

Artificial Intelligence-Assisted Learning, Study Practices, and Academic Performance among Undergraduate Medical Students: A Cross-Sectional Study in Physiology Education

Jeya Lisha John¹, Bhavithra Bharathi S¹, Delinda Linu Swornila D.¹, N. Parvatha Rani¹, Lavanya V.¹

¹Department of Physiology, Kanyakumari Medical Mission Research Centre, Muttom, Kanyakumari District

Received: 01-07-2026 / Revised: 15-08-2026 / Accepted: 21-09-2026

Corresponding author: Dr. Lavanya V.

Conflict of interest: Nil

Abstract

Background: Artificial intelligence (AI) tools are increasingly being incorporated into medical education, but their association with academic performance and students' perceptions in undergraduate physiology learning remains uncertain. This study evaluated the relationship between students' preferred learning methods involving AI and their academic performance, perceptions of AI-assisted learning, and study-related factors in undergraduate physiology.

Methods: A cross-sectional study was conducted among 60 undergraduate medical students learning Endocrine Physiology. Students were categorized according to their primary learning method as traditional learning alone, traditional learning supplemented with AI, or predominantly AI-based learning. Academic performance was assessed using a formative assessment comprising knowledge (Knows), case-based, reasoning, and Knows-how components. Students' perceptions regarding the educational benefits, responsible use, and concerns related to AI were assessed using a structured Likert-scale questionnaire. Group differences were assessed using the Kruskal–Wallis test, and associations between academic performance, study time, previous academic performance, and AI-related perceptions were examined using Spearman's correlation.

Results: Of the 60 students, 21 (35.0%) reported traditional learning alone, 31 (51.7%) traditional learning supplemented with AI, and 8 (13.3%) predominantly AI-based learning. Total formative assessment scores did not differ significantly across the three groups ($p = .189$), although the mean score was numerically higher in the traditional + AI group ($52.4 \pm 19.7\%$) than in the traditional-only ($42.5 \pm 17.7\%$) and AI-based ($43.5 \pm 26.3\%$) groups. No significant differences were observed for the Knows, case-based, reasoning, or Knows-how components. Overall perceptions of AI were favourable, with 63.4% of students rating AI as good or excellent and 81.7% recommending its use for future batches. Most students agreed that AI should supplement rather than replace teachers (88.3%) and that faculty guidance was necessary for effective AI use (83.4%). Previous academic performance showed significant positive correlations with Total FA score ($p = 0.599$, $p < .001$) and all individual performance components. Study time was also positively correlated with Total FA score ($p = 0.333$, $p = .014$), the Knows component ($p = 0.406$, $p = .002$), and reasoning performance ($p = 0.369$, $p = .006$). Perceived educational benefits were not associated with improved academic performance.

Conclusion: Students demonstrated favourable attitudes towards the use of AI in physiology learning, particularly when AI was viewed as a supplement to conventional teaching. However, the learning method involving AI was not significantly associated with academic performance. Previous academic achievement and study time showed stronger associations with current performance, suggesting that AI may serve as a supportive educational tool rather than an independent determinant of academic achievement. Faculty guidance and appropriate integration of AI into conventional teaching may be important for its effective use in undergraduate medical education.

Keywords: Artificial Intelligence; Medical Education; Physiology; Undergraduate Medical Students; Academic Performance; Generative AI; Student Perceptions.

DOI: 10.25258/ijcpr.18.10.75

This is an Open Access article that uses a funding model which does not charge readers or their institutions for access and distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>) and the Budapest Open Access Initiative (<http://www.budapestopenaccessinitiative.org/read>), which permit unrestricted use, distribution, and reproduction in any medium, provided original work is properly credited.

Introduction

Artificial intelligence (AI), particularly generative AI and large language models (LLMs), has rapidly become accessible to medical students and is increasingly being incorporated into everyday learning practices. Tools such as ChatGPT, Gemini, Claude, Grok, and other AI-based platforms can provide rapid explanations, summarise complex topics, generate examples, answer questions, and facilitate self-directed learning. These capabilities have created new opportunities for personalised and interactive learning in medical education. Recent evidence indicates that generative AI is being used across several learning activities, including knowledge acquisition, inquiry, practice, and production, although its role in collaborative learning remains less explored [1].

The increasing availability of AI-based learning tools has also raised concerns regarding their educational consequences. Although AI can improve accessibility and efficiency, inaccurate or fabricated information, excessive dependence on generated answers, reduced critical thinking, superficial learning, and academic integrity remain important concerns. Systematic reviews of AI in medical education have highlighted both its potential as a learning aid and the need for appropriate guidance regarding its limitations and responsible use [2,3].

The educational effectiveness of generative AI remains uncertain. Evidence from recent systematic reviews and meta-analyses suggests that AI-supported learning may improve some educational outcomes, but findings are heterogeneous and appear to depend on the learning context, intervention, learner characteristics, and outcome measured. A recent meta-analysis of randomized controlled trials found no statistically significant overall difference in knowledge acquisition between generative AI-based and conventional teaching approaches, emphasising that AI should currently be viewed as a potential supplement rather than a replacement for established educational strategies [4].

This issue is particularly relevant to undergraduate physiology, where students must develop not only factual knowledge but also conceptual understanding and physiological reasoning. AI tools may provide rapid explanations and facilitate independent exploration of difficult concepts; however, the extent to which their use is associated with measurable academic performance among undergraduate medical students remains unclear. Previous work has also demonstrated that generative AI can perform well on physiology examinations, highlighting both its educational

potential and the importance of understanding how students themselves use such tools for learning [5].

Therefore, the present study was undertaken to assess the pattern of AI use among undergraduate medical students, examine their preferred learning approaches, and evaluate the association between AI-assisted learning, study practices, and academic performance in an Endocrine Physiology formative assessment. The study also explored students' perceptions regarding the usefulness and potential limitations of AI as a learning resource.

Methodology

Study Design and Setting: A cross-sectional questionnaire-based observational study was conducted among undergraduate medical students in the Department of Physiology. The study focused on students' learning practices in relation to an Endocrine Physiology formative assessment (FA).

Study Participants: Students who participated in the relevant formative assessment and responded to the study questionnaire were included. A total of 60 valid responses were obtained and included in the analysis.

Data Collection: Data were collected using a structured questionnaire administered to the students following the Endocrine Physiology formative assessment. Participation was voluntary, and the questionnaire collected information regarding students' use of AI-based and conventional learning resources.

Students were asked whether they used AI for preparation and, where applicable, which AI-based tool they used. The AI platforms reported by students included ChatGPT, Gemini, Claude, Grok, and Google AI. Information regarding the frequency of AI use and the student's primary learning method was also collected.

For analysis, students were categorised according to their reported learning approach, including traditional learning, AI-based learning, and combined use of AI and traditional learning resources, as applicable to their responses.

Assessment of Academic Performance: Academic performance was assessed using the students' marks obtained in the Endocrine Physiology formative assessment.

The assessment outcome was analysed both as an overall FA score and, according to components assessing different cognitive domains, including knowledge ("Knows") and reasoning.

Assessment of Study Practices: Students reported their average study time, expressed as the

approximate number of hours spent studying. The relationship between reported study time and academic performance was assessed using the total FA score.

Assessment of Perceptions: The questionnaire also included items assessing students' perceptions regarding AI-assisted learning and its potential advantages and concerns. Responses included Likert-type options assessing perceptions such as usefulness, understanding, learning efficiency, confidence, and concerns related to dependence on AI, accuracy of AI-generated information, and superficial learning.

Statistical Analysis: Data were analysed using jamovi version 2.7.38. Categorical variables were summarised using frequencies and percentages, while continuous variables were summarised using appropriate measures of central tendency and dispersion. Associations between categorical variables were assessed using appropriate tests based on the distribution and expected cell frequencies. Academic performance between learning-method groups was compared using appropriate inferential tests.

Because study-time and academic-performance variables were not assumed to follow a normal distribution and the variables included ordinal/continuous measures, Spearman's rank correlation coefficient (ρ) was used to assess associations between average study time and

academic performance. Correlations were also examined separately for the "Knows" and reasoning components of the assessment. A two-sided p value <0.05 was considered statistically significant.

Results

A total of 60 students participated in the study. Based on their reported primary learning method, 21 students (35.0%) used traditional learning methods alone, 31 (51.7%) used traditional learning supplemented with AI, and 8 (13.3%) reported predominantly AI-based learning.

Previous Academic Performance and Study Time: Previous IA performance was most commonly in the 50–65% category (30/60, 50.0%), followed by $<50\%$ (18/60, 30.0%), 66–75% (7/60, 11.7%), and $>75\%$ (5/60, 8.3%). The distribution of previous IA performance categories did not differ significantly across the three learning-method groups ($\chi^2(6) = 5.90$, $p = .435$; Fisher's exact test, $p = .476$) (Table 1). The mean reported study time was 4.55 ± 2.95 hours, with a median of 4.0 hours (range, 1–16 hours). Mean study time was 4.54 ± 3.71 hours among students using traditional learning alone, 4.77 ± 2.42 hours among those using traditional learning supplemented with AI, and 3.75 ± 2.82 hours among those using AI-based learning. Study time did not differ significantly among the three groups (Kruskal–Wallis $\chi^2(2) = 2.02$, $p = .365$).

Table 1: Characteristics of the study participants according to learning method

Characteristic	Traditional only (n=21)	Traditional + AI (n=31)	AI-based (n=8)	Total (N=60)
Previous IA performance, n (%)				
$<50\%$	4 (19.0)	9 (29.0)	5 (62.5)	18 (30.0)
50–65%	12 (57.1)	16 (51.6)	2 (25.0)	30 (50.0)
66–75%	3 (14.3)	3 (9.7)	1 (12.5)	7 (11.7)
$>75\%$	2 (9.5)	3 (9.7)	0 (0.0)	5 (8.3)
Average study time (hours)				
Mean $\pm$ SD	4.54 ± 3.71	4.77 ± 2.42	3.75 ± 2.82	4.55 ± 2.95
Median (range)	3.0 (1–16)	5.0 (1–10)	3.5 (1–10)	4.0 (1–16)

Academic performance according to learning method: Academic performance measures are presented in Table 2. Students using traditional learning supplemented with AI had numerically higher mean scores across all assessed components.

The mean Total FA score was $52.4 \pm 19.7\%$ in the traditional + AI group, compared with $42.5 \pm 17.7\%$ in the traditional-only group and $43.5 \pm 26.3\%$ in the AI-based group. However, the difference was not statistically significant (Kruskal–Wallis $p = .189$). Similarly, no statistically significant differences were observed

for the Knows component, with mean scores of $48.7 \pm 19.4\%$, $56.6 \pm 20.6\%$, and $50.0 \pm 24.5\%$ in the traditional-only, traditional + AI, and AI-based groups, respectively ($p = .218$). Mean case-based scores were $35.9 \pm 19.8\%$, $47.8 \pm 26.0\%$, and $37.1 \pm 29.8\%$, respectively ($p = .288$). Mean reasoning scores were $35.5 \pm 37.3\%$, $47.8 \pm 31.0\%$, and $33.3 \pm 30.3\%$, respectively ($p = .323$). Mean Knows-how scores were $35.9 \pm 19.6\%$, $47.8 \pm 24.5\%$, and $36.5 \pm 29.0\%$, respectively ($p = .189$). Thus, none of the assessed academic performance measures differed significantly according to learning method.

Table 2: Academic performance according to learning method.

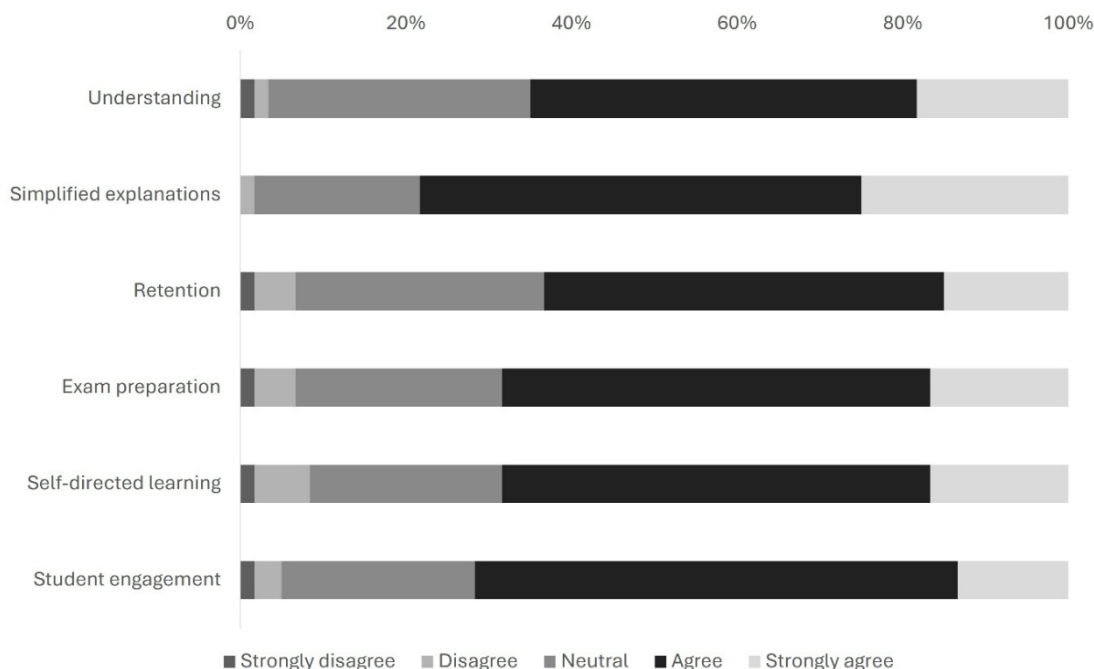
Characteristic	Traditional only (n=19)	Traditional + AI (n=29)	AI-based (n=6)	p
Total FA, mean $\pm$ SD	42.5 $\pm$ 17.7	52.4 $\pm$ 19.7	43.5 $\pm$ 26.3	0.189
Knows component, mean $\pm$ SD	48.7 $\pm$ 19.4	56.6 $\pm$ 20.6	50.0 $\pm$ 24.5	0.218
Case-based marks, mean $\pm$ SD	35.9 $\pm$ 19.8	47.8 $\pm$ 26.0	37.1 $\pm$ 29.8	0.288
Reasoning marks, mean $\pm$ SD	35.5 $\pm$ 37.3	47.8 $\pm$ 31.0	33.3 $\pm$ 30.3	0.323
Knows-how component, mean $\pm$ SD	35.9 $\pm$ 19.6	47.8 $\pm$ 24.5	36.5 $\pm$ 29.0	0.189

Values are expressed as percentages and presented as mean $\pm$ standard deviation. p values are from Kruskal–Wallis tests. None of the academic performance measures differed significantly across learning-method groups.

Students' perceptions regarding AI use in Physiology: The item-wise distribution of perceptions is presented in Figures 1-3. Overall, students reported favourable perceptions regarding the educational benefits of AI. A total of 39 students (65.0%) agreed or strongly agreed that AI helped them understand physiology concepts, while 47 (78.3%) agreed or strongly agreed that AI provided simplified explanations. Similarly, 38 students (63.3%) reported that AI aided retention, 41 (68.4%) perceived benefit for examination preparation, 41 (68.4%) reported support for self-directed learning, and 43 (71.7%) agreed or strongly agreed that AI improved student

engagement. Students also expressed favourable views regarding the appropriate integration of AI into medical education. Forty-three students (71.7%) agreed or strongly agreed that AI use should be encouraged in undergraduate medical education. Most students (53/60, 88.3%) agreed or strongly agreed that AI should be used as a supplement rather than a replacement for teachers, while 50 students (83.4%) agreed or strongly agreed that faculty guidance was necessary for effective use of AI tools.

Regarding potential concerns, 48 students (80.0%) disagreed or strongly disagreed that overdependence on AI may reduce critical thinking. Similarly, 47 students (78.3%) disagreed or strongly disagreed that the accuracy of AI-generated information was a concern, and 44 students (73.4%) disagreed or strongly disagreed that AI may promote superficial learning.

**Figure 1: Students' perceptions of the educational benefits of artificial intelligence in physiology.**

The 100% stacked bar chart illustrates the distribution of responses across the five-point Likert scale for perceived improvement in understanding, simplification of explanations, retention, examination preparation, self-directed learning, and student engagement.

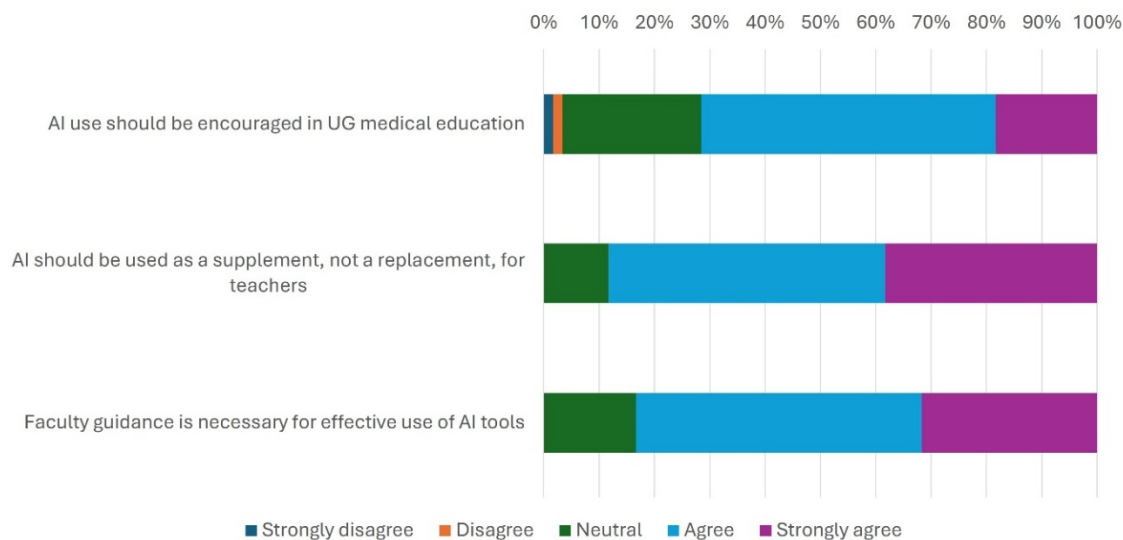


Figure 2: The 100% stacked bar chart presents the distribution of responses to items concerning encouragement of AI use, use of AI as a supplement rather than a replacement for teachers, and the need for faculty guidance.

Responses are shown on a five-point Likert scale from strongly disagree to strongly agree.

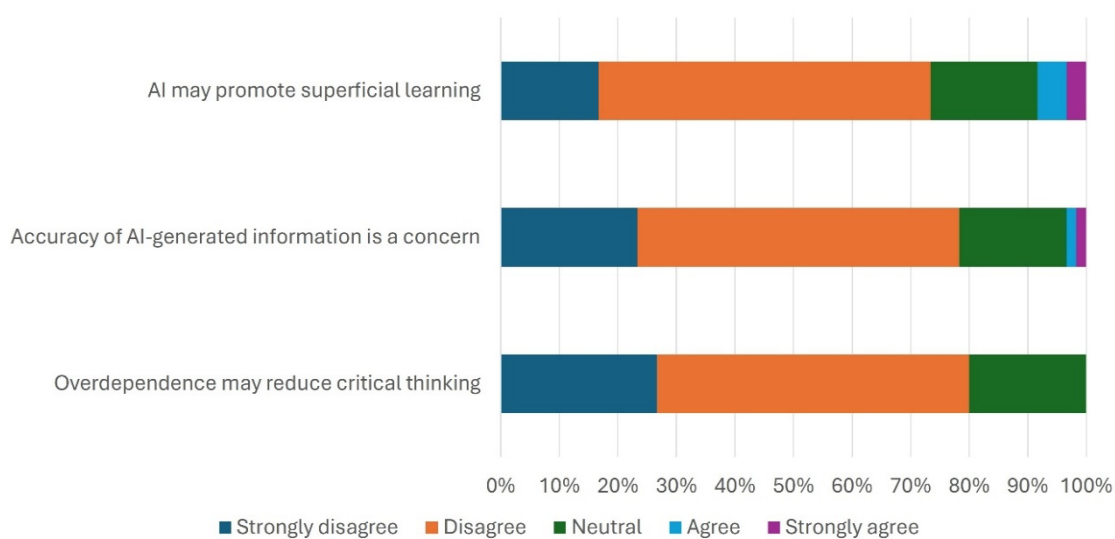


Figure 3: Students' perceptions of potential concerns associated with artificial intelligence use in physiology education.

The figure shows the distribution of responses concerning overdependence, accuracy of AI-generated information, and the potential for superficial learning.

AI-related perception scores according to learning method: The composite perception scores are shown in Table 3. The mean educational-benefit score was 3.71 ± 0.62 among students using traditional learning, 3.81 ± 0.73 among those using traditional learning supplemented with AI, and 3.98

± 0.90 among those using AI-based learning. There was no significant difference between the three groups (Kruskal-Wallis $p = .885$). Mean responsible-AI-use scores were 4.02 ± 0.48 , 4.17 ± 0.55 , and 3.96 ± 0.88 in the traditional-only, traditional + AI, and AI-based groups, respectively, with no statistically significant difference ($p = .655$). Similarly, AI-concern scores were 2.02 ± 0.54 , 2.01 ± 0.81 , and 2.38 ± 0.74 , respectively, without a significant between-group difference ($p = .287$).

Table 3: AI-related perceptions according to learning method.

Characteristic	Traditional only	Traditional + AI	AI-based	p
Educational benefits	3.71 ± 0.62	3.81 ± 0.73	3.98 ± 0.90	.885
Responsible AI use	4.02 ± 0.48	4.17 ± 0.55	3.96 ± 0.88	.655
AI concern score	2.02 ± 0.54	2.01 ± 0.81	2.38 ± 0.74	.287

Values are expressed as mean ± standard deviation. Scores were derived from Likert-scale questionnaire items, with higher scores indicating greater agreement with the respective construct. p values are from Kruskal–Wallis tests. None of the comparisons between learning-method groups were statistically significant.

Overall opinion and recommendation regarding AI:

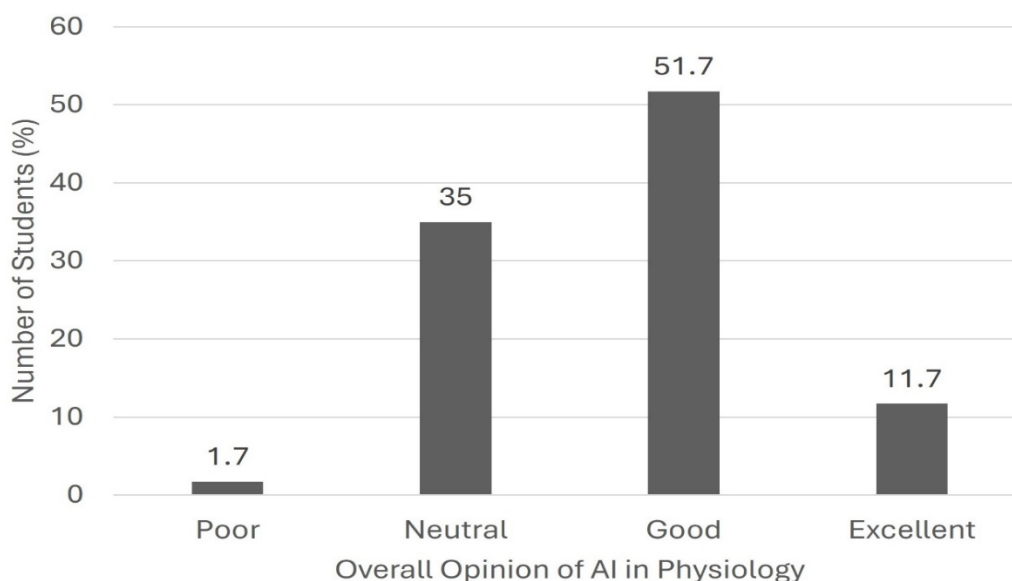
Overall opinion regarding AI use in Physiology was predominantly positive (Figure 4). Among the 60 students, 31 (51.7%) rated AI as good, 7 (11.7%) as excellent, 21 (35.0%) as neutral, and 1 (1.7%) as poor. The mean overall opinion score was 2.73 ± 0.69 , with a median of 3.0. Overall opinion did not differ significantly according to

learning method (Kruskal–Wallis $\chi^2(2) = 0.588$, $p = .745$).

Overall, 49 students (81.7%) recommended the use of AI in Physiology for future batches, whereas 11 (18.3%) did not.

The proportion recommending AI was 81.0% among students using traditional learning alone, 80.6% among those using traditional learning supplemented with AI, and 87.5% among those using AI-based learning.

The association between learning method and recommendation of AI was not statistically significant ($\chi^2(2) = 0.211$, $p = .900$; Fisher's exact test, $p = 1.000$).

**Figure 4: Overall opinion of and recommendation regarding AI use in Physiology.**

Overall opinion was graded as Poor = 1, Neutral = 2, Good = 3, and Excellent = 4. The mean overall opinion score was 2.73 ± 0.69 (median 3.0). The association between recommendation of AI use and learning method was not statistically significant ($\chi^2 = 0.211$, $df = 2$, $p = .900$; Fisher's exact test $p = 1.000$).

Correlation of previous academic performance and study time with current performance:

Previous IA performance demonstrated significant positive correlations with all current academic outcomes (Table 4, Figures 5, 6). Higher previous

IA performance was associated with higher Total FA scores (Spearman's $\rho = 0.599$, $p < .001$), Knows component scores ($\rho = 0.613$, $p < .001$), case-based scores ($\rho = 0.441$, $p < .001$), reasoning scores ($\rho = 0.481$, $p < .001$), and Knows-how scores ($\rho = 0.495$, $p < .001$). Reported study time was also positively correlated with Total FA score ($\rho = 0.333$, $p = .014$), Knows component score ($\rho = 0.406$, $p = .002$), and reasoning score ($\rho = 0.369$, $p = .006$). Correlations with case-based performance ($\rho = 0.126$, $p = .365$) and Knows-how performance ($\rho = 0.213$, $p = .122$) were not statistically significant.

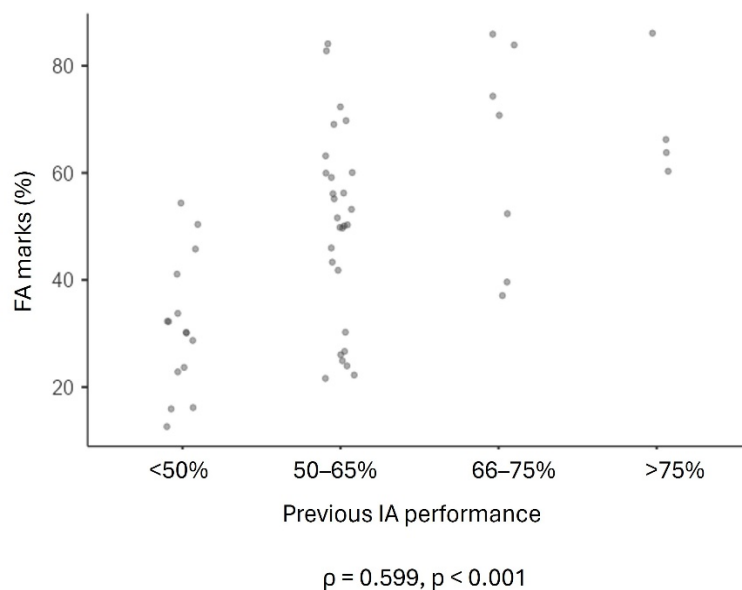


Figure 5: Association between previous internal assessment performance and subsequent formative assessment performance.

The jittered dot plot shows the distribution of total formative assessment (FA) scores across categories of previous internal assessment (IA) performance. A positive association was observed between previous IA performance and total FA score (Spearman's $\rho = 0.599, p < 0.001$), indicating that students with higher previous IA performance tended to achieve higher subsequent FA scores.

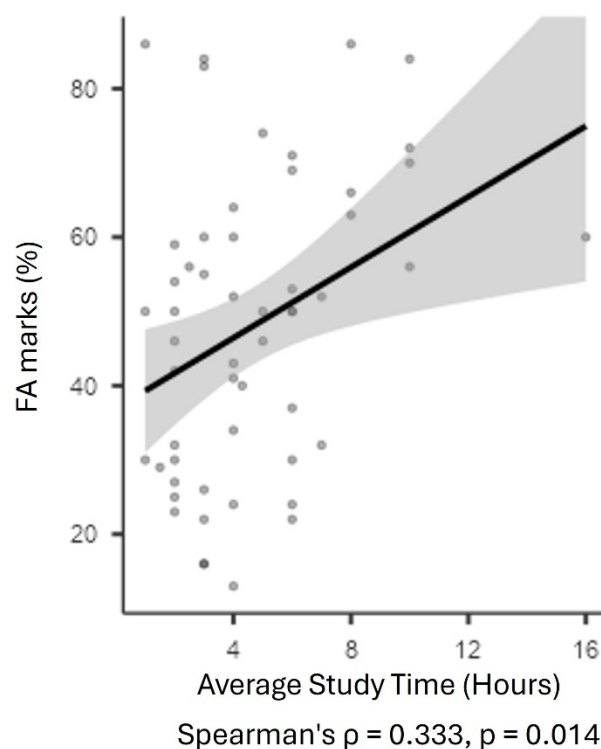


Figure 6: Association between average study time and academic performance.

The scatter plot demonstrates the relationship between reported average study time and total formative assessment score. Spearman's rank correlation coefficient and corresponding p-value are presented in the figure.

Table 4: Correlations between previous AI performance, study time, academic performance and AI-related perceptions.

Variables	Spearman's ρ	p value
Previous FA performance		
↔ Total FA score	0.599	<0.001
↔ Knows component	0.613	<0.001
↔ Case-based marks	0.441	<0.001
↔ Reasoning marks	0.481	<0.001
↔ Knows-how component	0.495	<0.001
Average study time		
↔ Total FA score	0.333	0.014
↔ Knows component	0.406	0.002
↔ Case-based marks	0.126	0.365
↔ Reasoning marks	0.369	0.006
↔ Knows-how component	0.213	0.122
AI-related perceptions		
Educational benefits ↔ Responsible AI use	0.698	<0.001
Educational benefits ↔ AI concern	-0.449	<0.001
Responsible AI use ↔ AI concern	-0.542	<0.001
Educational benefits ↔ Overall opinion	0.641	<0.001
Responsible AI use ↔ Overall opinion	0.609	<0.001
AI concern ↔ Overall opinion	-0.260	0.045

Correlations were assessed using Spearman's rank correlation coefficient. Positive coefficients indicate that higher values of one variable were associated with higher values of the other, whereas negative coefficients indicate an inverse association. Statistical significance was set at $p < 0.05$.

Correlations among AI-related perceptions:

Perceived educational benefits showed a strong positive correlation with responsible AI use ($\rho = 0.698$, $p < .001$) and overall opinion of AI ($\rho = 0.641$, $p < .001$), while showing moderate inverse correlations with AI-related concerns ($\rho = -0.449$, $p < .001$). Responsible AI use was positively correlated with overall opinion ($\rho = 0.609$, $p < .001$) and inversely correlated with AI-related concerns ($\rho = -0.542$, $p < .001$). AI-related concerns showed a weak inverse correlation with overall opinion ($\rho = -0.260$, $p = .045$).

Perceived educational benefits showed weak inverse correlations with Total FA score (Spearman's $\rho = -0.299$, $p = .028$) and the Knows component ($\rho = -0.320$, $p = .018$). No significant associations were observed between perceived

educational benefits and case-based performance ($\rho = -0.180$, $p = .193$), reasoning performance ($\rho = -0.268$, $p = .050$), or Knows-how performance ($\rho = -0.223$, $p = .105$).

Perceptions regarding responsible AI use were not significantly correlated with Total FA score or any individual academic performance component. Similarly, study time was not significantly associated with educational-benefit perceptions ($\rho = -0.071$, $p = .591$), responsible AI use ($\rho = -0.036$, $p = .782$), AI-related concerns ($\rho = 0.004$, $p = .976$), or overall opinion of AI ($\rho = 0.046$, $p = .727$). (Table 4)

Reliability of the AI perception scales: The six-item Educational Benefits scale demonstrated excellent internal consistency, with Cronbach's $\alpha = 0.932$ and McDonald's $\omega = 0.932$. The three-item Responsible AI Use scale demonstrated acceptable internal consistency (Cronbach's $\alpha = 0.723$; McDonald's $\omega = 0.779$), while the three-item AI Concerns scale demonstrated good internal consistency (Cronbach's $\alpha = 0.870$; McDonald's $\omega = 0.881$) (Table 5).

Table 5: Reliability of the AI perception scales.

Scale	Items	Cronbach's α	McDonald's ω
Educational benefits	Understanding; Simplified explanations; Retention; Exam preparation; SDL; Student engagement	0.723	0.779
Responsible AI use	Encouraged; Supplement/not replacement; Faculty guidance	0.723	0.779
AI concerns	Overdependence; Accuracy concern; Superficial learning	0.870	0.881

Cronbach's α and McDonald's ω represent measures of internal consistency reliability. The negatively worded concern items were reverse-coded before calculation of the corresponding composite scores.

Discussion

The present study explored the relationship between AI-assisted learning practices, conventional learning approaches, study behaviour, and academic performance among undergraduate medical students preparing for an Endocrine Physiology formative assessment. An important finding was the positive association between study time and academic performance. Students who reported spending more time studying tended to achieve higher total FA scores, with a statistically significant positive correlation between average study time and total FA performance ($\rho = 0.333$, $p = 0.014$).

The association was particularly evident for the knowledge-based "Knows" component ($\rho = 0.406$, $p = 0.002$), while a significant positive correlation was also observed with reasoning marks ($\rho = 0.369$, $p = 0.006$). These findings suggest that the amount of time students devote to academic preparation may remain an important determinant of performance even in an increasingly AI-supported learning environment. AI may provide rapid access to information, but effective learning still depends on adequate engagement with the subject matter.

The positive association between study time and reasoning performance is noteworthy. Physiology learning requires integration of concepts and application of basic principles rather than simple memorisation. AI tools can potentially support this process by providing explanations, examples, alternative ways of presenting concepts, and opportunities for question-based interaction. Recent reviews have found that generative AI is commonly used for inquiry and practice, suggesting that these tools may complement self-directed learning activities [1]. However, the present study cannot determine whether AI use itself improved reasoning ability; the observed association between study time and reasoning performance may reflect greater overall academic engagement.

The findings should also be interpreted in the context of the rapidly expanding use of AI among medical students. Generative AI can provide immediate explanations and personalised responses, making it attractive for students who are seeking clarification outside scheduled teaching sessions. At the same time, the reliability of generated information remains an important concern. Previous studies have demonstrated that although AI systems can perform well in medical examinations, their performance is not uniformly

reliable, particularly when questions require complex reasoning or interpretation [5,6]. Therefore, students should be encouraged to use AI as a supplementary learning resource while verifying information against authoritative textbooks, lecture material, and faculty guidance. An important educational implication of the present findings is that AI use should not be equated with effective learning. The amount of time spent engaging with learning material showed a measurable relationship with performance, whereas simply having access to AI does not necessarily ensure deeper understanding. Recent evidence from controlled studies similarly suggests that the educational advantage of generative AI over conventional approaches is not consistently demonstrated across knowledge outcomes [4]. This supports an approach in which AI is integrated into existing learning strategies rather than replacing textbooks, classroom teaching, or faculty interaction.

The coexistence of AI-based and traditional learning approaches may therefore be particularly relevant in undergraduate physiology. Traditional resources provide structured and curriculum-aligned content, whereas AI tools can potentially offer additional explanation and personalised clarification. A combined approach may allow students to benefit from the accessibility of AI while retaining the accuracy and structure of established educational resources. Current evidence increasingly supports this complementary role rather than unrestricted replacement of conventional learning methods [7].

The study also has implications for medical educators. Rather than discouraging students from using AI, educators may need to provide guidance on how AI should be used effectively and critically. Students should be taught to verify AI-generated information, recognise potential inaccuracies, ask appropriately constructed questions, and use AI to support rather than substitute their own reasoning. Systematic reviews of AI education have similarly emphasised the need for structured AI literacy and responsible integration into undergraduate medical education [2,8].

Strengths and Limitations: A strength of this study was the examination of AI use in relation to an actual formative assessment rather than relying solely on students' perceptions of academic benefit. The study also considered study time and separate cognitive components of performance, allowing a more detailed assessment of learning behaviour.

However, several limitations should be acknowledged. The study was conducted at a single institution with a relatively small sample of 60 students, limiting the generalisability of the findings. AI use and study time were self-reported

and may therefore be affected by recall or reporting bias. The cross-sectional design also prevents determination of causality. Other factors, including prior academic ability, motivation, preferred learning style, and access to learning resources, may have influenced performance.

Conclusion

AI-based learning has become an emerging component of undergraduate medical education, but its educational value should be considered alongside students' broader learning behaviour.

In the present study, greater reported study time was significantly associated with higher total formative assessment performance as well as better knowledge and reasoning scores.

These findings suggest that effective engagement with learning material remains important irrespective of the learning technology used. AI may be most appropriately positioned as a supplementary tool that enhances explanation, exploration, and self-directed learning while conventional resources and faculty guidance continue to provide the foundation for accurate and meaningful learning.

Acknowledgement: We express our sincere gratitude to all the students and support staff whose participation was essential for this study. We also thank the management of our institution for permitting us to conduct this study.

References

1. Pham TD, Karunaratne N, Exintaris B, Liu D, Lay T, Yuriev E, et al. The impact of generative AI on health professional education: A systematic review in the context of student learning. *Med Educ*. 2025 Dec;59(12):1280–9. doi:10.1111/medu.15746 PubMed PMID: 40533396; PubMed Central PMCID: PMC12686775.
2. Gordon M, Daniel M, Ajiboye A, Uraiby H, Xu NY, Bartlett R, et al. A scoping review of artificial intelligence in medical education: BEME Guide No. 84. *Med Teach*. 2024 Apr;46(4):446–70. doi:10.1080/0142159X.2024.2314198 PubMed PMID: 38423127.
3. Tozsin A, Ucmak H, Soyuturk S, Aydin A, Gozen AS, Fahim MA, et al. The Role of Artificial Intelligence in Medical Education: A Systematic Review. *Surg Innov*. 2024 Aug;31(4):415–23. doi:10.1177/15533506241248239 PubMed PMID: 38632898.
4. Li J, Yin K, Wang Y, Jiang X, Chen D. Effectiveness of generative artificial intelligence-based teaching versus traditional teaching methods in medical education: a meta-analysis of randomized controlled trials. *BMC Med Educ*. 2025 Aug 19;25:1175. doi:10.1186/s12909-025-07750-2 PubMed PMID: 40830476; PubMed Central PMCID: PMC12362899.
5. Soulage CO, Van Coppenolle F, Guebregziabher F. The conversational AI “ChatGPT” outperforms medical students on a physiology university examination. *Adv Physiol Educ*. 2024 Dec 1;48(4):677–84. doi:10.1152/advan.00181.2023 PubMed PMID: 38991037.
6. Sumbal A, Sumbal R, Amir A. Can ChatGPT-3.5 Pass a Medical Exam? A Systematic Review of ChatGPT’s Performance in Academic Testing. *J Med Educ Curric Dev*. 2024;11:23821205241238641. doi:10.1177/23821205241238641 PubMed PMID: 38487300; PubMed Central PMCID: PMC10938614.
7. Wang C, Sun N, Pan X, Liu P, Ma Z, Wang W, et al. The impact of integrating generative artificial intelligence into medical education on short-term learning outcomes: a systematic review and meta-analysis of randomized controlled trials. *BMC Med Educ*. 2026 Apr 30;26(1):983. doi:10.1186/s12909-026-09320-6 PubMed PMID: 42063033; PubMed Central PMCID: PMC13274167.
8. Pupic N, Ghaffari-Zadeh A, Hu R, Singla R, Darras K, Karwowska A, et al. An evidence-based approach to artificial intelligence education for medical students: A systematic review. *PLOS Digit Health*. 2023 Nov;2(11):e0000255. doi:10.1371/journal.pdig.0000255 PubMed PMID: 38011214; PubMed Central PMCID: PMC10681314.