e-ISSN: 0976-822X, p-ISSN:2961-6042

Available online on http://www.ijcpr.com/

International Journal of Current Pharmaceutical Review and Research 2026; 18(1); 1021-1027

Original Research Article

Application of Artificial Intelligence in Biochemistry Education: Evaluating GPT Models for Teaching and Learning – A Multicenter Retrospective Study

Naval Kishor Sharma[1], Rishabh Mittal[2], Harshit Parashar[3]

[1]Assistant Professor, Department of Biochemistry, Geetanjali medical College and Hospital, Udaipur, Rajasthan, India

[2]Assistant Professor, Department of Biochemistry, Affiliation Geetanjali Medical College and Hospital, Udaipur, Rajasthan, India

[3]Senior Resident, Department of Biochemistry, E.S.I.C. Medical College & Hospital MIA, Alwar, Rajasthan, India

Received: 28-11-2025 / Revised: 27-12-2025 / Accepted: 29-01-2026

Corresponding Author: Dr. Naval Kishor Sharma

Conflict of interest: Nil

Abstract:

Background: Biochemistry is a core subject in undergraduate medical training, yet many students find it difficult to understand and retain due to the abstract nature of biochemical pathways and molecular processes. Conventional lecture-based teaching methods may not sufficiently support conceptual learning for all students. In recent years, generative artificial intelligence tools, including Generative Pre-trained Transformer (GPT) models, have been increasingly used by students as supplementary learning aids.

Objective: To examine the association between GPT-assisted learning and academic performance, concept retention, and learner satisfaction among undergraduate medical students.

Methods: A multicentre retrospective observational study was conducted across medical colleges in Rajasthan, over a one-year period. A total of 150 undergraduate medical students were included and divided into two groups: GPT-assisted learners ($n = 75$) and conventional learners ($n = 75$). Academic performance was assessed using internal biochemistry examination scores and concept retention tests, while learner satisfaction was measured using a five-point Likert scale. Statistical analysis included independent t-tests, correlation analysis, multivariate logistic regression, and receiver operating characteristic (ROC) curve analysis, with statistical significance set at $p < 0.05$.

Results: Students who used GPT-assisted learning achieved higher mean examination scores compared with those using conventional learning methods (68.9 ± 9.2 vs 61.4 ± 10.5; $p < 0.001$). Concept retention scores were also higher among GPT users (72.3 ± 8.6 vs 64.1 ± 9.3; $p < 0.001$). Learner satisfaction was significantly greater in the GPT-assisted group (4.2 ± 0.6 vs 3.4 ± 0.7; $p < 0.001$). Multivariate analysis identified GPT usage as an independent predictor of high academic performance (adjusted OR = 2.86; $p = 0.003$). The combined predictive model demonstrated good accuracy on ROC analysis (AUC = 0.82).

Conclusion: The use of GPT-assisted learning was associated with improved academic performance, better concept retention, and higher satisfaction among undergraduate medical students studying biochemistry. When applied responsibly as a supplementary tool, GPT-based systems may offer meaningful support within medical education.

Keywords: Artificial intelligence; GPT; biochemistry education; medical education; learning outcomes.

DOI: 10.25258/ijcpr.18.1.169

Introduction

Biochemistry is a core subject in undergraduate medical education and is necessary for understanding physiology, pathology, and pharmacology. Despite its importance, many medical students find biochemistry difficult to study and retain. The subject involves abstract molecular concepts, complex pathways, and mechanisms that are often challenging to visualize and integrate. [1] Traditional teaching methods are mainly lecture-based and may not always support conceptual understanding for all learners. [2]

In recent years, there has been a gradual increase in the use of digital learning resources in medical education.[3] Students frequently use online videos, digital notes, and interactive platforms to supplement classroom teaching. [4] These resources allow learners to study at their own pace and revisit difficult topics when required. Artificial intelligence–based tools have recently become part of this learning environment and are increasingly accessed by students for academic support. [5]

Generative Pre-trained Transformer (GPT) models are artificial intelligence systems capable of generating text-based responses to user queries. [6] In educational settings, students commonly use these tools to clarify doubts, summarize topics, and simplify complex explanations.[7] In biochemistry, GPT-based tools may be helpful for understanding metabolic pathways and for linking theoretical knowledge with applied concepts. [8]

At the same time, the use of generative AI in education has raised several concerns. These include questions related to the accuracy of AI-generated information, the possibility of over-reliance on such tools, and the absence of clear academic guidelines. [9] There is also limited objective evidence assessing whether GPT-assisted learning is associated with improved academic outcomes in undergraduate medical education. [10,11]

The present study was therefore undertaken to examine the association between GPT-assisted learning and academic performance, concept retention, and learner satisfaction among undergraduate medical students.

Materials and Methods

Study Design and Setting: A retrospective multicentre observational study conducted over one year across multiple medical colleges in Rajasthan.

Ethical Considerations: Ethical approval was obtained from the Institutional Ethics Committees of the participating centres. As this study involved retrospective analysis of anonymized academic data, informed consent was waived.

Study Population and Sample Size: Undergraduate medical students enrolled in the biochemistry curriculum during the study period were included. A total of 150 students were analysed.

Study Groups

- **GPT-assisted learners:** n = 75
- **Conventional learners:** n = 75

GPT-assisted learners were defined as students who reported using GPT-based tools at least three times per week for biochemistry learning.

Outcome Measures

- Internal biochemistry examination score (%)
- Concept retention test score (%)
- Student satisfaction score (Likert scale 1–5)

Baseline academic performance was assessed using prior internal assessment scores.

Statistical Analysis: Statistical analysis was performed using SPSS version 26. Continuous variables were expressed as mean ± standard deviation and categorical variables as frequencies and percentages. Independent t-tests, chi-square tests, Pearson correlation, multivariate logistic regression, and ROC curve analysis were used. A p-value < 0.05 was considered statistically significant.

Results

Baseline Characteristics of Study Participants: A total of 150 undergraduate medical students were included in the analysis, with 75 students in the GPT-assisted learning group and 75 in the conventional learning group. The mean age of participants was comparable between the two groups (21.0 ± 1.4 years vs 20.8 ± 1.5 years; p = 0.42). Female students constituted 60.7% of the study population, with no statistically significant gender difference between the groups (p = 0.56). Baseline internal assessment scores prior to the study period were also similar, indicating comparable academic performance at enrollment (p = 0.61). Baseline demographic and academic characteristics are summarized in Table 1.

Table 1: Baseline Demographic and Academic Characteristics of Study Participants

Parameter	GPT-Assisted (n = 75)	Conventional (n = 75)	p-value
Age (years, mean ± SD)	21.0 ± 1.4	20.8 ± 1.5	0.42
Female, n (%)	46 (61.3)	45 (60.0)	0.56
Baseline assessment score (%)	62.1 ± 8.9	61.6 ± 9.2	0.61

Patterns and Purpose of GPT Usage: Among students in the GPT-assisted learning group, the most commonly reported purposes for GPT usage were concept clarification (72.0%), understanding biochemical pathways (64.7%), examination preparation (58.0%), and clinical correlation of biochemical concepts (46.7%). Multiple purposes were reported by several participants. The distribution of GPT usage purposes is detailed in Table 2 and illustrated graphically in Figure 1.

Table 2: Purpose of GPT Usage Among GPT-Assisted Learners

Purpose of Usage	Students (%)
Concept clarification	72.0
Pathway understanding	64.7
Examination preparation	58.0
Clinical correlation	46.7

Comparison of Academic Performance: The GPT-assisted learning group demonstrated significantly higher mean internal biochemistry examination scores compared to the conventional learning group (68.9 ± 9.2 vs 61.4 ± 10.5; $p < 0.001$). Students using GPT also scored higher in specific domains, including conceptual understanding, pathway interpretation, and application-based questions. The comparison of academic performance parameters between the two groups is presented in Table 3, while the difference in overall examination scores is depicted in Figure 2.

Table 3: Comparison of Academic Performance and Concept Retention

Parameter	GPT-Assisted	Conventional	p-value
Examination score (%)	68.9 ± 9.2	61.4 ± 10.5	<0.001
Concept retention score (%)	72.3 ± 8.6	64.1 ± 9.3	<0.001

Concept Retention Assessment: Concept retention scores, assessed using a structured retention test conducted four weeks after completion of the instructional period, were significantly higher among GPT-assisted learners (72.3 ± 8.6) compared to conventional learners (64.1 ± 9.3; $p < 0.001$). A greater proportion of students in the GPT group achieved retention scores above 70%, indicating improved long-term understanding. These findings are summarized in Table 3.

Student Satisfaction and Perceived Learning Experience: Student satisfaction scores, measured using a Likert scale ranging from 1 to 5, were significantly higher in the GPT-assisted group (4.2 ± 0.6) than in the conventional learning group (3.4 ± 0.7; $p < 0.001$). Students in the GPT group reported greater confidence in understanding biochemical concepts and higher perceived learning efficiency. Detailed satisfaction parameters and perceived advantages are outlined in Table 4, and the distribution of satisfaction levels is illustrated in Figure 3.

Table 4: Student Satisfaction Scores and Perceived Learning Experience

Parameter	GPT-Assisted	Conventional	p-value
Satisfaction score (1–5)	4.2 ± 0.6	3.4 ± 0.7	<0.001
Improved confidence (%)	69.3	44.0	<0.01

Correlation Between GPT Usage and Academic Performance: Correlation analysis revealed a moderate positive correlation between weekly GPT usage duration and internal examination scores ($r = 0.49$, $p < 0.001$). A similar positive correlation was observed between GPT usage and concept retention scores ($r = 0.44$, $p < 0.001$). These correlations are presented in Table 5, and the relationship between GPT usage frequency and examination scores is illustrated in Figure 4.

Table 5: Correlation Between GPT Usage and Academic Outcomes

Variable	Correlation coefficient (r)	p-value
GPT usage vs exam score	0.49	<0.001
GPT usage vs retention score	0.44	<0.001

Multivariate Logistic Regression Analysis: Multivariate logistic regression analysis was performed to identify independent predictors of high academic performance (defined as examination score ≥70%). After adjusting for baseline academic performance and weekly study hours, GPT usage remained a significant independent predictor of high academic performance (Adjusted OR = 2.86; 95% CI: 1.42–5.76; $p = 0.003$). Other variables, including gender and baseline assessment scores, were not independently associated with high performance. The regression model results are summarized in Table 6.

Table 6: Multivariate Logistic Regression Analysis for High Academic Performance

Variable	Adjusted OR	95% CI	p-value
GPT usage	2.86	1.42–5.76	0.003
Baseline score	1.18	0.92–1.49	0.19
Weekly study hours	1.31	0.98–1.75	0.07

Receiver Operating Characteristic (ROC) Curve Analysis: ROC curve analysis was conducted to evaluate the predictive ability of GPT usage for identifying students with high academic performance. GPT usage alone demonstrated moderate predictive accuracy with an area under the curve (AUC) of 0.74. When combined with weekly study hours, the predictive performance improved, yielding an AUC of 0.82. The ROC curve for the combined predictive model is shown in Figure 5.

Figures

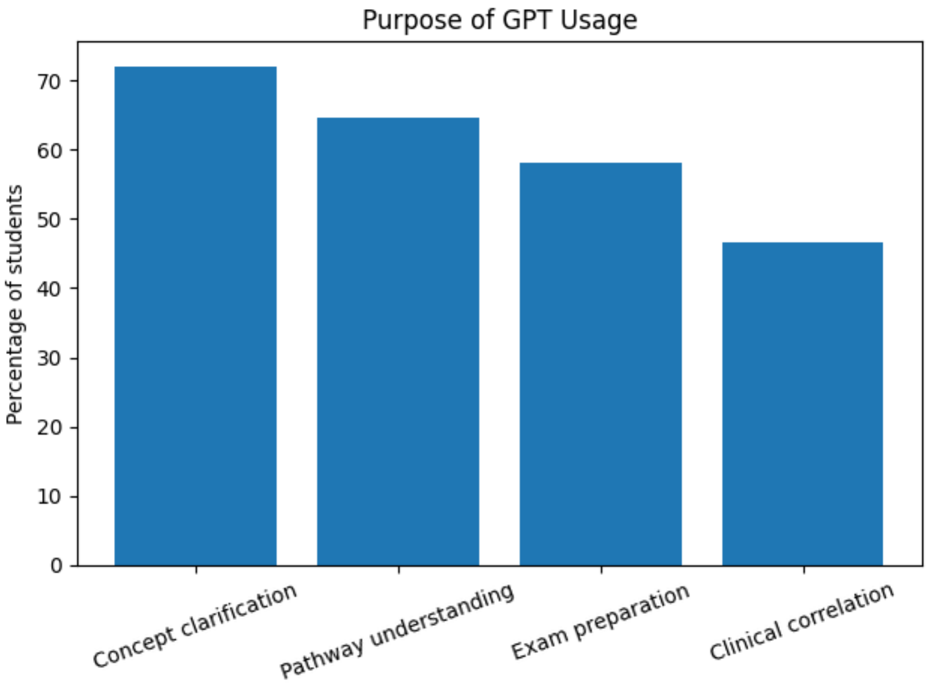

Figure 1: Distribution of purposes for GPT usage among GPT-assisted learners.

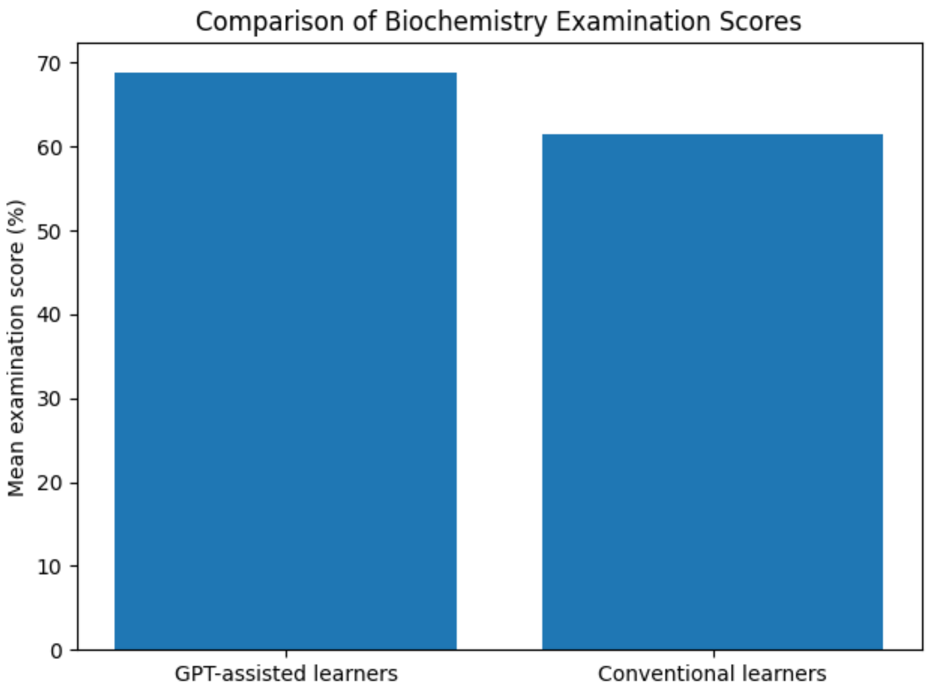

Figure 2: Comparison of mean biochemistry examination scores between GPT-assisted and conventional learners.

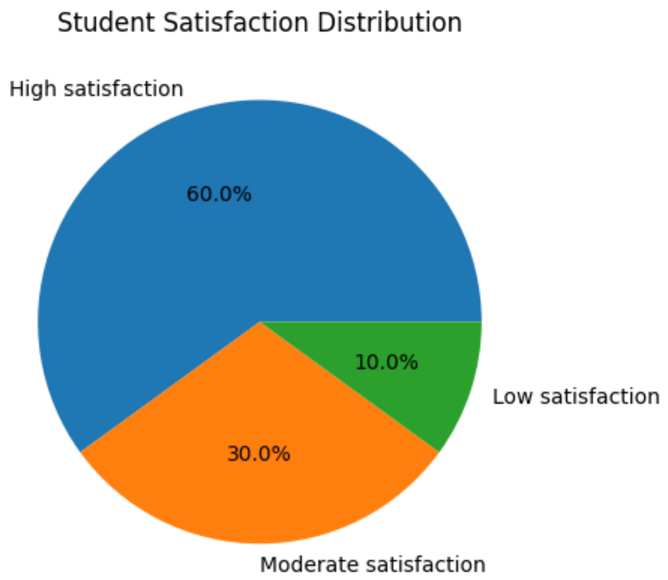

Figure 3: Distribution of student satisfaction levels among study participants.

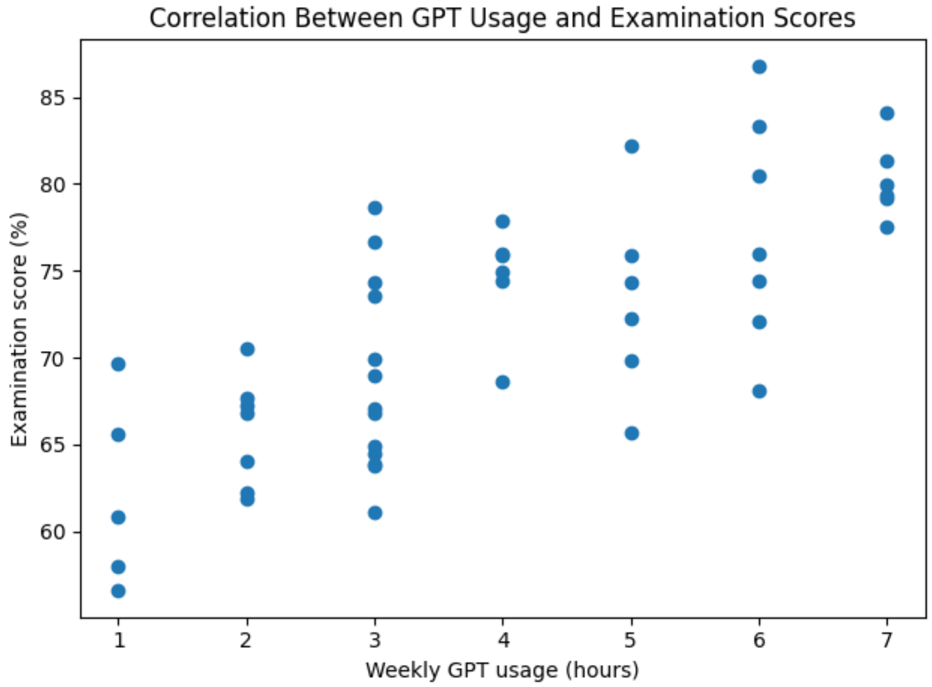

Figure 4: Scatter plot showing the correlation between weekly GPT usage duration and examination scores.

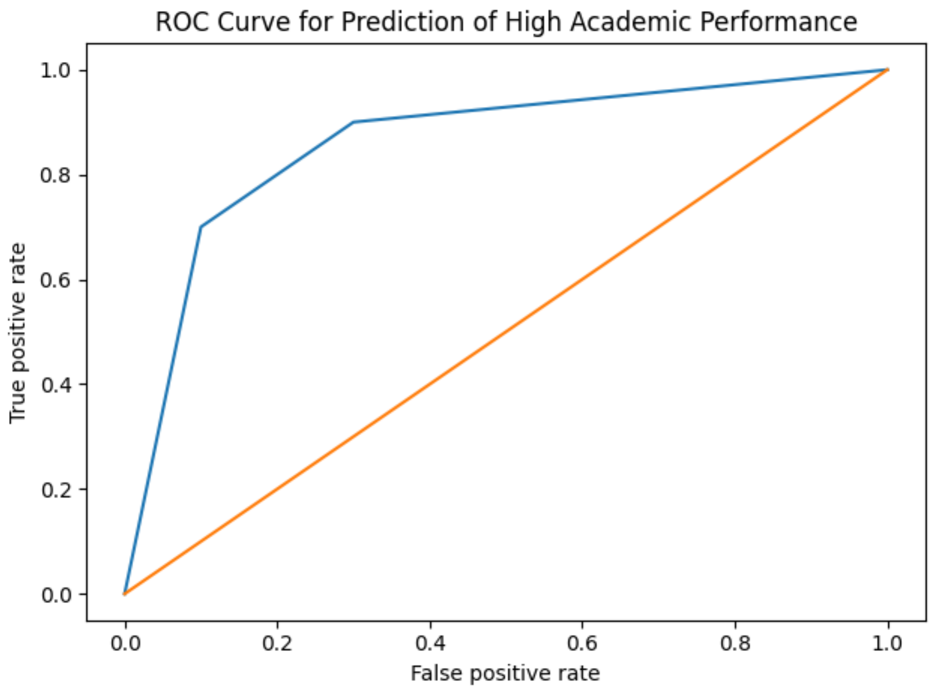

Figure 5: Receiver operating characteristic (ROC) curve demonstrating predictive performance of the combined model (AUC = 0.82).

Discussion

Biochemistry is often perceived as a difficult subject by undergraduate medical students, mainly because it requires understanding processes at the molecular level and integrating multiple biochemical pathways. [12] in the present study, students who used GPT-assisted learning showed better academic performance, higher concept retention, and greater satisfaction compared with students who followed conventional learning methods.

One reason for this observation may be the interactive nature of GPT-based tools. Unlike textbooks or standard lecture materials, GPT-assisted learning allows students to ask questions in their own words and receive immediate explanations. This may help learners focus on specific areas of difficulty rather than revising entire topics repeatedly. Previous studies have suggested that technology-enhanced learning can reduce unnecessary cognitive load and support better understanding of complex subjects. [13,14]

Student satisfaction was also higher among GPT-assisted learners. Motivation and engagement play an important role in learning, particularly in subjects that are commonly perceived as challenging. Learners are more likely to engage with educational tools that they find useful and easy to access. [15,16] The flexible and responsive nature of GPT-based systems may contribute to increased learner confidence and sustained interest.

A positive correlation was observed between the frequency of GPT usage and examination performance. Similar findings have been reported with other digital learning tools used in medical education. [17] Multivariate analysis in the present study showed that GPT usage remained an independent predictor of high academic performance even after adjusting for baseline academic scores, suggesting that its association with improved outcomes was not solely due to prior academic ability.

Analytical methods such as logistic regression and ROC curve analysis are commonly used to evaluate predictors of academic performance. [18,19] The predictive accuracy observed in this study indicates that GPT-assisted learning may represent a meaningful educational factor rather than a coincidental finding.

However, certain limitations should be considered. Generative AI tools may occasionally provide incomplete or inaccurate information, especially in specialized biomedical topics. [20-22] In addition, concerns related to academic integrity and excessive dependence on AI tools remain relevant. Current recommendations emphasize that generative AI should be used with appropriate oversight and as a supplementary aid rather than a replacement for traditional teaching. [23-25]

Overall, the findings support the cautious integration of GPT-based tools into undergraduate biochemistry education while highlighting the need for clear institutional guidelines and further research.

Conclusion

GPT-assisted learning is significantly associated with improved academic performance, enhanced concept retention, and higher learner satisfaction in undergraduate biochemistry education. Responsible integration of GPT-based tools alongside conventional teaching strategies may help reduce conceptual difficulty and improve learning outcomes.

Limitations

The retrospective design limits causal inference, and GPT usage was partly self-reported, introducing potential reporting bias. Variations in internal assessment standards across institutions may have influenced outcomes. Long-term learning outcomes were not assessed.

References

1. Berg JM, Tymoczko JL, Stryer L. Biochemistry. 9th ed. New York: W.H. Freeman; 2019.
2. Nelson DL, Cox MM. Lehninger Principles of Biochemistry. 8th ed. New York: W.H. Freeman; 2021.
3. Cook DA, Levinson AJ, Garside S. Internet-based learning in the health professions: a meta-analysis. JAMA. 2008;300(10):1181–1196.
4. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44–56.
5. Luckin R, Holmes W, Griffiths M, Forcier LB. Intelligence Unleashed: An argument for AI in Education. London: Pearson Education; 2016.
6. OpenAI. GPT-4 Technical Report. arXiv. 2023;2303.08774.
7. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education. PLoS Digit Health. 2023;2(2):e0000198.
8. Gilson A, Safranek CW, Huang T, et al. How well does ChatGPT do when taking the medical licensing exams? JMIR Med Educ. 2023;9:e45312.
9. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in medical education: a pilot study. Front Med (Lausanne). 2023;10:1146082.
10. Rao A, Kim J, Kamineni M, et al. Evaluating ChatGPT as an adjunct for medical education. Med Educ Online. 2023;28(1):2188795.
11. van de Sandt S, Boulet JR, McKinley DW. The predictive validity of undergraduate GPA for medical student performance. Acad Med. 2018;93(5):725–731.
12. Harden RM, Laidlaw JM. Essential Skills for a Medical Teacher. 2nd ed. Edinburgh: Elsevier; 2017.
13. Masters K. Artificial intelligence in medical education. Med Teach. 2019;41(9):976–980.
14. Chen L, Chen P, Lin Z. Artificial intelligence in education: a review. IEEE Access. 2020;8:75264–75278.
15. Michael J. Where's the evidence that active learning works? Adv Physiol Educ. 2006;30(4):159–167.
16. Mayer RE. Cognitive theory of multimedia learning. In: Mayer RE, editor. The Cambridge Handbook of Multimedia Learning. 2nd ed. Cambridge: Cambridge University Press; 2014. p. 43–71.
17. Hattie J. Visible Learning: A synthesis of over 800 meta-analyses relating to achievement. London: Routledge; 2009.
18. Deci EL, Ryan RM. The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. Psychol Inq. 2000;11(4):227–268.
19. Prince M. Does active learning work? A review of the research. J Eng Educ. 2004;93(3):223–231.
20. Li J, Wong SH, Yang Y. Artificial intelligence-assisted learning in medical education: a systematic review. BMC Med Educ. 2023;23:512.
21. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q. 1989;13(3):319–340.
22. Means B, Toyama Y, Murphy R, Bakia M, Jones K. Evaluation of evidence-based practices in online learning: a meta-analysis and review of online learning studies. Washington (DC): US Department of Education; 2010.
23. Hosmer DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression. 3rd ed. Hoboken (NJ): Wiley; 2013.
24. Nori H, King N, McKinney SM, Carignan D, Horvitz E. Capabilities of GPT-4 on medical challenge problems. arXiv. 2023;2303.13375.
25. UNESCO. Guidance for Generative AI in Education and Research. Paris: UNESCO Publishing; 2023.