

HARNESSING CHATGPT FOR ASSESSMENT IN MEDICAL EDUCATION

¹DR. SULTHAN AL RASHID, ²DR. ANBARASU KANCHANA MALA,
³DR.B.SARAVANAKUMAR

¹ASSISTANT PROFESSOR, DEPARTMENT OF PHARMACOLOGY, SAVEETHA MEDICAL COLLEGE AND HOSPITAL, SIMATS, CHENNAI, INDIA

²POSTGRADUATE, DEPARTMENT OF PHARMACOLOGY, , SAVEETHA MEDICAL COLLEGE AND HOSPITAL, SIMATS, CHENNAI, TAMIL NADU, INDIA.

³PROFESSOR, DEPARTMENT OF ORAL& MAXILLOFACIAL SURGERY , SREE BALAJI DENTAL COLLEGE & HOSPITAL, CHENNAI, INDIA

INTRODUCTION

Medical education is undergoing a major shift. Traditional methods that emphasized memorization and high-stakes final exams are being replaced by competency-based medical education (CBME), which focuses on the holistic development of knowledge, skills, attitudes, and values. At the heart of CBME is assessment—not merely as a tool for grading, but as a dynamic, ongoing process that encourages, guides, and evaluates meaningful learning. In this evolving framework, assessments are designed to be authentic, continuous, and aligned with real-world clinical competencies.

Despite progress in medical education, educators still encounter numerous hurdles in applying effective assessment strategies. Challenges include developing well-constructed questions, ensuring fairness and consistency, providing timely feedback, and customizing assessments to meet the needs of diverse learners. These tasks are often labor-intensive and require advanced teaching expertise, especially when dealing with large student cohorts and an expanding curriculum. Additionally, the post-pandemic era has heightened the demand for flexible, digital, and technology-enabled tools to support both educators and students.

Artificial Intelligence (AI), particularly language-based tools like ChatGPT by OpenAI, is increasingly being recognized as a useful aid in education (Ahmed, 2023; Masters et al., 2025). Trained on vast amounts of text, ChatGPT can understand prompts, generate human-like responses, and assist with various academic tasks. While it is widely used for tasks such as content development and brainstorming, its potential in the areas of assessment design, execution, and feedback delivery in medical education is only beginning to be explored.

From my own experience as a pharmacology educator, I have found ChatGPT to be an effective, creative, and surprisingly insightful tool. It has been helpful not only in generating complex, integrated essay questions but also in simulating viva voce scenarios, developing OSCE structures, and offering feedback on reflective writing. In this personal review, I intend to present a reflective and optimistic view of ChatGPT's role in medical education assessments—highlighting its potential, sharing practical examples, discussing current limitations, and imagining its future applications.

The Changing Landscape of Assessment in Medical Education

Historically, medical education assessments have primarily used standard formats like multiple-choice questions (MCQs), structured short answer questions (SAQs), long essays, viva voce exams, and clinical evaluations such as OSCEs and ward-based assessments. While these approaches have long been the backbone of evaluation, they often fall short in assessing key real-world competencies like clinical reasoning, communication skills, professionalism, and a dedication to continuous learning.

Moreover, traditional assessment has largely focused on summative evaluations, typically conducted at the end of a course or term to decide pass or fail results. These high-stakes exams can increase student stress, promote surface-level learning, and often fail to accurately reflect a learner's actual capabilities—especially in handling complex clinical situations.

The transition to CBME, advocated by the National Medical Commission (NMC) of India and embraced worldwide, places a strong emphasis on competency-based, formative, and programmatic approaches to assessment. This transformation requires assessments that:

- Are continuous and aligned with learning milestones
- Encourage reflection and self-directed learning

- Integrate knowledge across disciplines (horizontal and vertical)
- Are authentic and close to real clinical contexts
- Provide individualized, timely feedback to promote growth

Bringing this vision to life presents significant challenges. Faculty members frequently struggle with limited time, insufficient training, and a lack of resources, which makes it difficult to create diverse and personalized assessments for large groups of students. Maintaining consistency and high quality adds to the difficulty, as activities such as evaluating reflective journals, assessing OSCEs, or developing integrated questions are often time-consuming and prone to variability.

At the same time, learners are evolving. Today's students, as digital natives, are accustomed to interacting with smart technologies. They expect immediate assistance, personalized feedback, and engaging, context-specific assessments. Meeting these expectations through traditional methods is becoming more and more challenging.

AI tools such as ChatGPT hold the promise to revolutionize assessment practices. Its ability to rapidly generate content, mimic authentic interactions, and provide context-aware feedback makes it a strong fit for the evolving needs of competency-based medical education (CBME). Whether it's automating preliminary grading, collaborating on question creation, simulating clinical situations, or designing diagnostic quizzes, ChatGPT can support every stage of the assessment process—from design and delivery to feedback.

The current challenge is no longer about whether AI can support us, but about how to integrate it responsibly—striking the right balance between innovation and ethics, efficiency and precision, and technological advancement and human judgment. This review seeks to examine these critical aspects in depth.

ChatGPT in Question Generation

Developing high-quality assessment questions is among the most time-intensive tasks for educators. It requires careful alignment with learning objectives, adherence to Bloom's taxonomy, and the inclusion of clinically relevant content—each requiring meticulous planning and peer review. In this regard, ChatGPT can serve as a helpful and time-saving tool to simplify and support the question creation process (Indran et al., 2024).

In my personal teaching experience, I have used ChatGPT to develop integrated pharmacology questions, such as connecting pharmacokinetics with clinical case scenarios. By using a simple prompt like, "Create a clinical case-based essay question on beta-blockers that includes horizontal integration with physiology and vertical integration with cardiology," the tool generated well-organized questions that aligned with our assessment standards. With additional prompts, it also suggested appropriate distractors for MCQs and provided explanations for the correct answers.

This AI-driven assistance reduces the cognitive load on faculty and encourages creative question design. More importantly, it allows educators to quickly produce multiple variations of the same question for use in formative assessments and practice tests, thereby helping to reduce academic dishonesty and promote a more equitable evaluation process.

Creating varied and Bloom's taxonomy-aligned assessment questions can be time-consuming. ChatGPT can generate clinical, integrated, and analytical questions based on prompts provided by faculty.

Example: MCQ Generation (Basic Knowledge)

Prompt to ChatGPT:

"Create a single-best answer MCQ on beta-blockers with options and correct answer highlighted."

Response:

Question: A 55-year-old male with a history of myocardial infarction is prescribed a beta-blocker. Which of the following beta-blockers is cardioselective?

- A. Propranolol
- B. Carvedilol
- C. Atenolol ☒
- D. Labetalol
- E. Nadolol

This saved significant time in creating question banks and allowed easy alignment with learning objectives.

Example: Essay Question with Integration

Prompt:

"Frame an integrated essay question on thiazide diuretics incorporating horizontal integration with renal physiology and vertical integration with hypertension management."

Response:

Essay Question:

Discuss the mechanism of action, pharmacological properties, and adverse effects of thiazide diuretics. Correlate their role with renal physiology, particularly sodium handling in the distal tubule, and explain how they are used in the management of hypertension according to national treatment guidelines.

This aligned perfectly with both preclinical and clinical curricula and allowed testing of higher-order thinking skills.

Enabling Formative Feedback

Effective formative assessment depends on prompt, targeted, and constructive feedback. ChatGPT can support this process by offering immediate feedback on short responses, reflective pieces, or progress assessments. For instance, I used ChatGPT to evaluate a student's brief reflection on antibiotic stewardship. It pointed out flaws in reasoning, recommended areas for expansion, and acknowledged strengths—tasks that would typically require considerable faculty time and effort.

Students gain significantly from this interaction as well. ChatGPT provides a non-judgmental environment where they can explore their thoughts, get constructive input, and refine their answers before submission. For those who are shy or face academic challenges, it serves as a supportive resource that boosts their confidence and encourages independent learning.

Nevertheless, human supervision remains crucial. Although ChatGPT can identify conceptual mistakes or flaws in logic, it may overlook nuanced context or emotional depth—elements that are particularly important in assessments involving reflection or ethical reasoning.

ChatGPT's ability to review text and offer suggestions makes it valuable for formative feedback on student responses or reflective writing (Krumsvik, 2025).

Example: Reflective Writing Evaluation

Student Input:

"I saw a patient with asthma today. I learned about the importance of using inhalers properly and how empathy matters when explaining this to the patient."

Prompt:

"Evaluate this reflection using Gibbs' Reflective Cycle and suggest improvements."

ChatGPT Response:

- **Description:** Adequate
- **Feelings:** Not clearly expressed
- **Evaluation:** Could elaborate on challenges
- **Analysis:** Needs depth
- **Conclusion:** Missing
- **Action Plan:** Could benefit from identifying specific future steps

Suggested improvement: Add emotional response (e.g., "I felt unsure when explaining"), discuss the patient's reaction, and state how the experience will change future behavior.

This helps students develop structured reflection skills with immediate guidance, even outside classroom hours.

Supporting Higher-Order Learning

Assessment should go beyond simple recall and aim to promote critical thinking, synthesis, and practical application. When used thoughtfully, ChatGPT can facilitate this type of learning through Socratic-style questioning. For example, a student might present a clinical case and ask ChatGPT to quiz them on differential diagnoses, diagnostic steps, and treatment options. This interactive dialogue encourages deeper reasoning and helps identify areas of weakness.

Additionally, in problem-based learning (PBL) and case-based discussions (CBD), ChatGPT can serve as a co-facilitator. In our pharmacology sessions, students used the tool to generate potential questions based on clinical vignettes and then reviewed them with their peers. This collaborative approach between students and AI made the learning process more interactive and engaging, while also aligning with CBME's focus on active, student-centered learning.

OSCE and Simulation Support

Objective Structured Clinical Examinations (OSCEs) continue to be the benchmark for evaluating clinical competencies. Although ChatGPT cannot replicate a physical patient, it can assist in developing OSCE stations (Misra & Suresh, 2024). I have personally used it to create standardized patient (SP) scripts, checklists, and examiner instructions with minimal prompting. For example, a simple request like, "Create a 5-minute OSCE station on counseling a diabetic patient about insulin therapy, including a checklist and patient prompts," yielded practical drafts that could be easily adapted to suit local requirements.

Likewise, in virtual simulations, ChatGPT can take on the role of a virtual patient or caregiver, responding to student queries and encouraging deeper exploration. This enhances realism and supports reflective learning, particularly when in-person simulations are constrained by budget or logistical challenges. It can also be used to generate initial drafts of clinical scenarios, checklists, and patient instructions for simulation-based learning.

Example: OSCE Station Blueprint

Prompt:

"Design a 5-minute OSCE station for counseling a newly diagnosed diabetic patient on insulin use. Include checklist and patient instructions."

Response Summary:

- **Scenario:** Patient is worried about starting insulin.
- **Checklist for examiner:**
 1. Greets patient
 2. Explains need for insulin
 3. Demonstrates injection technique
 4. Addresses concerns about pain and stigma
 5. Asks for patient understanding and consent
- **Patient Instructions:** Act anxious and reluctant. Ask about lifelong need for insulin and alternatives.

This station was directly usable with minor modifications and helped in rapid OSCE preparation during internal assessments.

Self-generated Quiz Practice

A student studying for pharmacology asked:

"Quiz me on antiepileptic drugs with case-based MCQs."

ChatGPT provided:

Case Question:

A 24-year-old woman with epilepsy becomes pregnant. Which antiepileptic drug should be avoided due to teratogenicity?

- A. Lamotrigine
- B. Valproic Acid ☒
- C. Carbamazepine
- D. Levetiracetam

It then gave rationale for each option, turning it into an active learning experience. Students shared that such AI interactions improved their confidence and preparation.

While not replacing human judgment, ChatGPT can assist in providing a first-pass evaluation or content-based analysis of essays.

Pharmacology Long Answer Evaluation

Input:

"Evaluate the following essay on ACE inhibitors for content accuracy and coherence."

ChatGPT Response Summary:

- Clear explanation of mechanism ✓
- Missed mention of cough as a common adverse effect ✗
- Good integration with RAAS system ✓
- Repetition noted in clinical uses section ✗
- Coherence: Moderate, needs better transitions

This can be used for initial drafts, self-editing by students, and training junior faculty in assessment standards.

ChatGPT in Viva/Oral Exam Preparation

Oral exams require spontaneous thinking. Students can simulate vivas using ChatGPT.

Example: Pharmacology Viva Simulation

Prompt:

"Pretend you are a pharmacology examiner. Ask me questions on aminoglycosides and give feedback based on my answers."

ChatGPT asked:

1. What is the mechanism of action of aminoglycosides?
2. Name two major side effects and explain their mechanism.
3. Why are aminoglycosides given with beta-lactams?

This allowed students to rehearse under pressure-like conditions, improving verbal articulation and clinical reasoning.

Benefits for Educators

From a faculty perspective, the integration of ChatGPT brings several benefits:

- Time-saving: It accelerates question creation, OSCE station design, and feedback delivery.
- Consistency: Reduces subjectivity and maintains uniform quality in questions and evaluations.

- Capacity-building: Enables faculty to focus on mentorship and deeper analysis rather than routine tasks.
- Professional Development: Encourages educators to upskill in prompt engineering, AI ethics, and blended learning.

However, institutional support and training are essential to realize these benefits. Educators must be empowered to explore AI tools while being guided on their ethical and pedagogical use.

Challenges and Ethical Concerns

Despite its promise, using ChatGPT in assessment also raises several concerns:

1. Accuracy and Hallucination: ChatGPT may occasionally provide factually incorrect or misleading content. This is dangerous in medical education, where precision is crucial.
2. Bias: The model may reflect inherent biases from its training data. Without careful review, such bias could propagate into assessments or feedback.
3. Over-reliance by Students: There is a risk that students may use ChatGPT to generate answers rather than develop their own reasoning, thus undermining learning.
4. Data Privacy: Using student data or uploading confidential content into AI platforms must comply with privacy regulations.
5. Assessment Security: In high-stakes exams, unregulated access to ChatGPT can enable academic dishonesty.

Hence, the integration of ChatGPT must be accompanied by clear guidelines, supervision, and ethical safeguards. AI should augment, not replace, human judgment.

Future Directions

To harness the full potential of ChatGPT in assessment, medical schools should:

- Develop AI-assisted Assessment Policies: Clearly define acceptable uses of AI for students and faculty.
- Train Faculty and Students in Prompt Crafting: The quality of output depends on the quality of input.
- Embed AI in LMS Platforms: Seamless integration with existing tools will enhance usability and oversight.
- Conduct Research: Evidence-based studies are needed to validate ChatGPT's impact on learning outcomes, assessment quality, and faculty workload.
- Promote Ethical Use: Through workshops, case studies, and simulations, cultivate a culture of responsible AI use.

CONCLUSION

Integrating ChatGPT into assessment practices in medical education goes beyond mere convenience—it signifies a transformative change in how we design, interpret, and implement evaluations of learning. Through my own experience as an educator, I have observed how this AI-driven tool significantly improves the efficiency, innovation, and flexibility of assessments, ranging from MCQs to integrated essays and reflective responses.

Fundamentally, ChatGPT acts as a digital partner, supporting educators by generating a variety of question formats, simulating clinical cases, developing OSCE checklists, and providing personalized formative feedback. These functions help bridge persistent gaps in assessment, especially in large medical cohorts where individualized support and prompt feedback are often limited by resource challenges.

More significantly, when used appropriately, ChatGPT fosters active learning and critical thinking. It enables students to participate in AI-supported self-assessment, obtain immediate clarifications, and practice clinical reasoning through simulated dialogues—shifting assessment from a passive, stress-inducing experience to an engaging, student-centered learning process.

Nonetheless, adopting ChatGPT for assessment comes with important considerations. It demands well-planned faculty training, strong institutional guidelines, and a firm commitment to ethical practices and content accuracy. Caution must be taken to prevent over-dependence on AI, uncritical use, or substituting human feedback with automated responses. Educators should remain at the helm of the teaching process, using AI as a supportive aid to enrich—not substitute—their professional judgment and expertise.

Furthermore, the utility of ChatGPT must be backed by systematic research. There is a pressing need for evidence-based studies to assess the reliability, accuracy, and educational impact of AI in assessments. Medical schools should encourage pilot projects, cross-disciplinary collaborations, and workshops that explore AI's full potential in a controlled and reflective manner.

In the future, as AI continues to evolve, tools like ChatGPT could be integrated directly into learning management systems (LMS), allowing for adaptive assessments, personalized feedback loops, and real-time analytics to monitor student progress. The promise is enormous—but realizing it depends on thoughtful, intentional, and ethically sound implementation.

In conclusion, ChatGPT presents a valuable opportunity to transform assessment in medical education. It enhances efficiency for educators, motivates students to actively engage in their own learning, and aligns seamlessly with the CBME approach of ongoing, competency-driven evaluation. As we navigate the evolving landscape of technology and education, it is essential to embrace such tools thoughtfully and strategically—shaping assessments that go beyond measuring learning to actively fostering it.

REFERENCES

- Ahmed Y. (2023). Utilization of ChatGPT in Medical Education: Applications and Implications for Curriculum Enhancement. *Acta informatica medica :AIM : journal of the Society for Medical Informatics of Bosnia & Herzegovina :casopisDrustva za medicinskuinformatiku BiH*, 31(4), 300–305. <https://doi.org/10.5455/aim.2023.31.300-305>
- Indran IR, Paranthaman P, Gupta N, Mustafa N. Twelve tips to leverage AI for efficient and effective medical question generation: A guide for educators using Chat GPT. *Med Teach*. 2024 Aug;46(8):1021-1026. doi: 10.1080/0142159X.2023.2294703. Epub 2023 Dec 26. PMID: 38146711.
- Krumsvik R. J. (2025). GPT-4's capabilities for formative and summative assessments in Norwegian medicine exams-an intrinsic case study in the early phase of intervention. *Frontiers in medicine*, 12, 1441747. <https://doi.org/10.3389/fmed.2025.1441747>
- Masters, K., MacNeil, H., Benjamin, J., Carver, T., Nemethy, K., Valanci-Aroesty, S., Taylor, D. C. M., Thoma, B., &Thesen, T. (2025). Artificial Intelligence in Health Professions Education assessment: AMEE Guide No. 178. *Medical teacher*, 1–15. Advance online publication. <https://doi.org/10.1080/0142159X.2024.2445037>
- Misra, S. M., & Suresh, S. (2024). Artificial Intelligence and Objective Structured Clinical Examinations: Using ChatGPT to Revolutionize Clinical Skills Assessment in Medical Education. *Journal of medical education and curricular development*, 11, 23821205241263475. <https://doi.org/10.1177/23821205241263475>