this study, though such an analysis could provide valuable insights into communication patterns and the quality of the simulations. Incorporating transcript analysis and student participation in future research could offer a more comprehensive understanding of ChatGPT's effectiveness in enhancing OSCEs.

Discussion

The introduction of practice sessions using ChatGPT for OSCEs enhances accessibility and provides a cost-effective opportunity for students to practice clinical encounters. This integration of technology into medical training aligns with the learning preferences of a new generation of students and offers several advantages:

Standardized Assessment: ChatGPT ensures consistent and unbiased assessment of communication skills and clinical knowledge across students, minimizing examiner bias.

Scalability: The use of ChatGPT allows for the simulation of patient interactions on a large scale, making it an ideal solution for institutions with a high number of medical learners. This allows an institution to manage the resource of standardized patients more effectively.

Flexibility: OSCE scenarios can be easily adapted and customized with ChatGPT to focus on specific learning objectives and competencies, providing tailored learning experiences.

Immediate Feedback: Students benefit from immediate feedback on their performance, which supports continuous improvement in both communication skills and medical knowledge.

Resource Efficiency: The implementation of ChatGPT reduces the reliance on standardized patients, which can be both costly and logistically challenging to manage.

Data-Driven Insights: The data collected from student interactions with ChatGPT can offer valuable insights into areas where students may require additional training, allowing for targeted educational interventions.

Overall, the integration of ChatGPT into OSCEs presents a promising advancement in medical education, offering a scalable, flexible, and resource-efficient approach to enhancing student learning and assessment.

Conclusion

The integration of AI language models, specifically training ChatGPT to act as a patient, represents a promising innovation in the realm of Objective Structured Clinical Examinations (OSCEs) within medical education. By incorporating ChatGPT, which can simulate a wide range of patient scenarios (enriched learning experience) and provide realistic responses, there is potential to enhance both the educational and assessment aspects of OSCEs (enhanced training effectiveness). This approach aims to improve the accuracy of skill assessments, offer diverse patient interactions, and ultimately better prepare medical students for real-world clinical settings.

References

1. Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE) (1979).

2. Ramchandani R, Biglou S, Gupta M, et al. Using AI to revolutionize clinical training through OSCE-GPT: a focused exploration of user feedback on otolaryngology and neurology cases (2024).

3. Hopkins L, Hampton BS, et al. To the point: medical education, technology and the millennial learner (2017).

4. Choi B, Jegatheeswaran L, Minocha A, et al. The impact of the COVID-19 pandemic on final year medical students in the United Kingdom: a national survey (2020).

5. Twenge JM. Generational changes and their impact in the classroom: teaching Generation Me (2009).

6. Khalil MK, Elkhider IA. Applying learning theories and instructional design models for effective instruction (2016).

7. Jones O, Saunders H, Mires G. The E-learning revolution in obstetrics and gynaecology (2010).

8. Sweller J. Cognitive load during problem solving: effects on learning (1988).

9. Gandhi MH, Mukherji P. Learning Theories (2024).

10. Mitchell SA, Boyer TJ. Deliberate Practice in Medical Simulation (2024).

11. Kolb DA. Experiential learning: experience as the source of learning and development (1984).

12. Shen W, Liang X, Xiang X. Using AI as standardized patients in pediatric surgeon training program: a tentative exploration (2024).

13. Maicher KR, Zimmerman L, Wilcox B, Liston B, et al. Using virtual standardized patients to accurately assess information gathering skills in medical students (2019).