Research Square

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

The design, practice, and evaluation of AI-assisted cloud platform in medical biochemistry education

Yang-Wuyue Liu

leo90lau@tmmu.edu.cn

Army Medical University

Zhizhen Xu

Army Medical University

Tao Li

Army Medical University

Dan Zhong

Army Medical University

Wenhui He

Army Medical University

Gang Huang

Army Medical University

Shan Chen

Army Medical University

Shuang-Shuang Dai

Army Medical University

Research Article

Keywords: AI-assisted teaching, Medical biochemistry, Cloud platform, Higher-order thinking, Knowledge retention, Teacher workload

Posted Date: November 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10990443/v1

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

Additional Declarations: No competing interests reported.

Abstract

Aim

This study evaluated the effectiveness of a multi-module AI-assisted teaching cloud platform in medical biochemistry education.

Background

Biochemistry is a core medical course characterized by abstract, knowledge-dense content that challenges traditional teaching. Empirical evidence on systematically integrating multiple AI interaction modules within a complete biochemistry course remains limited.

Methods

A 20-week controlled experiment was conducted with 151 eight-year clinical medicine students, assigned to an AI platform group (n = 69, using the Chaoxing AI platform with Basic Learning, Case Discussion, Virtual Practice, and Comprehensive Assessment modules) or a control group (n = 82, using traditional tools). Final exam scores, higher-order thinking question performance, delayed post-test (6 weeks post-course) for knowledge retention, platform usage data, and questionnaire responses were collected and analyzed.

Results

The AI group significantly outperformed the control group on higher-order thinking questions (7.52 ± 0.78 vs. 5.95 ± 0.99, $p = 0.015$). Platform usage time positively correlated with final scores (R = 0.4082, $p < 0.0001$), but no significant group difference emerged in knowledge retention (65.75 ± 13.05 vs. 63.24 ± 13.42). Although 77.78% of teachers agreed the platform enriched teaching, 55.56% disagreed it reduced their workload.

Conclusions

The AI platform enhances higher-order thinking but does not boost basic knowledge retention or reduce teacher burden. Future optimization should target personalized algorithms and teacher training to improve human-computer collaboration.

1. Introduction

Biochemistry, as a core foundational course in medical education, is characterized by abstract content, strong logical coherence, and high knowledge density, making it a persistent challenge in teaching. Traditional instructional models face ongoing difficulties in enhancing learning efficiency, promoting deep knowledge internalization, and achieving personalized instruction. In recent years, the rapid advancement of artificial intelligence (AI) technology has brought transformative opportunities to medical education [1]. Research indicates that AI-assisted teaching demonstrates significant advantages in reinforcing the learning process, improving knowledge retention (the testing effect), enabling precise curriculum mapping and dynamic management, and supporting personalized learning pathways with immediate feedback. For instance, the empirical study by Schneider *et al* provided robust evidence for the "testing effect," demonstrating that active retrieval practice (answering questions) paired with instructional videos during the pre-class phase of a flipped classroom promotes knowledge understanding and long-term memory more effectively than passive reading [2]. AI platforms can seamlessly integrate this mechanism. By automating the generation, distribution, grading, and formative assessment of exercises, they can normalize and personalize "learning through testing," thereby addressing the challenge of easily forgotten biochemical knowledge points. Furthermore, the "Twelve Tips" for curriculum design proposed by Raskob *et al* reveal AI's strong capability in constructing teaching knowledge graphs, linking learning objectives with assessments, and rapidly identifying content gaps within a curriculum, which helps ensure alignment between instruction and evaluation [3]. More importantly, by analyzing student learning behaviors and assessment data, AI can outline individual knowledge mastery profiles. Combined with intelligent Q&A bots providing instant support, this moves beyond a "one-size-fits-all" teaching model, creating possibilities for personalized learning [4].

The application of AI in medical education has expanded to multiple domains. In disciplines such as clinical skills training, anatomy, and pathology, AI tools based on virtual reality, augmented reality, or adaptive learning systems have been proven effective in enhancing operational skills, spatial understanding, and diagnostic thinking [5]. However, in the teaching of biochemistry—particularly core modules like metabolism, which are highly abstract and mechanistically complex—high-quality empirical research on how to systematically design and evaluate AI-assisted teaching programs remains relatively limited [6]. Existing studies predominantly focus on clinical skills training or single instructional phases (pre-class self-study, post-class adaptive practice). There is a lack of research within a complete course system that systematically integrates multiple AI interaction modules and compares their differential effects. Simultaneously, the actual impact of AI-assisted teaching on both teachers and students remains unclear. First, its long-term instructional effectiveness, especially advantages in graduation examinations, licensing exams, and long-term knowledge retention, requires more evidence. Second, determining which AI interaction forms (Basic Learning, Case Discussion, Virtual Practice, Comprehensive Assessment) are most suitable for biochemistry learning and teaching needs further in-depth investigation. Finally, whether AI application can genuinely reduce teachers' workload—rather than increase their hidden labor—while maintaining or even enhancing teaching effectiveness still lacks quantitative analysis. This is crucial for optimizing the teacher's role and enabling them to focus on higher-value teaching activities, such as personalized tutoring and pedagogical research.

Therefore, to address these research gaps, this study conducted a controlled teaching experiment aiming to systematically evaluate the application effectiveness of a multi-module, AI-assisted teaching cloud platform in medical biochemistry education. Specifically, this study aimed to investigate the following three core questions: 1) Compared with the traditional teaching model, how effective is the AI-platform assisted teaching in improving student performance (especially higher-order thinking skills) and long-term knowledge retention? 2) What are the differences in the effectiveness and acceptance by both teachers and students among the different AI interaction modules (Basic Learning, Case Discussion, Virtual Practice, Comprehensive Assessment) in practical application? 3) What is the actual impact of the introduction of the AI platform on teachers' teaching burden and work patterns? The findings of this study are expected to provide empirical evidence and optimization directions for the in-depth and effective application of AI in basic medical science disciplines.

2. Methods and Experimental Design

2.1 Experimental Design

Experimental Grouping:

Based on existing teaching class divisions at the Army Medical University, 151 eight-year clinical medicine program students enrolled in 2024 were assigned to either the experimental group (n = 69) or the control group (n = 82). A pre-course basic knowledge test confirmed balanced foundational levels between the groups. Students in the experimental group used the Chaoxing AI Teaching Platform as the primary auxiliary tool, with each participant having a personal cloud platform login account. The control group used traditional teaching aids (PPT, textbooks, conventional online resources). Aside from this, teaching faculty, textbooks, syllabus, and assessment criteria were consistent between the two groups. The participants were medical students (not patients), the intervention was an AI-assisted teaching cloud platform, and the measured outcomes were academic performance, higher-order thinking scores, knowledge retention, and satisfaction questionnaire responses. None of these constitute health or biological outcomes, which does fall within the ICMJE definition of a clinical trial.

Data Collection:

Data was collected in three aspects throughout the experiment (as shown in Fig. 1): 1) Learning outcome data, including mid-term test scores at week 12, final exam scores at week 20, and scores from a retrospective knowledge retention test administered 6 weeks after the course concluded for both groups; 2) Learning process data, collected exclusively for the experimental group from the AI teaching platform, including usage frequency, duration, self-test scores, individual semester analysis, and preferences for different modules; 3) In-class performance rates, each students’ attendance, interaction, and accuracy in answering questions during every class session are recorded in real time as their in-class performance rate.; 4) Teacher and student experience ratings and suggestions, collected via questionnaires from both students and teachers in both groups at the end of the course, covering

satisfaction with the entire teaching process/platform use, perceived teaching load, and subjective suggestions. The details of all the questionnaires were attached on the Supplementary File 1.

2.2 Organization of Course Content

Following the knowledge system of biochemistry, the teaching content was divided into three modules based upon our previous teaching organization [7]: the Molecular Basis of Life (biological macromolecules, enzyme kinetics), the Basis of Material Metabolism (carbohydrate metabolism, lipid metabolism, amino acid metabolism, biological oxidation, metabolism of non-nutrient substances), and the Interaction Between Life and Its Periphery (fundamentals of molecular biology, cell signal transduction). Both teaching groups underwent 20 weeks of instruction, totaling 140 credit hours. The AI teaching platform integrated an intelligent teaching assistant Q&A system, knowledge graphs with online video explanations, virtual practice with standardized patient consultations, and adaptive exercise generation. Without increasing the total teaching hours, it enhanced in-class online interaction and provided real-time statistics on student participation. The control group employed methods such as centralized Q&A sessions, distribution of teaching key-point outlines, traditional offline case discussions, and periodic in-class tests.

2.3 Application Design of the Functional Modules in the Chaoxing Learning AI Application Platform

The Chaoxing Learning Platform (https://app.chaoxing.com) is an AI-enhanced smart teaching environment built upon the DeepSeek large language model [8]. It integrates over 35 AI teaching applications via the Fanya Smart Course Platform, providing three core frameworks: exclusive intelligent agents, knowledge graphs, and an AI Workbench. The usage framework includes: 1) Establishing a knowledge base for each course to train a dedicated intelligent agent; 2) Constructing knowledge graphs to structure knowledge points and personalize learning paths; 3) Integrating over 30 application scenarios—such as intelligent exercise generation, lesson plan creation, and automated assignment grading—through the AI Workbench, covering the entire "teaching, learning, research, and assessment" process.

Based on the Chaoxing Learning Platform, the teaching team designed a complete AI-empowerment workflow comprising four segments: Basic Learning, Case Discussion, Virtual Practice, and Comprehensive Assessment. The application forms of each module differ for students and teachers (see Fig. 2):

1) Basic Learning Module: Students can learn via knowledge graphs, links to digitized textbooks, and online video explanations. Teachers, by reviewing platform backend data on students' knowledge point click frequency, textbook note indexing, and questions raised during video watching, can preliminarily analyze students' learning status and key areas of focus. This provides evidence for adjusting the emphasis of in-class teaching on difficult points.

2) Case Discussion Module: Students use the platform's teaching assistant to index background information for cases assigned in advance by the teacher and to summarize discussions with group members promptly, thereby improving the efficiency of pre-class/in-class seminars. Teachers can utilize keyword cloud extraction and thematic guidance from the AI assistant to promptly grasp the key terms of student discussions and provide corrective guidance when necessary, ensuring the seminar content remains closely aligned with the teaching objectives.

3) Virtual Practice Module: Students either complete "research task-driven exploration" and "standardized patient simulated consultations" via AI practice, or they can self-assemble tests by selecting knowledge points. Based on pre-established scoring criteria and relevant knowledge points, teachers have the AI assistant grade each student participating in the virtual practice. This provides materials for teachers to comprehensively assess students' mastery of basic theory, knowledge, and skills, and to offer personalized follow-up learning suggestions.

4) Comprehensive Assessment Module: The cloud platform scores each student across three dimensions—class participation activity, knowledge point mastery, and innovative assignment evaluation—based on assessment parameters (check-in frequency, platform usage duration, discussion contribution frequency, in-class quiz accuracy, learning knowledge coverage). Each student can only view their own scores. Teachers, by collecting backend statistics of the scores across these three dimensions, can monitor the in-class/after-class learning status of the teaching cohort in real-time, focus on struggling students and weak knowledge points, promptly collect and archive students' innovative work during the learning period, and provide a basis for assigning innovation credits and evaluating Grade Point Average (GPAs) for each student.

2.4 Assessment of Learning Outcomes and Learning Process

For assessing learning outcomes, we collected the formative assessment scores (mid-term and final exams) of both groups and conducted an in-depth comparative analysis of higher-order thinking questions (comprehensive application questions). Additionally, a basic knowledge retention test was administered 6 weeks after the semester ended, and the scores of both groups were collected and compared. For assessing the learning process, the control group's grades were composed of (usual performance, seminar performance), primarily assigned directly by the instructor. The grades for students in the AI platform group were composed of (class participation activity, knowledge point mastery, innovative assignment evaluation), scored directly by the platform based on students' learning habits and knowledge mastery.

2.5 Assessment of Experience and Attitudes Towards the AI Learning Platform

Upon course completion, feedback from students and teachers was collected via questionnaire surveys (using a 5-point Likert scale), covering satisfaction and frequency of use. For the satisfaction items,

scores 1–5 respectively represented: 1: Strongly Dissatisfied; 2: Dissatisfied; 3: Neutral; 4: Satisfied; 5: Strongly Satisfied. For the frequency of use items, scores 1–5 respectively represented: 1: Unwilling to use; 2: Hardly ever use; 3: Use passively; 4: Occasionally use proactively; 5: Frequently use proactively. Additionally, anonymous comments were collected regarding students' learning experiences, difficulties, gains, and suggestions for improvement during their use of the AI platform.

2.6 Data Presentation and Analysis

For data such as GPAs and usage duration, descriptive statistics were presented as the mean and standard deviation for each variable. For the questionnaire survey (5-point Likert scale), results were presented as percentage bar chart. For personalized comments and suggestions, keywords were extracted and sorted by frequency, with the most frequently mentioned opinions and suggestions highlighted. For between-group comparisons, an independent samples t-test was used. The relationship between platform usage frequency and academic performance was analyzed using correlation analysis.

3. Results

3.1 The four AI-empowered teaching modules demonstrated different usage frequencies and satisfaction levels among students and teachers.

The four core teaching modules designed for this study—Basic Learning, Case Discussion, Virtual Practice, and Comprehensive Assessment—exhibited differentiated satisfaction levels and active usage tendencies during application. In Fig. 3, quantitative analysis of student surveys revealed that the item "4.2 Analysis of knowledge point mastery and self-assessment exercises" within the "4.0 Comprehensive Assessment" module had the highest weekly average access duration (26.54 ± 15.45 min). The overall usage proportion of the "Comprehensive Assessment" module (35.91 min,43.58%) was also significantly higher than that of the "Basic Learning (22.32 min, 27.09%)", "Case Discussion (5.87 min, 7.12%)", and "Virtual Practice (18.3 min, 22.21%)" modules. Post-course satisfaction questionnaire surveys indicated that student satisfaction was highest for the "Comprehensive Assessment" module (59.42%), particularly praising its effectiveness in promptly helping students identify knowledge weaknesses. Furthermore, the "Virtual Practice" module showed a clearly polarized reception. Some students noted in anonymous comments that activities like "standardized patient simulated consultations" and "research task-driven explorations" made abstract metabolic pathways (glucose metabolism disorders in diabetic patients, hepatic encephalopathy, ketoacidosis) vivid and concrete, greatly enhancing the sense of immersion and accomplishment in learning. However, other students reported that issues during the experience, such as "information overload provided by the AI, making it easy to lose track of the main investigative thread," were key reasons contributing to this module having the highest dissatisfaction rate (23.19%).

As shown in Fig. 4, the teacher group most frequently used the "1.6 In-class check-in and quiz" feature within the Basic Learning module and the "4.4 Full-coverage learning analysis" feature within the Comprehensive Assessment module (with weekly average usage counts of 3.11 ± 0.93 and 3.22 ± 2.54, respectively). Through the back-end data dashboard of the "Comprehensive Assessment" module,

teachers could access real-time learning analytics dashboards for the entire class and individual students. For example, after the "Enzyme Kinetics" chapter, a platform-generated report automatically indicated that 25% of students had a high error rate on exercises related to "allosteric regulation," enabling the teacher to conduct a targeted review in the subsequent class. In the "Case Discussion" module, the word cloud generated by the AI teaching assistant could quickly visualize the key thematic foci of each group's discussion, helping teachers swiftly grasp students' thinking trajectories and provide focused guidance. This data-driven approach to teaching decisions was widely recognized by teachers as enhancing the efficiency and precision of their insight into the learning situation. However, some teachers also pointed out that the massive volume of behavioral data generated by the "Basic Learning" module (video pause points, note indexing frequency) and the detailed, per-student knowledge mastery profiles from the "Comprehensive Assessment" module, while comprehensive, required additional time for interpretation, and their direct instructional guidance value needed further exploration. Additionally, although the AI assistant could autonomously grade online assignments, it required further teacher confirmation, which conversely increased the teachers' workload. More importantly, adding AI-facilitated interactive segments in class demanded that teachers invest extra time in designing corresponding teaching formats and content transitions to ensure the "close alignment of classroom discussion content with the instruction." These reasons collectively contributed to the suboptimal user experience reported by teachers.

3.2 The AI platform-empowered teaching model can enrich instructors' teaching methods and enhance student interaction, but it does not reduce the actual teaching activity burden.

Teacher & Student survey data from Fig. 5A and 5B indicate that 77.78% of teachers agreed that the platform enriched teaching methodology (Q2), and 77.78% agreed it enhanced classroom interaction (Q3). However, a majority (55.56%) of teachers disagreed that the platform reduced the teaching burden (Q1), with only 11.11% in agreement. This is further reflected in the activity analysis (Fig. 5D), where teachers in the AI group reported spending more time on "Marking and reviewing assignments" (22.22% in AI vs. 11.11% in control) and "Writing and updating teaching materials" (18.52% in AI vs. 7.41% in control). These findings demonstrate a clear dichotomy. The AI platform successfully diversified instructional methods and fostered interaction, as perceived by teachers. However, it redistributed rather than reduced their workload. The burden shifted from traditional grading to more demanding tasks like designing and adapting materials for AI-enhanced activities and managing more complex interactive sessions, leading to an overall increase in perceived workload.

3.3 In the experimental group, the usage time of the AI platform showed a significant positive correlation with students' final academic performance.

Figure 6B presents a scatter plot showing the relationship between total platform usage time (hours) and final course scores. The analysis reveals a statistically significant positive correlation with a Pearson correlation coefficient of $R^2 = 0.4082$ ($p < 0.0001$). This significant correlation suggests that greater engagement with the AI platform is associated with better learning outcomes as measured by the final

exam. It implies that the platform's resources and activities provided meaningful learning benefits. Notably, Fig. 6A and 6C show there is no clear correlation between platform time and in-class performance rates, or between in-class performance rates and assessment score.

3.4 Students in the AI platform-empowered experimental group and those in the control group performed similarly on the biochemistry basic knowledge retention test.

Figure 7A and 7B show the score distributions for the delayed post-test (basic knowledge retention) taken six weeks after the course. The average score for the AI group was 65.75 ± 13.05, while the control group averaged 63.24 ± 13.42. Furthermore, the final score distribution (Fig. 7C), which is heavily weighted towards core knowledge, was highly similar between groups across all ranges. The lack of a significant difference in delayed post-test scores indicates that the AI platform did not confer a specific advantage in the long-term retention of foundational, factual biochemistry knowledge compared to traditional methods. Both instructional approaches were equally effective (or ineffective) in promoting the durability of this type of learning.

3.5 Compared to the control group, students in the AI platform-empowered experimental group performed better on biochemistry higher-order thinking questions.

As shown in Fig. 7D, the experimental (AI) group achieved a significantly higher average score (7.52 ± 0.78 out of 10) on comprehensive application questions (a proxy for higher-order thinking) compared to the control group (5.95 ± 0.98). An unpaired t-test confirmed this difference was statistically significant ($p = 0.015$). This result is a key positive finding. It suggests that the AI platform's interactive modules—such as virtual practice, case discussions, and adaptive assessments—were particularly effective in developing students' abilities to apply, analyze, and synthesize knowledge to solve complex problems, a crucial skill in medical education.

3.6 "Respecting students' autonomy in choosing comfortable learning modes" and "avoiding information overload" are key aspects that require attention in the future application of AI platforms to biochemistry teaching.

These two points are explicitly listed among the top four suggestions from students in Fig. 7E. They emerged from qualitative feedback collected at the end of the course. The prominence of these suggestions in student feedback highlights critical user experience challenges. "Avoiding information overload" likely relates to the design of content delivery and notification systems within the platform. "Respecting...autonomy" suggests that while the platform offered new pathways, some students felt constrained by its structure or desired more control over their learning sequence and interaction styles. Addressing these aspects is essential for improving student acceptance and the effective personalization of learning.

4. Discussion

This 20-week controlled teaching experiment systematically evaluated the application effectiveness of an AI-assisted teaching cloud platform in medical biochemistry education. The findings reveal the multifaceted value and complex impact of the AI platform in teaching: Firstly, the four platform-empowered teaching modules (Basic Learning, Case Discussion, Virtual Practice, Comprehensive Assessment) demonstrated differentiated application scenarios and satisfaction levels. Among them, the "Basic Learning" and "Comprehensive Assessment" modules showed relatively high usage frequency/duration from both students and teachers. However, the "Comprehensive Assessment" module on the teacher's end (involving confirmation of online assignment grading and statistics) actually increased the teaching burden, leading to higher user dissatisfaction. The "Case Discussion" module was most relied upon by teachers, but due to reasons such as "imprecise information retrieval" and "interaction forms limited to fixed text or simple Q&A modes," it received average feedback from students. Secondly, platform usage time showed a significant positive correlation with final exam scores ($R^2 = 0.4082$, $p < 0.0001$). Specifically, active interactive learning activities (Virtual Practice, self-assessment tests) contributed the most to score improvement, whereas mandatory interactive activities (class participation metrics) showed no correlation with performance levels. Thirdly, students in the experimental group performed significantly better on higher-order thinking questions than the control group, but no significant difference was found between the two groups in the basic knowledge retention test. Fourthly, although the AI platform enriched teaching methods and increased classroom interaction frequency, it did not reduce the teacher's burden; instead, it introduced new work content.

These findings have important theoretical significance and practical value. Theoretically, this study provides new evidence for the application of the "testing effect" theory in real educational settings [9]. The AI platform realized the normalization and personalization of "learning through testing" via adaptive exercises and virtual practice, echoing the findings of Schneider *et al*. [2]. More importantly, this study revealed the differential impact of the AI platform on learning outcomes at different cognitive levels—its promoting effect on higher-order thinking skills was significantly better than its effect on the retention of basic knowledge. This suggests that AI educational tools may be more suitable for teaching scenarios requiring integrated application, innovation, and other higher-order cognitive abilities.

From a practical perspective, this study provides clear directions for optimizing the design of AI platforms. In the first place, in response to the differentiated needs of teachers and students for different modules, platforms should offer more flexible module combination options, allowing teachers to "call functions on demand" based on specific teaching objectives. Moreover, platforms need to establish more intelligent content delivery mechanisms, optimizing algorithms to provide "the right amount of information at the right time" and avoid information overload. For example, the depth and breadth of delivered content could be dynamically adjusted based on a student's learning progress and knowledge mastery. Last but not least, teacher training and support systems should be strengthened to help instructors better utilize platform data to guide teaching while alleviating their newly added workload.

Notably, the phenomenon of "increased teacher burden" discovered in this study warrants in-depth consideration. It reflects a common dilemma in current AI educational applications: while technological

tools enhance teaching effectiveness, they often require teachers to invest additional effort in course adaptation, data interpretation, and classroom management [10]. Future platform design should pay more attention to the user-friendliness of "human-computer collaboration," for example, by simplifying operational procedures and providing automated analysis reports to lower the technical barrier for teachers.

5. Study Limitations

This study has several limitations. First, the sample was relatively homogeneous, coming from a single medical university, which may affect the generalizability of the results. Second, the 20-week study period might be insufficient to observe the long-term effects of the AI platform, particularly its sustained impact on knowledge retention. Third, the platform's functionalities were not fully developed, especially the algorithms for consolidating basic knowledge, which require optimization. Fourth, this study primarily focused on quantitative metrics, with insufficient depth in the qualitative analysis of teacher and student experiences.

Future research could be improved in the following aspects: expanding the sample scope and diversity to include medical universities from different regions and types; extending the study period to 1–2 academic years to observe the sustainability of effects; strengthening the development of personalized platform algorithms, particularly spaced repetition mechanisms for basic knowledge consolidation; adopting mixed-methods research to conduct in-depth analysis of subjective experiences; and designing more refined controlled experiments targeting different knowledge types (such as declarative knowledge, procedural knowledge) to further clarify the optimal application scenarios for AI platforms.

6. Conclusion

This controlled experiment confirmed that the AI-assisted teaching cloud platform can significantly enhance the effectiveness of medical biochemistry education. One key finding was that students in the experimental group scored significantly higher on higher-order thinking questions than the control group, and platform usage time was positively correlated with performance. However, there was no significant difference between the two groups in basic knowledge retention, suggesting that solid mastery of foundational knowledge still relies on effective instructional design and sustained effort, not merely AI empowerment. A second finding revealed that the Virtual Practice and Case Discussion modules were effective in promoting knowledge application ability, while the AI-driven formative assessment effectively realized "learning through testing". The final finding indicated that the use of the platform did not reduce the teacher's burden; instead, it added to the teaching workload due to tasks such as course design and data interpretation. Based on the above findings, the authors recommend strengthening the development of personalized platform algorithms, particularly spaced repetition mechanisms for basic knowledge consolidation, balancing information provision with cognitive load to further improve student knowledge retention. Simultaneously, enhancing teacher training is crucial to maximize the efficiency of

human-computer collaboration, laying the foundation for truly alleviating teachers' burdens and further stimulating teaching innovation.

Declarations

Acknowledgements

The authors extend their sincere gratitude to the students who participated in this study and to the technician from Chaoxing who facilitated the research and data processing.

Authors' contributions

Y-W Liu was involved in conceptualization, methodology, investigation, data curation, formal analysis, and writing of the original draft as well as review and editing. Z Xu contributed to methodology, investigation, formal analysis, and writing—review and editing. Y-W Liu and Z Xu prepared all figures. T Li, D Zhong, W He and G Huang participated in teaching and data collection. S Chen and S-S Dai, as co-corresponding authors, were responsible for conceptualization, methodology, supervision, project administration, and manuscript review and editing. All authors read and approved the final manuscript.

Funding

This study is supported by The Key Construction Courses of Chongqing Municipality with the theme of "Artificial Intelligence+" (2025); Key Project of Chongqing Higher Education Teaching Reform Research (232167).

Data availability

The data in this paper are available upon reasonable request.

Ethics approval and consent to participate

The requirement for ethical approval was waived by the Medical Ethics Committee of the Army Medical University, as the study involved an anonymous questionnaire survey with no collection of identifiable personal information. Participation was entirely voluntary, and the data were used exclusively for research purposes. All participants were informed about the study prior to participation, and returning a completed questionnaire was considered indicative of informed consent. All procedures involving human participants were conducted in accordance with the ethical standards of the institutional research committee and the principles of the Declaration of Helsinki.

Consent for publication

All authors agree the publication.

Competing interests

The authors declare no competing interests.

References

1. Gordon M, Daniel M, Ajiboye A, Uraiby H, Xu NY, Bartlett R, Hanson J, Haas M, Spadafore M, Grafton-Clarke C, et al. A scoping review of artificial intelligence in medical education: BEME Guide 84. Med Teach. 2024;46(4):446–70.
2. Schneider A, Kühl M, Kühl SJ. Utilizing research findings in medical education: The testing effect within a flipped/inverted biochemistry classroom. Med Teach. 2019;41(11):1245–51.
3. Raskob K, Duman H, Kinder J, Lee J, Wilson J, Segerson K. Twelve tips to harness the power of AI for curriculum mapping. Med Teach. 2026;48(2):192–201.
4. Fortuna A, Prasetya F, Samala AD, Rawas S, Criollo-C S, Kaya D, Raihan M, Andriani W, Safitri D, Nabawi RA. Artificial intelligence in personalized learning: A global systematic review of current advancements and shaping future opportunities. Social Sci Humanit Open. 2025;12:102114.
5. Hui Z, Zewu Z, Jiao H, Yu C. Application of ChatGPT-assisted problem-based learning teaching method in clinical medical education. BMC Med Educ. 2025;25(1):50.
6. Bolgova O, Shypilova I, Mavrych V. Large Language Models in Biochemistry Education: Comparative Evaluation of Performance. JMIR Med Educ 2025, 11.
7. Zhong D, Liu YW, Huang G, Xu Z, Zhao Y, He W, Sun J, He F, Chen S, Dai SS. Change and challenge: An online course in Medical Biochemistry and Molecular Biology. Biochem Mol Biol Educ. 2023;51(5):557–65.
8. Yao Q, Zhu P, Zou Z, Cheng Y, Liang Y, Huang W. Synergising flipped classroom and case-based learning in resource-intensive anatomy education in China: a quasi-experimental study. BMJ Open. 2025;15(11):e102001.
9. Feldman M, Fernando O, Wan M, Martimianakis MA, Kulasegaram K. Testing Test-Enhanced Continuing Medical Education: A Randomized Controlled Trial. *Acad Med* 2018, 93(11S Association of American Medical Colleges Learn Serve Lead: Proceedings of the 57th Annual Research in Medical Education Sessions):S30-s36.
10. Phung JY, Chung AD, Kwan BY. Tackling Administrative Burden in Graduate Medical Education: Consider AI-Powered Tools. J Grad Med Educ. 2026;18(3):262–6.

Figures

Figure 1

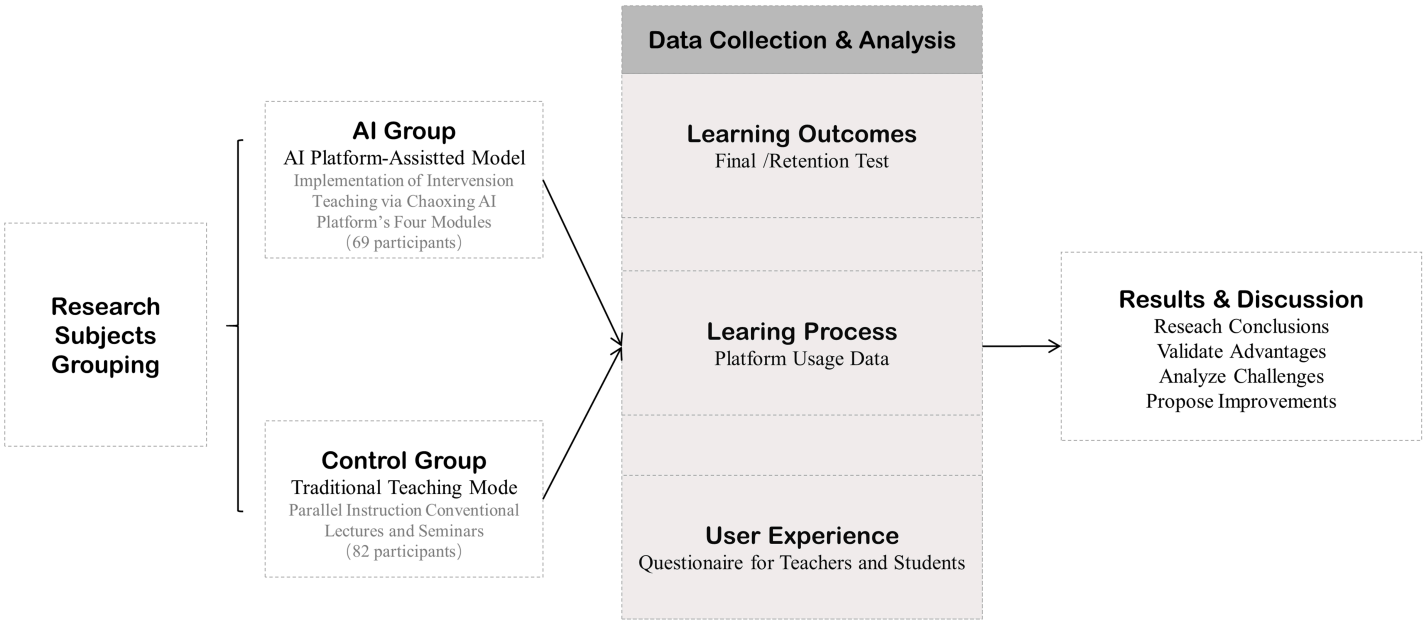


Figure 1

Schematic design of the research.

Students were assigned to either the AI-platform assisted group (n=69) or the control group (n=82) based on existing class divisions. Both groups were equivalent in terms of student demographics, teaching faculty, course schedule, and final assessment methods. Over the 20-week course, usage data for the AI-assisted group were collected via the ChaoXing platform. At the end of the course, final exam scores and questionnaire responses from both students and instructors were collected from all participants. Six weeks after the course concluded, both groups completed a delayed post-test on basic knowledge retention. The resulting data, organized into the categories of learning outcomes, teaching process, and teaching/learning experience, provided the empirical basis for addressing the study's three core research questions and formulating implementation recommendations.

AI-Empowerment Modules	AI-Empowerment Method (Student)	AI-Empowerment Method (Teacher)	AI-Empowerment Timing
1. Basic Learning	1.1 Knowledge Graph	1.4 Learning Situation Prediction	Before-class / In- Class/ After-class
	1.2 Textbook Association	1.5 Lesson Plan Writing	
	1.3 Video Analysis	1.6 In-class Checkin and Quiz	
2. Case Discussion	2.1 Background Search Engine	2.3 Discussion Data Analysis	Before-class
	2.2 Real-Time Summarization	2.4 Word Cloud Extraction	In-class
3. Virtual Practice	3.1 Research Task-Driven Exploration	3.3 Compilation of Cases and Standard Patients	In-class / After-class
	3.2 Standard Patient Simulation Consultation	3.4 Personalized Suggestions for Students	
4. Comprehensive Assessment	4.1 In Class Quiz	4.4 Full-Coverage Learning Analysis	In Class
	4.2 Knowledge Point Mastery Status and Self-Test	4.5 Weak Knowledge Point Marking	Before-class / After-class
	4.3 Innovative Assignment	4.6 Online Assignment Grading	After-class

Figure 2

Detailed AI-empowerment modules for students and teachers.

This table provides a comprehensive specification of the AI-empowered teaching platform's functional components. It categorizes the platform into four main modules (Basic Learning, Case Discussion, Virtual Practice, Comprehensive Assessment) and lists the specific AI-empowerment methods available for students and students within each module, along with their applicable timings in the teaching process (Before-class, In-class, After-class). **In basic learning module**, students could use knowledge graph, textbook association and video analysis to help with pre-class study preparation; teachers are able to use the AI platform to acquire learning situation prediction and finish lesson plan before class, as well as finish in-class checkin and quiz. **In case discussion module**, students can obtain relevant knowledge about the case background through search engines and form the group discussion conclusions through real-time summarization; teachers can achieve the alignment of the discussion content with the AI platform by summarizing the discussion information from each group and keyword extraction. **In virtual practice module**, students can complete the research task-driven exploration and standard patient simulation consultation based on the knowledge points they have learned; the teacher could obtain the "compilation of cases and standard patients" as well as "personalized learning suggestions for students" based on each student's usage records. **In comprehensive assessment**

module, students can not only participate in in-class quizzes and innovative assignments and earn corresponding course credits, but also test their learning outcomes through the "knowledge point mastery status and self-test" section of the platform; teachers can promptly grasp the basic learning situation of the classes they are teaching through "full-coverage learning analysis and weak knowledge point marking", as well as assist in the grading of online assignments through the platform.

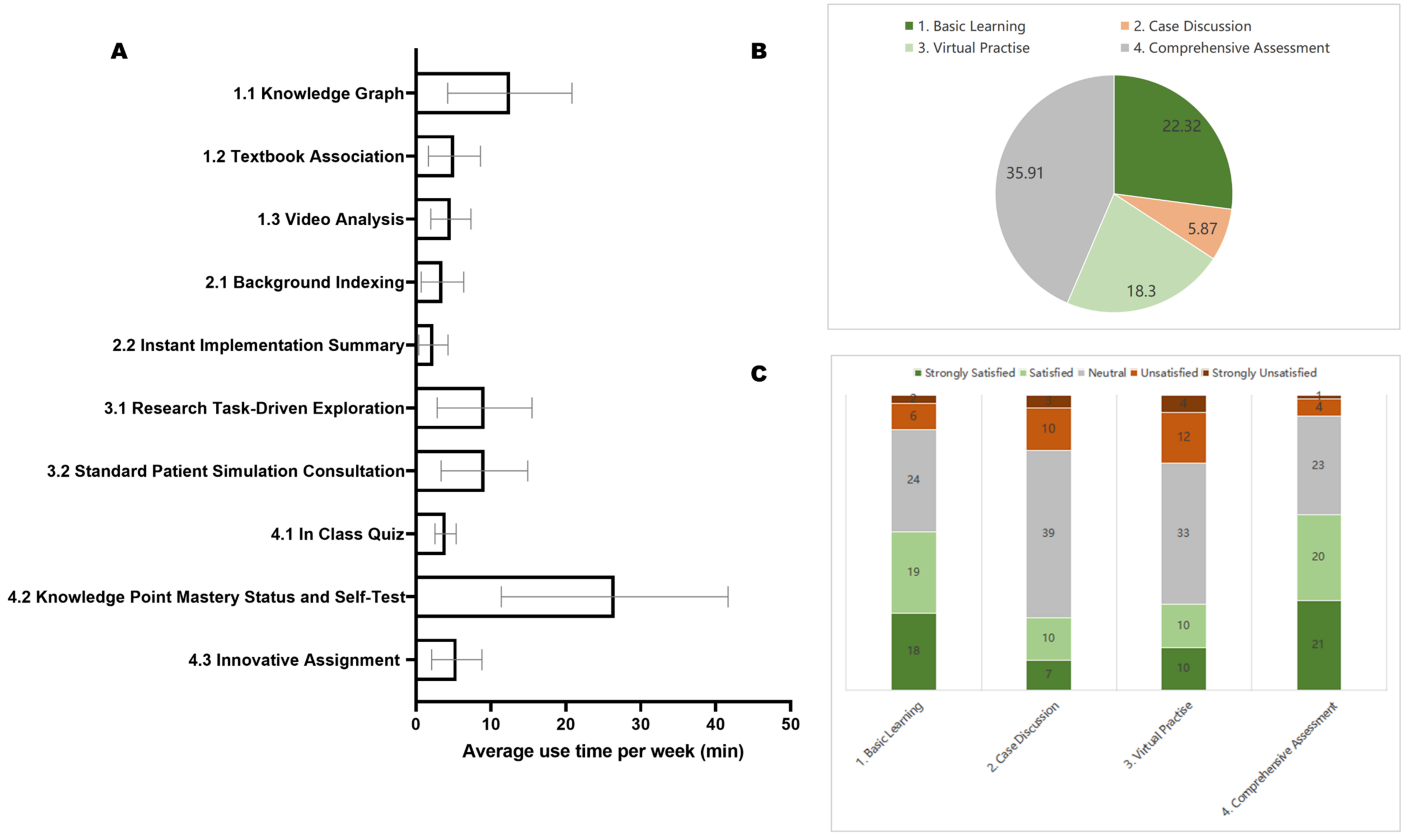


Figure 3

The usage situation and satisfaction level of students within different modules of the AI platform.

A. The average use time (min) per week from students for each AI platform associated activities, n=69. The data were represented as average ± SD, including 1.1 knowledge graph (13.41 ± 8.27), 1.2 textbook association(5.51 ± 3.34), 1.3 video analysis (4.61 ± 2.76), 2.1 background search engine (3.64 ± 2.89), 2.2 real-time summarization (2.35 ± 2.31), 3.1 research task-driven exploration (9.20 ± 6.30), 3.2 standard patient simulation consultation (9.35 ± 5.73), 4.1 in class quiz (3.90 ± 1.44), 4.2 knowledge point mastery status and self-test (26.54 ± 15.45), 4.3 innovative assignment (5.33 ± 3.37) . **B.** The comprehensive average usage duration of the four modules and their respective proportions. The students' average use time per week and proportion for each AI modules were: basic learning (22.32 min, 27.09%), case discussion (5.87 min, 7.12%), virtual practice (18.3 min, 22.21%), comprehensive assessment (35.91 min, 43.58%). **C.** The panels showed the distribution of students' usage satisfaction (strongly satisfied to strongly dissatisfied) to four AI modules. The numbers for each levels were marked,

n=69. In basic learning module, the total satisfied rate was 53.62%, the dissatisfied rate was 11.59%. In case discussion module, the total satisfied rate was 24.63%, the dissatisfied rate was 18.84%. In virtual practice module, the total satisfied rate was 28.99%, the dissatisfied rate was 23.19%. In comprehensive assessment module, the total satisfied rate was 59.42%, the dissatisfied rate was 7.25%.

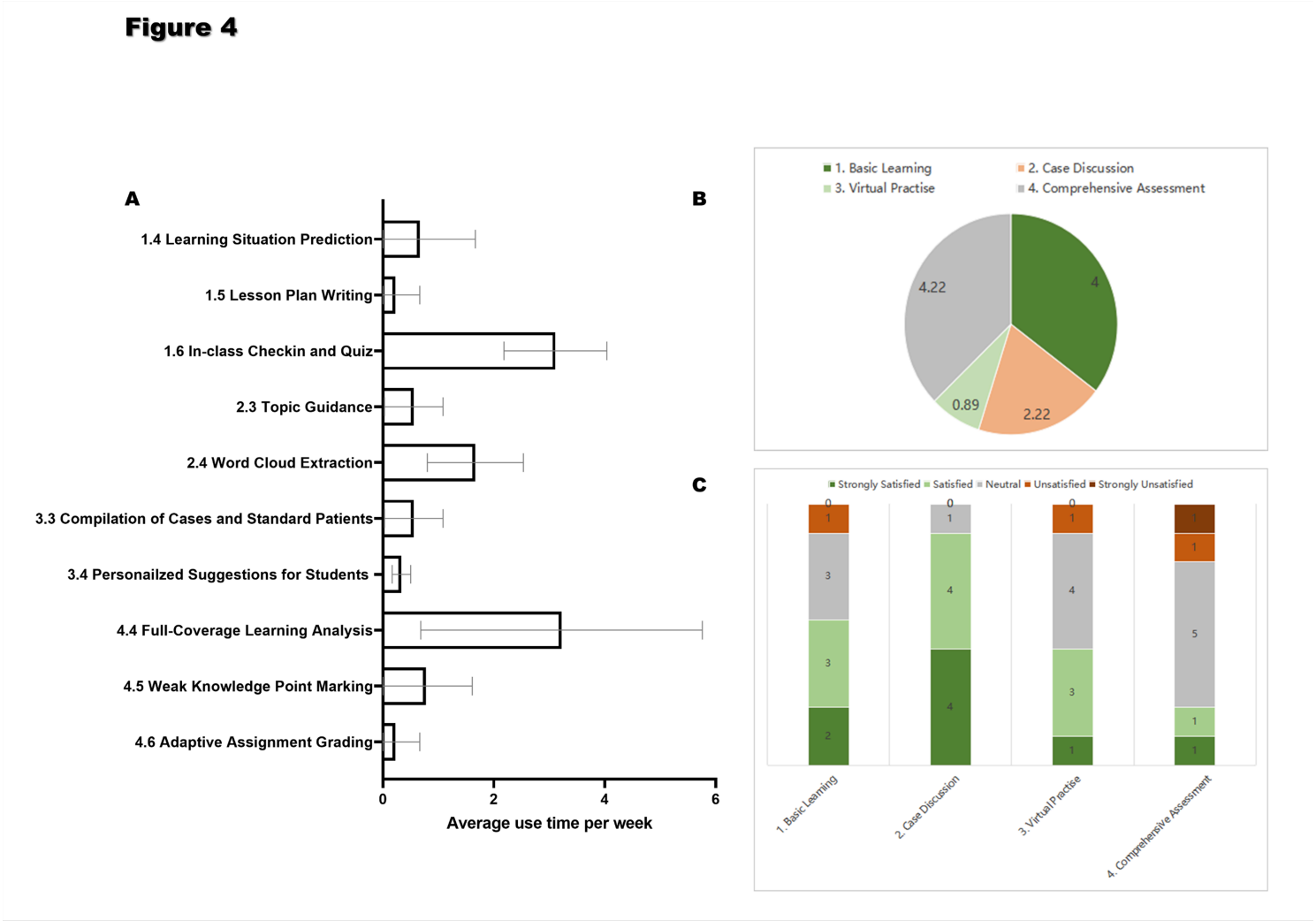


Figure 4

The usage situation and satisfaction level of teachers within different modules of the AI platform.

A. The average use time per week from teachers for each AI platform associated activities, n=9. The data were represented as average ± SD, including 1.4 learning situation prediction (0.67 ± 1.00), 1.5 lesson plan writing (0.22 ± 0.44), 1.6 in-class checkin and quiz (3.11 ± 0.93), 2.3 discussion data analysis (0.56 ± 0.53), 2.4 word cloud extraction (1.67 ± 0.87), 3.3 compilation of cases and standard patients (0.56 ± 0.53), 3.4 Personalized suggestions for students (0.33 ± 0.50), 4.4 full-coverage learning analysis (3.22 ± 2.54), 4.5 weak knowledge point marking (0.78 ± 0.83), 4.6 online assignment grading (0.22 ± 0.44). **B.** The comprehensive average usage time of the four modules and their respective proportions. The teachers' average use time per week and proportion for each AI modules were: basic learning (4, 35.30%), case discussion (2.22, 19.59%), virtual practice (0.89, 7.86%), comprehensive assessment

(4.22, 37.25%). **C.** The panels showed the distribution of teachers' usage satisfaction (strongly satisfied to strongly dissatisfied) to four AI modules. The numbers for each levels were marked, n=9. In basic learning module, the total satisfied rate was 55.56%, the dissatisfied rate was 11.11%. In case discussion module, the total satisfied rate was 88.89%, the dissatisfied rate was 0%. In virtual practice module, the total satisfied rate was 44.44%, the dissatisfied rate was 11.11%. In comprehensive assessment module, the total satisfied rate was 22.22%, the dissatisfied rate was 22.22%.

Figure 5

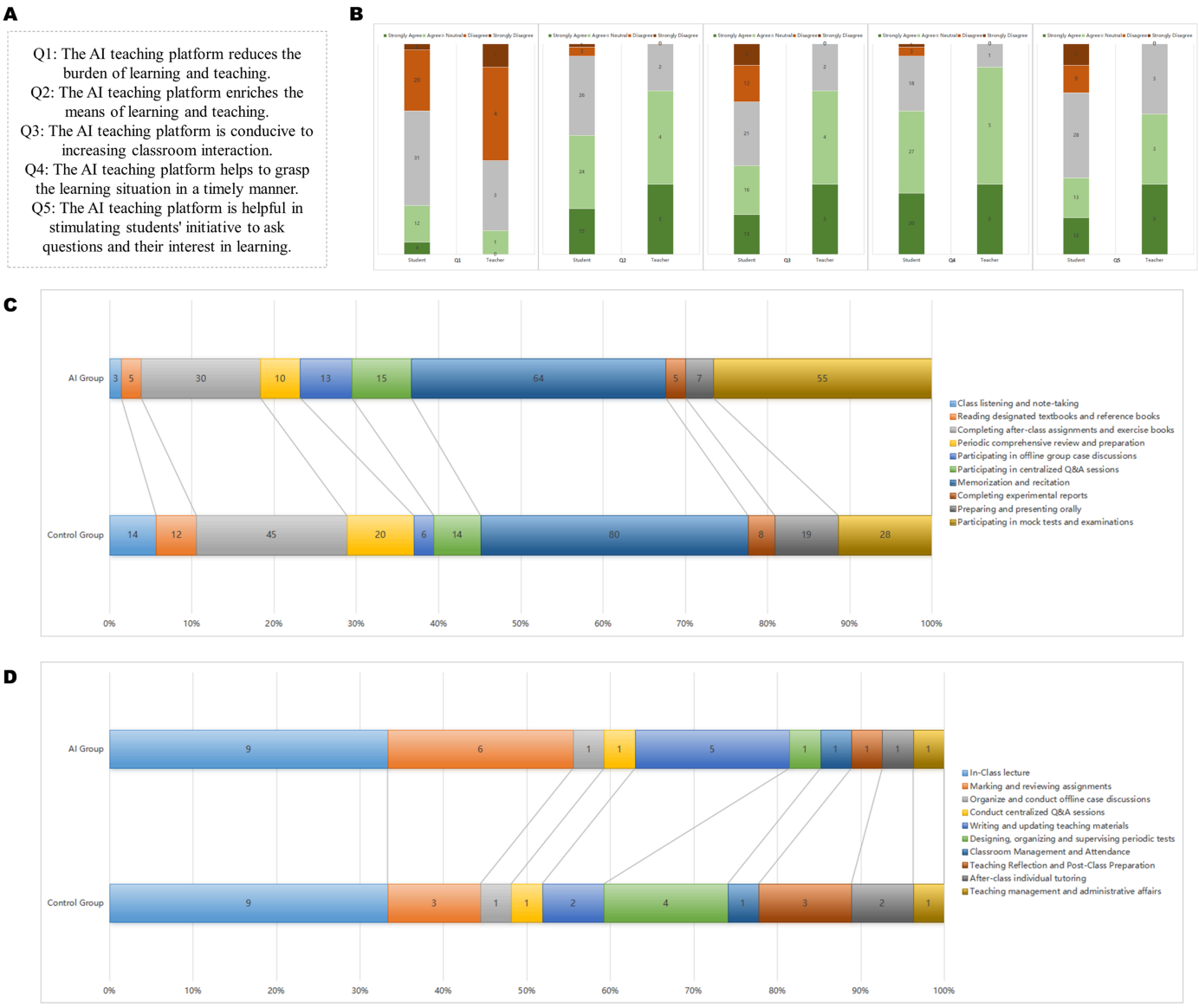


Figure 5

The perceptions of AI facilitated teaching from students and teachers.

A. The questionnaire for five key statements (Q1-Q5) regarding the platform's impact on learning/teaching burden (Q1), methodology enrichment (Q2), classroom interaction (Q3), learning situation awareness (Q4), and student initiative/interest (Q5). **B.** The panels showed the distribution of responses (Strongly Agree to Strongly Disagree) to five key statements (Q1-Q5) from students (n=69) and teachers (n=9) respectively. For Q1, the total agree rates were 23.19% (students) and 11.11% (teachers), total disagree rates were 31.88% (students) and 55.56% (teachers). For Q2, the total agree rates were 56.52% (students) and 77.78% (teachers), total disagree rates were 5.80% (students) and 0% (teachers). For Q3, the total agree rates were 42.03% (students) and 77.78% (teachers), total disagree rates were 27.54% (students) and 0% (teachers). For Q4, the total agree rates were 68.12% (students) and 88.89% (teachers), total disagree rates were 5.80% (students) and 0% (teachers). For Q5, the total agree rates were 36.23% (students) and 55.56% (teachers), total disagree rates were 23.19% (students) and 0% (teachers). **C.** At the end of the course, each student selected three of the ten main learning activities that they have participated in the most, n=69 for AI group, n=82 for control group. The proportion for each activities in each group (AI group *vs* control group) were: Class listening and note-taking (1.45% *vs*5.69%), Reading designated textbooks and reference books (2.42% *vs*4.89%), Completing after-class assignments and exercise books (14.49% *vs*18.29%), Periodic comprehensive review and preparation (4.83% *vs* 8.13%), Participating in offline group case discussions (6.28% *vs* 2.44%), Participating in centralized Q&A sessions (7.25% *vs* 5.69%), Memorization and recitation (30.92% *vs* 32.52%), Completing experimental reports (2.41% *vs* 3.25%), Preparing and presenting orally (3.38% *vs*7.72%), Participating in mock tests and examinations (26.57% *vs* 11.38%). **D.** At the end of the course, each teacher selected three of the ten main teaching activities that they have participated in the most, n=9 for both AI group and control group. The proportion for each activities in each group (AI group vs control group) were: In-Class lecture (33.33% *vs* 33.33%), Marking and reviewing assignments (22.22% *vs* 11.11%), Organize and conduct offline case discussions (3.70% *vs* 3.70%), Conduct centralized Q&A sessions (3.70% *vs* 3.70%), Writing and updating teaching materials (18.52% *vs* 7.41%), Designing, organizing and supervising periodic tests (3.70% *vs* 14.81%), Classroom Management and Attendance (3.70% *vs*3.70%), Teaching Reflection and Post-Class Preparation (3.70% *vs*11.11%), After-class individual tutoring (3.70% *vs* 7.41%), Teaching management and administrative affairs(3.70% *vs* 3.70%).

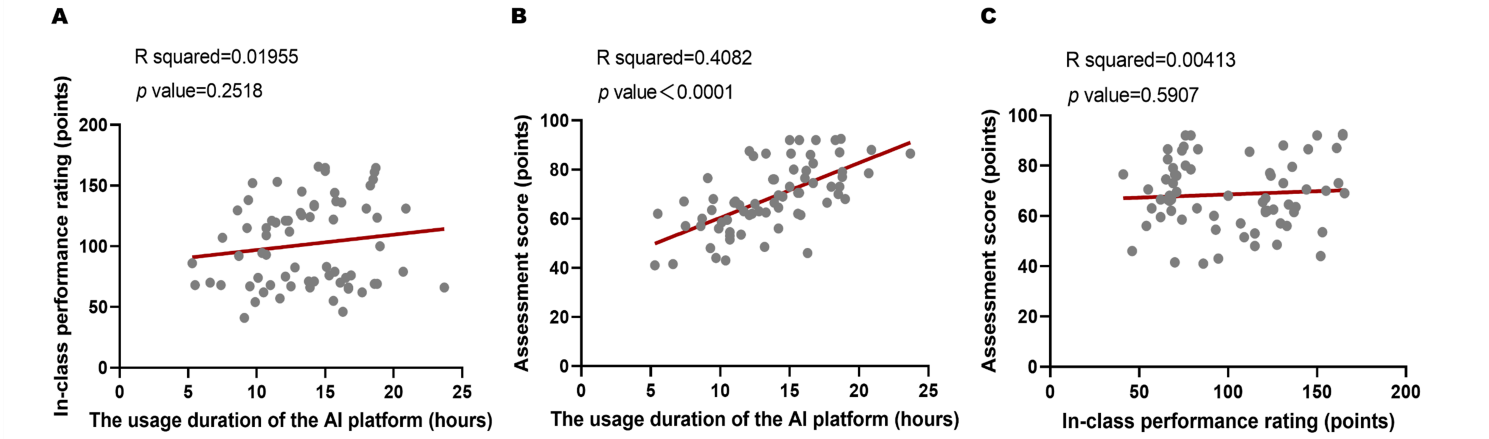


Figure 6

The correlation analysis between the usage of AI platforms and course grades.

A. The scatter plots displayed the relationship between the total time (hours) spent by students on the AI platform (horizontal axis) and their in-class performance rates (vertical axis). R^2=0.01955, p=0.2518. **B.** The scatter plots displayed the relationship between the total time (hours) spent by students on the AI platform (horizontal axis) and their final course scores (vertical axis). R^2=0.4082, p<0.0001. **C.** The scatter plots displayed the relationship between the in-class performance rates (horizontal axis) and their final course scores (vertical axis). R^2=0.00413, p=0.5907. The red trend lines were added for each graph, n=69.

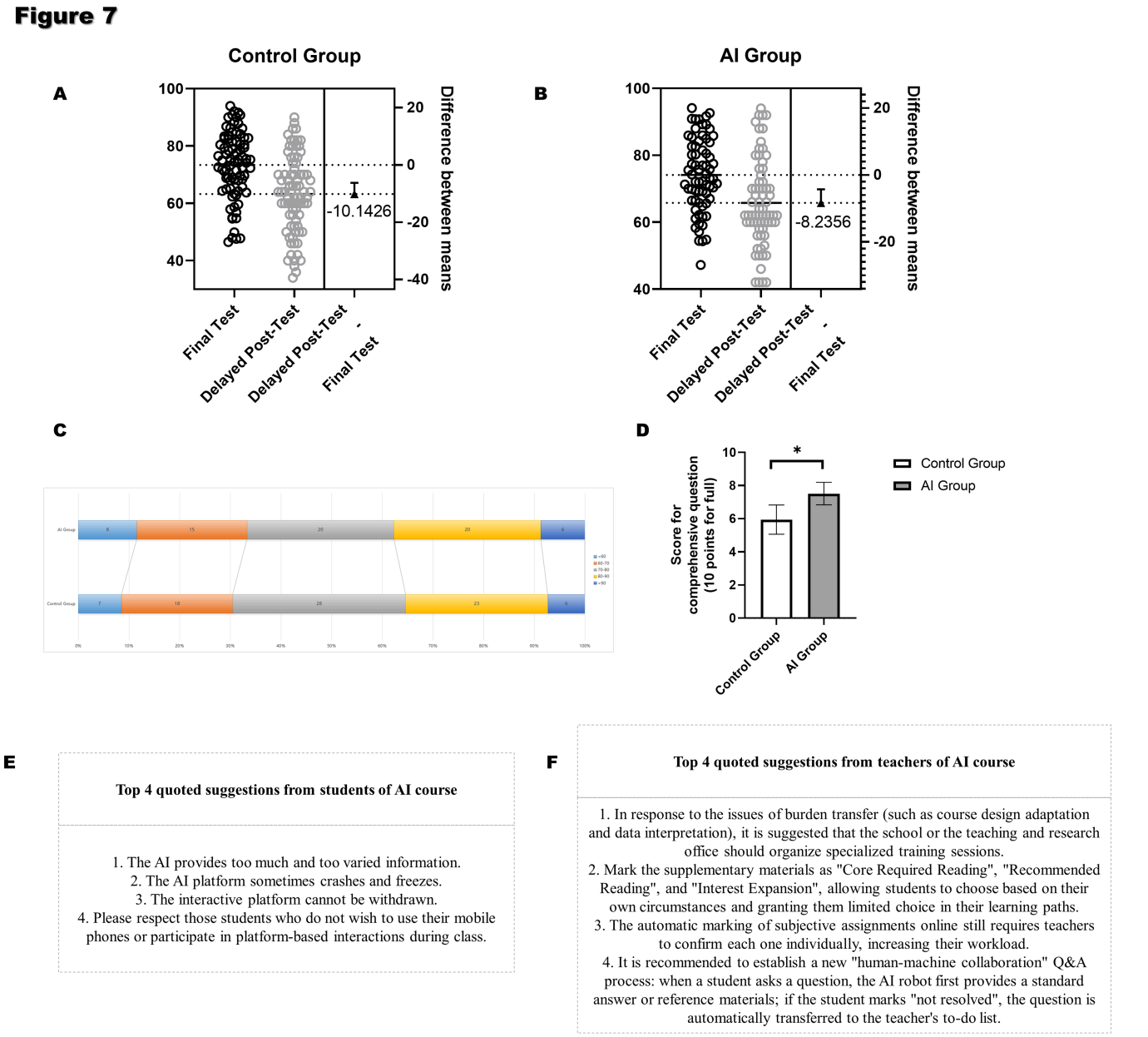


Figure 7

The analysis of knowledge retention and feed-backs after finishing the whole course.

The final test was made of multiple-choice questions, definition questions, short-answer questions and comprehensive application questions (100 score for full). Six weeks after the course concluded, the delayed post-test made of only multiple-choice questions (100 score for full) were exerted in both group. **A.** The score distribution of final test and delayed post-test in control group, n=82. The average scores for final test and delayed post-test were 73.39 ± 11.58 and 63.24 ± 13.42 respectively. **B.** The score distribution of final test and delayed post-test in AI group, n=69. The average scores for final test and delayed post-test were 74.13 ± 10.96 and 65.75 ± 13.05 respectively. **C.** In final test, the distribution of the number of students from different score ranges. The percentages of specific ranges from AI and

control groups were (AI vs control): <60 (11.59% vs 8.54%), 60-70 (21.74% vs 21.95%), 70-80 (28.99% vs 34.15%), 80-90 (28.99% vs 28.05%), and >90 (8.70% vs 7.32%). **D.** In final test, the scores of comprehensive application questions from AI (7.52 ± 0.78, n=69) and control group (5.95 ± 0.98, n=82), 10 points for full. The difference was compared with unpaired t test, p=0.015. **E.** The left column lists the top four suggestions from students (information overload, technical stability, interaction flexibility, respect for diverse participation preferences). **F.** The right column lists the top four suggestions from teachers (need for specialized training, structured content tagging, workload from AI-assisted grading, optimized human-AI Q&A workflow).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Questionnairefile.docx