ORIGINAL ARTICLE

pISSN 1226-6396, eISSN 2234-4942
J of Oriental Neuropsychiatry 2025;36(3):107-119
https://doi.org/10.7231/jon.2025.36.3.107

JON Journal of Oriental Neuropsychiatry

Exploring the Educational Potential of a Prototype AI System for Automated Research Note Generation: A Faculty Survey

Chan-Young Kwon[*,†]

*Department of Oriental Neuropsychiatry, Dong-eui University College of Korean Medicine, †Anti-Aging Research Center, Dong-eui University

Received: August 31, 2025
Revised: September 7, 2025
Accepted: September 8, 2025

Correspondence to
Chan-Young Kwon
Department of Oriental Neuropsychiatry, Dong-eui University College of Korean Medicine, 52-57 Yangjeong-ro, Busanjin-gu, Busan, Korea.
Tel: +82-51-850-8808
Fax: +82-51-867-5162
E-mail: beanalogue@naver.com

Acknowledgement
This work was partly supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP)-Innovative Human Resource Development for Local Intellectualization program grant funded by the Korea government (MSIT) (IITP-2025-RS-2020-II201791, 50%) and Dong-eui University Grant (202500980001, 50%).

Objectives: Traditional evidence-based medicine (EBM) education often presents research as a finished product, which can obscure the dynamic inquiry process and limit students' understanding of longitudinal decision-making. Because authentic research notes are rarely accessible, generative artificial intelligence (AI) can simulate the thought processes of researchers. This study evaluated the educational potential of a novel AI system that generates simulated research notes based on academic papers.

Methods: A cross-sectional survey was conducted with 16 EBM faculty members from Korean medicine colleges. Participants engaged in a hands-on evaluation of a web-based system (utilizing Google Gemini 1.5 Flash API) that generates simulated research notes. A mixed-methods questionnaire comprising 23 Likert-scale items across six domains and six open-ended questions was used for the evaluation.

Results: Participants provided moderately positive ratings, with 'Educational Effectiveness' (mean 3.97±0.84) and 'Overall Assessment' (3.97±0.96) receiving the highest scores. The system's ability to portray researcher concerns realistically was also well-rated (4.19±0.83), while 'Practical Applicability' received the lowest score (3.70±1.01). Qualitative analysis confirmed the system's educational novelty but highlighted significant concerns regarding AI reliability, including factual inaccuracies, and generated suggestions for practical implementation.

Conclusions: This proof-of-concept study demonstrates that AI-generated research note simulations address a gap in process-oriented EBM learning. Faculty acknowledged its conceptual value for teaching longitudinal research thinking; however, substantial technical reliability and implementation challenges must be addressed before it can be utilized by students. While technical limitations may improve with advancements in AI, specific educational challenges require focused pedagogical research. Future efforts should prioritize content accuracy and instructional design for effective classroom integration.

Key Words: Generative artificial intelligence, Evidence-based medicine, Korean medicine education, Research methodology, Educational technology.

Copyright © 2025 by The Korean Society of Oriental Neuropsychiatry. All rights reserved.
This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

I. INTRODUCTION

Evidence-based medicine (EBM), which integrates clinical expertise, patient values, and the best available research evidence, has become a foundational paradigm in modern medical decision-making[1,2]. Consequently, EBM education has been established as an essential component of the medical curriculum. However, the optimal strategies for teaching EBM to undergraduate students are still debated, as prior studies have reported inconsistent outcomes[2-4]. The predominant pedagogical approach focuses on the critical appraisal of published research articles[4]. Although this approach is essential for cultivating critical appraisal skills, it presents a major limitation: research is treated as a static final product, obscuring the dynamic and complex processes through which knowledge is generated[5-7]. For learners, the intellectual journey of a study, from the initial research question and methodological debates to the inevitable setbacks and discoveries, often remains a "black box."

This focus on the final product, however, can lead to a superficial understanding of scientific inquiry, leaving students without a deep, contextual grasp of the research process itself. A pedagogical shift is therefore necessary, moving beyond the static analysis of a finished paper toward a more holistic, process-oriented longitudinal perspective. This approach does not require extending the curriculum across multiple years; instead, it focuses on reconstructing the complete narrative of a single research project. Following a study's journey from its initial question to its conclusion allows learners to appreciate the context and complexities of scientific work, a skill essential for navigating the ambiguities of clinical reasoning[8]. As clinical decision-making becomes more complex, the need for such process-oriented learning grows ever more critical.

The recent emergence of generative artificial intelligence (AI) presents an unprecedented opportunity to realize this educational model. In medical education, generative AI is already being leveraged to create adaptive learning platforms, virtual patient scenarios, and interactive simulations that enhance personalized learning[9]. However, the application of these powerful tools has largely concentrated on clinical skills and student-teacher interactions[9]. To date, there has been a notable absence of AI tools designed to simulate the intricate cognitive and decision-making processes of a researcher. If AI could systematically deconstruct a final publication back into its procedural steps, recreating the researcher's probable thought process, challenges, and decisions, it could offer an innovative solution to the limitations of traditional EBM education. Such a tool would be particularly valuable in fields like Korean medicine (KM), where research often involves complex multicomponent interventions, individualized treatments, and the integration of traditional theory with modern scientific methodology[10].

Therefore, this study presents and evaluates an educational system that applies generative AI to reverse-engineer (i.e., reconstruct) published academic papers into a series of simulated research notes. The primary objective is to assess the educational utility and potential of this system from the perspective of expert educators in EBM and research methodology. Through this evaluation, this study aims to explore the value, limitations, and future directions of a process-oriented, AI-driven approach to EBM education.

II. METHODS

1. The educational system

The web-based educational tool evaluated in this study was developed on the Google Apps Script plat-

form, utilizing the Google Gemini 1.5 Flash API to automatically generate research notes. The system is designed to take a published academic paper as input and simulate the thought processes and decision-making steps a researcher might have experienced during the actual study. This process results in a series of educational research notes that reconstruct the study's progression over time (Fig. 1).

The system's architecture consists of three core components: a paper analysis engine, a research note generation engine, and an educational support module. The analysis engine parses uploaded research papers (in PDF or Word format) to extract key elements such as the research objective, methodology, results, and conclusion. It uses the Google Drive API for temporary file storage and employs optical character recognition and natural language processing to create a structured representation of the paper's content.

Based on this structured information, the generation engine applies specialized algorithms tailored to different research designs (e.g., literature reviews, clinical intervention studies). For instance, when processing a clinical trial, it simulates the stages of protocol development, institutional review board (IRB) approval, participant recruitment, intervention, and data analysis. Each generated note is designed to reflect the information realistically available to a researcher at that specific point in time, intentionally omitting future results or conclusions to provide an authentic, process-oriented learning experience (Fig. 2).

Finally, the educational support module automatically generates discussion questions for instructors based on the research notes, categorized into three difficulty levels (bachelor's, master's, and doctoral). It also includes a feature for providing AI-based feedback on student responses, aiming to reduce the instructor's workload while offering individualized learning support (Fig. 3).

2. Study design and participants

This study employed a cross-sectional, descriptive anonymous survey design to evaluate the educational utility of the system. Data were collected via an on-

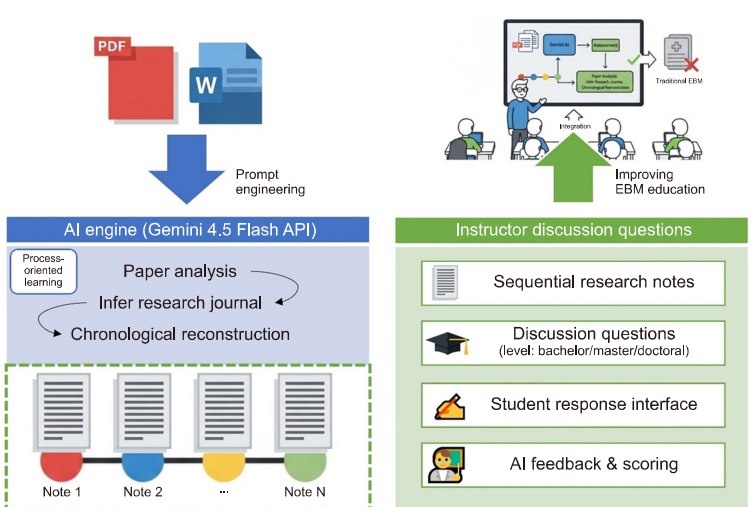


Fig. 1. Overview of the AI-powered process-oriented learning system for enhanced EBM education. This figure illustrates the workflow of the AI system. It analyzes an uploaded research paper using Google Gemini 1.5 Flash API to chronologically reconstruct the study into simulated research notes. It then generates educational outputs, including categorized discussion questions and automated AI feedback, to support a process-oriented approach to EBM education.

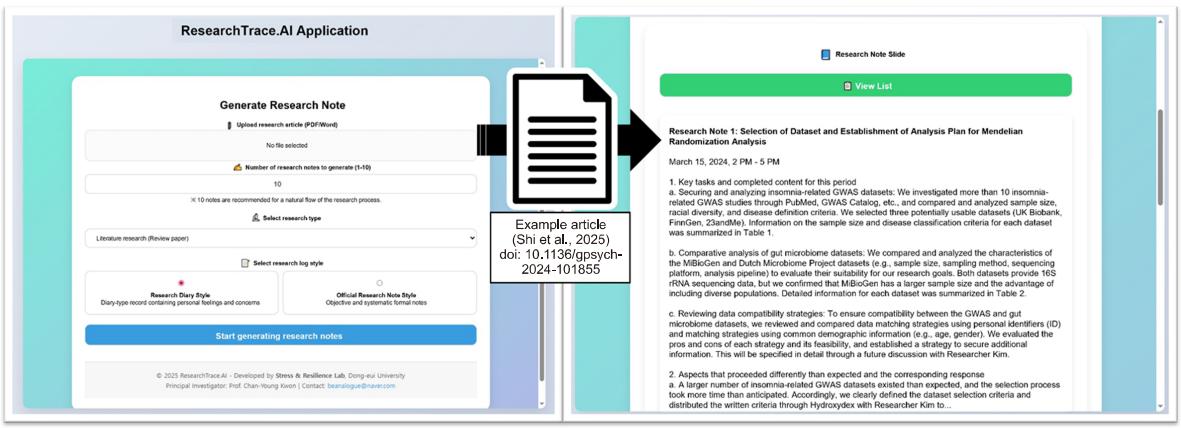


Fig. 2. Research note generation interface and output example.
This figure illustrates the process of generating research notes based on a user's uploaded research results. It shows how the system takes input parameters (e.g., number of notes, research type) and applies an algorithm to create a series of notes that simulate a realistic, process-oriented research journey. The generated notes are designed to reflect the information available at each stage, mirroring the actual research process.

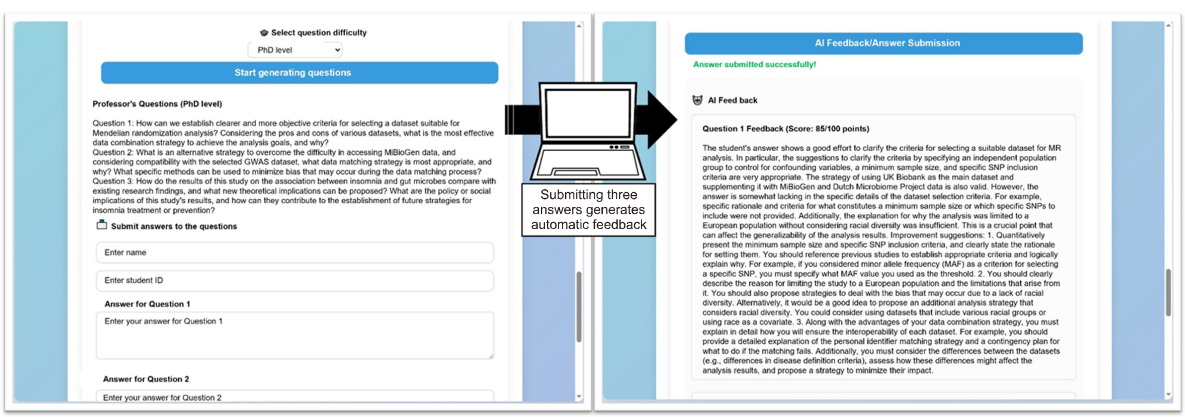


Fig. 3. Instructor question generation and AI-powered student feedback module.
This figure demonstrates the educational support functionality of the system. It shows how the platform automatically generates discussion questions for instructors at different academic levels (e.g., doctoral) and provides a dedicated interface for students to submit their answers. The system also includes an AI-based feedback module that offers personalized feedback on student responses, streamlining the teaching process and enhancing learning.

line survey conducted over one month, from July 31 to August 31, 2025. The target population consisted of full-time faculty members at KM colleges and graduate schools across the nation with experience teaching EBM or research methodology. Inclusion criteria were: (1) current full-time faculty status at a college of KM; (2) experience conducting at least one formal course or informal educational activity (e.g., seminar, workshop) related to EBM or research methodology within the last three years; and (3) voluntary consent to participate. Part-time lecturers and faculty with no relevant teaching experience were excluded.

Due to the specialized nature of the target population and the lack of a formal registry, participants were recruited using a snowball sampling strategy[11]. Initial contact was made with faculty at the researchers' university and affiliated institutions. These initial participants were then asked to recommend other eligible colleagues from their professional

networks. A target sample size of approximately 20 participants was set, considering the total number of KM colleges in the country. To encourage participation, a mobile gift voucher (equivalent to 10,000 KRW) was provided to each participant upon survey completion.

3. Survey instrument

As no standardized instrument existed for evaluating this type of educational tool, a new questionnaire was developed by the research team based on a review of relevant literature and expert opinion[12]. The survey consisted of three sections, as follows.

1) Respondent Demographics (6 items): Gender, age group, teaching experience, academic specialty, primary audience (e.g., undergraduate, graduate), and frequency of teaching EBM/research methodology.

2) Quantitative Evaluation (23 items): A 5-point Likert scale (1=strongly disagree, 5=strongly agree) was used to assess six domains, including Content Quality (5 items), Educational Effectiveness (4 items), System Usability (4 items), AI Content Reliability (4 items), Practical Applicability (4 items), and Overall Assessment (2 items).

3) Qualitative Feedback (6 open-ended questions): Questions explored perceived educational contributions, potential limitations, ethical considerations, feedback on specific features, and suggestions for improvement (Supplementary File 1).

4. Procedure

Participants were first instructed to interact directly with the educational system. They were asked to upload a research paper of their choosing, select the study type, and set the number of research notes to be generated. This hands-on approach allowed them to evaluate the system's performance on a topic relevant to their own teaching interests. After experiencing the system, participants completed the anonymous online survey to provide their feedback.

5. Ethical considerations

This study was approved by the Institutional Review Board of Dong-eui University Korean Medicine Hospital on July 25, 2025 (Approval No: DH-2025-09). All participants were provided with a detailed explanation of the study's purpose, procedures, potential risks and benefits, and data privacy measures before providing informed consent. All collected data were anonymized using research codes. Personally identifiable information was limited to mobile phone numbers for gift voucher distribution and was deleted immediately after disbursement.

6. Data analysis

Given the small sample size, the quantitative data were analyzed using descriptive statistics. Continuous variables from the Likert scale items were presented as means and standard deviations (SD). Categorical demographic data were summarized as frequencies and percentages. For the qualitative data from the open-ended questions, a thematic content analysis was conducted to identify and categorize key themes and opinions[13]. All statistical analyses were performed using Microsoft Excel 365 (Microsoft Corp., Redmond, WA, USA).

III. RESULTS

1. Respondent characteristics

A total of 20 potential participants were asked to access the system and participate in this survey, of which 16 members participated. A half of the participants were female (n=8, 50%) with the largest group being in their 40s (n=10, 62.5%). Regarding teaching experience, the majority had 5 years or less of experience (n=8, 50.0%), followed by 10～15 years (n=3, 18.75%), 5～10 years (n=3, 18.75%), and 15 years or

more (n=2, 12.5%). In terms of academic specialization, basic medicine had the highest representation (n=8, 50%), followed by clinical medicine (n=6, 37.5%) and medical humanities (n=2, 12.5%). Most participants taught undergraduate students, with 15 (93.75%) teaching clinical students and 7 (43.75%) teaching pre-clinical students. Ten participants (62.5%) also taught graduate students, while none reported teaching residents/trainees. Regarding EBM/research methodology teaching frequency over the past three years, 12 participants (75%) conducted regular coursework at least once annually, while 2 (12.5%) conducted regular coursework less than once annually, and 2 (12.5%) had no regular educational activities in the past three years. For informal educational activities, 9 participants (56.25%) engaged in activities such as seminars or workshops at least once yearly, 6 (37.5%) participated less than once annually, and 1 participant (6.25%) had no informal activities (Table 1).

2. Quantitative evaluation

The quantitative evaluation results across six domains are presented in Table 2. Overall, participants provided moderately positive ratings across most evaluation domains.

The highest-rated domain was Educational Effectiveness (3.97±0.84) and Overall Assessment (3.97±0.96), followed by System Usability (3.95±0.67). Content Quality received a mean score of 3.93±0.81, while AI Content Reliability scored 3.75±0.75. The Practical Applicability domain received the lowest mean score (3.70±1.01), with "immediate applicability to current courses" (Q24) receiving the lowest individual item score (3.44±1.09). The highest individual item scores were for "realistic portrayal of researcher concerns" (Q10: 4.19±0.83) and "intuitive and convenient interface" (Q16: 4.19±0.66). Other highly rated items included "improvement in research methodology un-

Table 1. Respondent Demographic Characteristics (N=16)

Characteristic	Category	n (%)
Gender	Male	8 (50)
	Female	8 (50)
Age group	30 s	4 (25)
	40 s	10 (62.5)
	50 s	2 (12.5)
Teaching experience	≤5 years	8 (50)
	5~10 years	3 (18.75)
	10~15 years	3 (18.75)
	≥15 years	2 (12.5)
Academic specialty	Basic medicine	8 (50)
	Clinical medicine	6 (37.5)
	Medical humanities	2 (12.5)
Primary teaching audience*	Pre-clinical students	7 (43.75)
	Clinical students	15 (93.75)
	Graduate students	10 (62.5)
	Residents/trainees	0 (0)
EBM/Research formal education freq. (last 3 years)	Regular coursework (≥1/year)	12 (75.0)
	Regular coursework (<1/year)	2 (12.5)
	No formal activities	2 (12.5)
EBM/Research informal education freq. (last 3 years)	Informal activities (≥1/year)	9 (56.25)
	Informal activities (<1/year)	6 (37.5)
	No informal activities	1 (6.25)

EBM: evidence-based medicine.
*Multiple selections allowed.

Table 2. Mean Scores by Evaluation Domain

Domain	Mean±SD	Range
A. Content quality (5 items)	3.93±0.81	1.60～5.00
Q7. Research notes captured key paper content	3.69±1.40	1～5
Q8. Natural progression of research process	4.00±0.82	2～5
Q9. Appropriate yet understandable academic expression	3.94±0.93	1～5
Q10. Realistic portrayal of researcher concerns	4.19±0.83	2～5
Q11. Educational utility of style options	3.81±0.91	2～5
B. Educational effectiveness (4 items)	3.97±0.84	1.50～5.00
Q12. Improvement in research methodology understanding	4.06±1.06	1～5
Q13. Effective tool for EBM education goals	3.88±0.81	2～5
Q14. Contribution to critical thinking development	4.00±0.89	2～5
Q15. Appropriate explanation for student level	3.94±0.93	1～5
C. System usability (4 items)	3.95±0.67	2.50～5.00
Q16. Intuitive and convenient interface	4.19±0.66	3～5
Q17. Suitability of instructor discussion questions	4.06±0.68	3～5
Q18. Usefulness of student response collection	3.75±0.93	2～5
Q19. Overall satisfactory user experience	3.81±1.11	1～5
D. AI content reliability (4 items)	3.75±0.75	2.00～5.00
Q20. Logical and consistent generated content	3.69±0.95	2～5
Q21. Sufficient quality for educational materials	3.81±0.98	2～5
Q22. Quality of automated student response evaluation	3.69±1.20	2～5
Q23. Applicability across various academic fields	3.81±0.75	2～5
E. Practical applicability (4 items)	3.70±1.01	1.25～5.00
Q24. Immediate applicability to current courses	3.44±1.09	1～5
Q25. Efficient reduction of class preparation time	3.63±0.96	1～5
Q26. Effective integration with existing methods	3.94±1.12	1～5
Q27. Willingness to recommend to colleagues	3.81±1.11	1～5
F. Overall assessment (2 items)	3.97±0.96	1.00～5.00
Q28. High educational value system overall	4.06±0.93	1～5
Q29. Contribution to EBM education improvement	3.88±1.02	1～5

AI: artificial intelligence, EBM: evidence-based medicine, SD: standard deviation.

derstanding" (Q12: 4.06±1.06), "suitability of instructor discussion questions" (Q17: 4.06±0.68), and "high educational value system overall" (Q28: 4.06±0.93). Notably, there was considerable variability in responses, with one participant providing consistently low scores (1～2 points) across most domains, while others provided predominantly high scores (4～5 points).

3. Qualitative findings

The qualitative analysis of open-ended responses revealed three major themes regarding participants' perspectives on the educational system.

1) Perceived educational novelty and potential

Participants acknowledged the system's innovative approach to providing a process-oriented learning experience. One clinical medicine faculty member noted, *"I was surprised by how well it incorporated the concerns researchers would likely have. The sequence and slight differences are similar to the actual concerns I had"* (Participant A). A medical humanities faculty member emphasized the novelty of this approach, stating, *"Research process education generally requires direct research experience, but experiencing it in advance through research notes is quite novel"* (Participant C). The system was seen as a valuable tool for research methodology education, offering students a chance to *"understand what kind of process research would involve, like a simulation in advance"* (Participant E). Another participant com-

mented that the tool could guide *"students who have not yet conducted research... step-by-step while reducing fear"* (Participant L) (Supplementary File 2).

2) Concerns regarding AI content and limitations

Despite the positive reception of its concept, participants reported some issues with the system's output and performance. The most critical concern was the accuracy of the AI-generated content. One basic medicine faculty member reported significant content distortions: *"The research notes generated by the current system do not properly understand the paper and write content different from the paper, distorting results (recording opposite results or stating that a systematic review paper marked as IRB-exempt needed IRB preparation)"* (Participant J). This participant further noted that when a general LLM was given the same task, *"there are fewer result distortions or interpretation errors"*. Technical limitations were also a recurring theme. Participants mentioned that *"research note generation took longer than expected (over 10 minutes)"* (Participant F) and that *"the system is unstable. It generated completely different content from the uploaded paper"* (Participant I). Another participant reinforced these concerns, noting that the generation speed was *"on the slow side"* and that there was *"a slight tendency for the AI to arbitrarily make up some of the content"* (Participant P). Minor inconsistencies, such as incorrect numbers of subjects or submitted journals, were also noted, though one participant found them reasonably acceptable: *"As with any AI, the details are somewhat different from the paper (the number of subjects, the journal to which they are submitted, etc.) but I think it can be sufficiently considered"* (Participant A) (Supplementary File 2).

3) Suggestions for practical implementation and improvement

Beyond identifying limitations, the participants offered extensive constructive feedback aimed at bridging the gap between the system's innovative concept and its current practical application. The suggestions ranged from immediate technical fixes to fundamental questions about the system's pedagogical framework. On a practical level, participants highlighted several ways to improve the user experience and the simulation's authenticity. A common point of friction was the system's text-heavy interface, which a participant noted made *"readability difficult"* (Participant H). To make the tool more versatile for both instructors and students, users requested basic functionalities like the ability to have *"research note generation results... downloaded as files"* (Participant A) and ensuring that *"tablets work smoothly"* (Participant M). Suggestions also aimed to enrich the educational content itself. To move beyond plain text, participants envisioned *"adding functions for generating tables, organizational charts, and task execution stages"* (Participant C), with another suggesting that if *"charts or tables were utilized, it would become more intuitive and increase understanding"* (Participant P). To make the simulation more immersive, another suggested allowing users to input the *"names of first authors, corresponding authors, and research assistants separately so they can be automatically used in research notes"* (Participant L).

The feedback on the educational features revealed a desire for more refined and targeted tools. The instructor question-generator, while appreciated in concept, was seen as needing improvement. One clinical medicine faculty member suggested that a structure based on the research process, *"'research questions,' 'major research methodology,' 'interpreta-*

tion,' and 'conclusion drawing'", would be more effective than the current academic levels (Participant E). Another faculty member recommended providing *"example answers along with questions"* (Participant G) to better guide instructors. The AI-powered feedback module received specific criticism, with one participant noting that its scoring system required *"much improvement,"* particularly because *"a high score is measured even when there is no answer content"* (Participant P).

Ultimately, these specific suggestions pointed to a larger, more fundamental issue: the need for a clearer pedagogical purpose. Even with technical improvements, some faculty were unsure how to integrate the system into their teaching. This challenge was best captured by one participant who asked, *"I don't know how to use it in class... if the system proceeds in its own way and doesn't match the direction of my class, it would be difficult to use"* (Participant B). For the system to be successfully adopted, he concluded, *"specific (detailed) purposes are needed, and learning objectives for specific topics should come first"* (Participant B) (Supplementary File 2).

IV. DISCUSSION

To our knowledge, this study is the first formal evaluation of a generative AI-based educational system designed to simulate the research process through automated research note generation. While recent scoping reviews have identified various AI applications in medical education including intelligent tutoring systems, virtual patients, and adaptive learning platforms, none have specifically addressed the simulation of longitudinal research processes[14,15]. As a proof-of-concept evaluation, the findings provide initial insights into the feasibility and potential of using AI to address a fundamental limitation in current EBM education, the inability to observe the longitudinal thought processes of researchers during actual studies[5-7].

The generally favorable ratings across most domains suggest that the underlying educational concept—using simulated research notes to provide process-oriented learning experiences—was well received by medical educators. The highest individual item score for "realistic portrayal of researcher concerns" (4.19±0.83) indicates that participants recognized the educational value of exposing students to the temporal progression of research decision-making. This approach addresses a long-standing pedagogical challenge in EBM education. Traditional methods rely heavily on retrospective analysis of published papers, which inevitably obscure the uncertainties, setbacks, and iterative decisions that characterize actual research processes[5-7]. Since authentic research notes from published studies are rarely available to educators, AI-generated simulation represents a potentially viable alternative for providing students with longitudinal exposure to research thinking. The qualitative findings support this conceptual framework, with participants describing the approach as "*quite novel*" and potentially valuable for students lacking research experience. One participant noted that "*research process education generally requires direct research experience, but experiencing it in advance through research notes is quite novel*," highlighting the unique educational niche this approach could fill.

The content reliability issues identified in this study must be understood within the context of the specific AI model employed (Gemini 1.5 Flash API) and the current state of generative AI technology. The AI Content Reliability domain scored moderately (3.75±0.75), but qualitative feedback revealed significant problems including content distortions and factual inaccuracies. Multiple participants reported instances where the system generated opposite re-

search results or inappropriate procedural recommendations. These technical limitations likely reflect the current capabilities of the specific AI model used rather than fundamental flaws in the educational concept itself. Given the rapid pace of large language model development, many of these accuracy issues may be addressed through future technological improvements rather than changes in research design[16]. The reported superiority of general-purpose language models in avoiding content distortions (*"When I use a commonly available LLM to upload the same paper and ask it to write 10 days' worth of research notes, there are not many result distortions or interpretation errors"*) may indicate that newer, more sophisticated AI systems could perform substantially better.

However, certain challenges will require targeted educational research solutions regardless of technological improvements. Key challenges include establishing pedagogical frameworks that integrate AI-generated content with existing EBM curricula and developing instructor training protocols to support process-oriented discussions. Equally important is the creation of assessment methods that evaluate students' understanding of research processes rather than solely their final outcomes. Additionally, educational researchers must address content validation challenges specific to academic contexts, determining how to verify that simulated research notes accurately reflect realistic decision-making patterns, establishing quality control standards for educational AI applications, and developing methods to customize content for different learning levels and institutional contexts. These education-specific challenges require domain expertise in medical pedagogy and cannot be resolved simply through general advances in AI capabilities, as they involve fundamental questions about how students best learn complex, longitudinal thinking processes that are distinct from the technical problems of content generation accuracy[17].

The study design appropriately focused on expert educator evaluation before student deployment, providing essential formative feedback for system development. The Practical Applicability domain's lower scores (3.70±1.01) and qualitative concerns about classroom integration reflect the expected limitations of a prototype system rather than insurmountable barriers to the educational approach. Faculty requests for enhanced functionality (e.g., downloadable outputs, tablet compatibility, visual elements) represent straightforward development priorities that could significantly improve usability. More substantive pedagogical concerns, such as the need for clearer learning objectives and structured implementation guides, highlight the importance of comprehensive educational design beyond the technological components. The considerable variability in participant responses, including one consistently negative evaluation, suggests that successful implementation will require addressing diverse faculty perspectives and comfort levels with educational technology. This heterogeneity is typical in educational innovation adoption and emphasizes the need for flexible implementation approaches[18]. Research on educational technology adoption among faculty has demonstrated that early adopters and late majority groups differ significantly in their approaches and needs, suggesting that effective implementation strategies must account for these individual differences rather than adopting a one-size-fits-all approach[19].

This study's implications extend beyond individual classroom applications to the broader medical education ecosystem. The demonstration that AI can simulate authentic research processes suggests potential applications in continuing medical education, where practicing physicians need to stay current with evolving research methodologies. Healthcare institu-

tions investing in evidence-based practice initiatives may find process-oriented AI tools valuable for training clinical staff in research literacy. Additionally, the concept of reverse-engineering published studies into learning experiences could inform medical journal publishers and professional societies in developing educational supplements to research publications. Moreover, this research raises important questions about standardization and quality control in AI-assisted medical education. Traditional assessment methods in medical education may be insufficient for measuring the complex thinking processes that this type of process-oriented simulation aims to develop. Furthermore, as AI tools become more prevalent in medical education[14,15,17], regulatory bodies may need to establish guidelines for validating educational AI applications, particularly those that simulate professional thinking processes.

Several limitations constrain the interpretation of these findings. The small sample size (n=16) and convenience sampling approach limit generalizability, particularly given recruitment through professional networks. The evaluation represents initial impressions of a prototype system rather than sustained use or measured learning outcomes. The absence of student perspectives represents a significant limitation, as educator perceptions may not predict actual learning effectiveness. The focus on KM faculty, while appropriate for initial evaluation, raises questions about applicability across different medical disciplines. More critically, this study evaluated a specific technical implementation rather than the broader educational concept. The technology-dependent limitations identified here may not reflect the ultimate feasibility of AI-assisted process-oriented learning as AI capabilities continue advancing. Furthermore, the survey instrument combined questions assessing the educational concept with those evaluating the system's technical performance. This design means that participants' opinions on one aspect could have influenced their evaluation of another. For instance, technical frustrations with the prototype, such as slow generation speed or instability, may have led to a more critical assessment of the underlying educational methodology. Conversely, enthusiasm for the novel teaching approach might have resulted in a more lenient evaluation of the system's technical flaws. Finally, it is worth noting a limitation in the survey's design itself, where the conceptual overlap between domains like 'Content Quality' and 'AI Content Reliability' could have led to ambiguity in how participants responded. This lack of clear distinction may have influenced the consistency of our findings.

Future research should distinguish between technology development and educational research priorities. Immediate technical improvements (e.g., enhanced accuracy, better content validation, improved user interfaces) may emerge naturally from general AI advancement[16]. Educational research should focus on areas requiring domain-specific investigation: optimal pedagogical integration strategies, instructor training approaches, student learning outcome measurement, and comparative effectiveness studies. Longitudinal evaluation of actual classroom implementation, preferably using more advanced AI systems, represents a critical next step. Such studies should include student learning outcomes, comparative analyses with traditional teaching methods, and assessment of long-term impact on research methodology understanding. The development of standardized evaluation frameworks for educational AI systems would facilitate more rigorous assessment of future iterations. Additionally, research into optimal content curation and validation processes for AI-generated educational materials could inform broader applications beyond research methodology training.

V. CONCLUSION

This study provides a proof-of-concept evaluation of a novel AI-based system designed to simulate the research process for EBM education. Faculty educators recognized the conceptual value of the system in addressing a key gap in traditional pedagogy: the lack of exposure to the longitudinal, process-oriented nature of scientific inquiry. The potential to provide students with a simulated experience of a researcher's journey was seen as a significant educational innovation. However, the evaluation also revealed critical barriers to practical implementation, most notably the significant concerns regarding the reliability and accuracy of the AI-generated content. While participants acknowledged the educational potential, the system's current technical instability and the need for a more defined pedagogical framework for classroom integration prevent its immediate adoption. Ultimately, these findings suggest that while AI-generated research note simulation is a promising approach to enrich EBM education, its successful deployment is contingent upon overcoming both technology-dependent limitations, which may improve with AI advancements, and education-specific challenges that require dedicated pedagogical research and instructional design.

CONFLICTS OF INTEREST

None.

SUPPLEMENTARY MATERIALS

Supplementary materials can be found via https://doi.org/10.7231/jon.2025.36.3.107

REFERENCES

1. Nielsen LD, Løwe MM, Mansilla F, Jørgensen RB, Ramachandran A, Noe BB, et al. Interventions, methods and outcome measures used in teaching evidence-based practice to healthcare students: an overview of systematic reviews. BMC Med Educ. 2024;24(1):306. https://doi.org/10.1186/s12909-024-05259-8
2. Bello JO, Grant P. A systematic review of the effectiveness of journal clubs in undergraduate medicine. Can Med Educ J. 2023;14(4):35-46. https://doi.org/10.36834/cmej.72758
3. Ahmadi SF, Baradaran HR, Ahmadi E. Effectiveness of teaching evidence-based medicine to undergraduate medical students: a BEME systematic review. Med Teach. 2015;37(1):21-30. https://doi.org/10.3109/0142159x.2014.971724
4. Tikkinen KAO, Guyatt GH. Understanding of research results, evidence summaries and their applicability-not critical appraisal-are core skills of medical curriculum. BMJ Evid Based Med. 2021;26(5):231-3.10. https://doi.org/1136/bmjebm-2020-111542
5. Halalau A, Holmes B, Rogers-Snyr A, Donisan T, Nielsen E, Cerqueira TL, et al. Evidence-based medicine curricula and barriers for physicians in training: a scoping review. Int J Med Educ. 2021;12:101-24. https://doi.org/10.5116/ijme.6097.ccc0
6. Mathieson K, Weemer M, Lipke L. Approaches to teaching evidence-based medicine in residency: a systematic review. Med Educ Online. 2025;30(1):2504467. https://doi.org/10.1080/10872981.2025.2504467
7. Simons M, Rapport F, Zurynski Y, Stoodley M, Cullis J, Davidson AS. Links between evidence-based medicine and shared decision-making in courses for doctors in training: a scoping review. BMJ Open. 2022;12(4):e057335. https://doi.org/10.1136/bmjopen-2021-057335
8. Norman G. Research in clinical reasoning: past history and current trends. Med Educ. 2005;39(4):418-27. https://doi.org/10.1111/j.1365-2929.2005.02127.x
9. Tian S, Jin Q, Yeganova L, Lai PT, Zhu Q, Chen X, et al. Opportunities and challenges for ChatGPT and large language models in biomedicine and health. Brief Bioinform. 2023;25(1):bbad493. https://doi.org/10.1093/bib/bbad493
10. Park HL, Lee HS, Shin BC, Liu JP, Shang Q, Yamashita H, et al. Traditional medicine in china, Korea, and Japan: a brief introduction and comparison. Evid Based Complement Alternat Med. 2012;2012:429103. https://doi.org/10.1155/2012/429103
11. Patton MQ. Qualitative research & evaluation methods: integrating theory and practice. 4th ed. Thousand Oaks (CA): SAGE Publications, Inc.; 2015. 832 p.
12. Artino AR Jr, La Rochelle JS, Dezee KJ, Gehlbach H. Developing questionnaires for educational research: AMEE

Guide No. 87. Med Teach. 2014;36(6):463-74. https://doi.org/10.3109/0142159x.2014.889814
13. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res. 2005;15(9):1277-88. https://doi.org/10.1177/1049732305276687
14. 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;46(4):446-70. https://doi.org/10.1080/0142159x.2024.2314198
15. Sapci AH, Sapci HA. Artificial intelligence education and tools for medical and health informatics students: systematic review. JMIR Med Educ. 2020;6(1):e19285. https://doi.org/10.2196/19285
16. Wang L, Li J, Zhuang B, Huang S, Fang M, Wang C, et al. Accuracy of large language models when answering clinical research questions: systematic review and network meta-analysis. J Med Internet Res. 2025;27:e64486. https://doi.org/10.2196/64486
17. Saroha S. Artificial intelligence in medical education: promise, pitfalls, and practical pathways. Adv Med Educ Pract. 2025;16:1039-46. https://doi.org/10.2147/amep.S523255
18. Sahin I. Detailed review of Rogers' diffusion of innovations theory and educational technology-related studies based on Rogers' theory. Turkish Online Journal of Educational Technology-TOJET. 2006;5(2):14-23.
19. Romano ML. School of education faculty and their use of technology: are early adopters and late majority faculty different? [dissertation]. New York: Pace University; 2010 [cited 2025 Aug 31].

Supplementary File 1. Questionnaire

Part 1: Respondent Demographics

(Items Q1-Q6 covered respondent's gender, age group, teaching experience, academic specialty, primary teaching audience, and frequency of teaching EBM/research methodology.)

Part 2: Quantitative Evaluation Items

(5-point Likert scale: 1=Strongly Disagree, 5=Strongly Agree)

A. Content Quality

- **Q7.** The generated research notes effectively capture the core content of the paper.
- **Q8.** The progression of the research process is naturally expressed in each entry.
- **Q9.** The academic expressions used are appropriate yet easy to understand.
- **Q10.** The researcher's concerns and decision-making processes are portrayed realistically.
- **Q11.** The option to choose between a 'diary style' and a 'formal note style' is educationally useful.

B. Educational Effectiveness

- **Q12.** It will be helpful for improving students' understanding of research methodology.
- **Q13.** It can be an effective tool for achieving EBM educational goals.
- **Q14.** It can contribute to the development of students' critical thinking skills.
- **Q15.** It explains the complex research process well, tailored to the students' level.

C. System Usability

- **Q16.** The system interface is intuitive and convenient to use.
- **Q17.** The discussion questions for instructors are suitable for use in class.
- **Q18.** The feature for collecting student responses is useful for educational evaluation.
- **Q19.** The overall user experience is satisfactory.

D. AI Content Reliability

- **Q20.** The generated content is logical and consistent.
- **Q21.** The quality is sufficient for use as educational material.
- **Q22.** The automated evaluation of student answers is of sufficient quality to be useful.
- **Q23.** It seems applicable to various academic fields.

E. Practical Applicability

- **Q24.** It can be immediately used in the courses I am currently teaching.
- **Q25.** It can efficiently reduce class preparation time.
- **Q26.** It can be effectively used in conjunction with existing teaching methods.
- **Q27.** I am willing to recommend it to my colleagues.

F. Overall Assessment

- **Q28.** Overall, this is a system with high educational value.
- **Q29.** It can contribute to the improvement of EBM education.

Part 3: Qualitative Evaluation Items

- **Q30. Educational Contribution:** How do you think this system can contribute to EBM/research methodology education? Please describe specific applications or expected effects.
- **Q31. Suggestions for Improvement & Limitations:** Please freely describe any limitations or areas for improvement you foresee when using the system.
- **Q32. Ethical Considerations:** Please share your opinions on ethical aspects or reliability issues related to using generative AI in education.
- **Q33. Evaluation of Instructor Question Feature:** Please evaluate the discussion questions for instructors generated by the system (e.g., usability, difficulty, improvement directions).

- **Q34. Suggestions for Additional Features:** Please suggest any features or improvement ideas you would like to see added to the system.
- **Q35. Additional Comments (Optional)**

Supplementary File 2. Questionnaire Responses from the participants

ID	Gender	Age Group	Teaching Experience (since full-time faculty appointment)	Primary Subject Area	Primary Teaching Audience (multiple selections possible)	Frequency of Formal EBM/Research Methodology Education in the Last 3 Years	Frequency of Informal EBM/Research Methodology Education in the Last 3 Years
A	Male	30s	<5 years	Clinical Medicine	Clinical students, Graduate students	Regular coursework (undergraduate/graduate) - At least once a year	Informal educational activities (seminars, lectures, workshops, etc.) - At least once a year
B	Male	50s	>15 years	Basic Medicine	Clinical students, Graduate students	Regular coursework (undergraduate/graduate) - At least once a year	Informal educational activities (seminars, lectures, workshops, etc.) - At least once a year
C	Female	40s	<5 years	Medical Humanities	Pre-clinical students, Graduate students	Regular coursework (undergraduate/graduate) - At least once a year	Informal educational activities (seminars, lectures, workshops, etc.) - At least once a year
D	Female	30s	<5 years	Clinical Medicine	Clinical students	No formal educational activities in the last 3 years	Informal educational activities (seminars, lectures, workshops, etc.) - Less than once a year (e.g., once every 2 years)
E	Male	40s	10-15 years	Clinical Medicine	Clinical students, Graduate students	Regular coursework (undergraduate/graduate) - At least once a year	Informal educational activities (seminars, lectures, workshops, etc.) - At least once a year
F	Female	40s	5-10 years	Basic Medicine	Clinical students	Regular coursework (undergraduate/graduate) - At least once a year	No informal educational activities in the last 3 years
G	Male	40s	<5 years	Clinical Medicine	Clinical students, Graduate students	Regular coursework (undergraduate/graduate) - Less than once a year (e.g., once every 2 years)	Informal educational activities (seminars, lectures, workshops, etc.) - Less than once a year (e.g., once every 2 years)
H	Male	40s	10-15 years	Basic Medicine	Pre-clinical students, Clinical students, Graduate students	Regular coursework (undergraduate/graduate) - At least once a year	Informal educational activities (seminars, lectures, workshops, etc.) - At least once a year
I	Male	40s	10-15 years	Clinical Medicine	Clinical students	No formal educational activities in the last 3 years	Informal educational activities (seminars, lectures, workshops, etc.) - Less than once a year (e.g., once every 2 years)
J	Female	50s	>15 years	Basic	Pre-clinical students,	Regular coursework	Informal educational activities

				Medicine	Clinical students, Graduate students	(undergraduate/graduate) - At least once a year	(seminars, lectures, workshops, etc.) - At least once a year
K	Female	40s	<5 years	Medical Humanities	Pre-clinical students, Clinical students, Graduate students	Regular coursework (undergraduate/graduate) - Less than once a year (e.g., once every 2 years)	Informal educational activities (seminars, lectures, workshops, etc.) - At least once a year
L	Female	40s	<5 years	Basic Medicine	Clinical students	Regular coursework (undergraduate/graduate) - At least once a year	Informal educational activities (seminars, lectures, workshops, etc.) - At least once a year
M	Female	30s	<5 years	Basic Medicine	Pre-clinical students, Clinical students	Regular coursework (undergraduate/graduate) - At least once a year	Informal educational activities (seminars, lectures, workshops, etc.) - Less than once a year (e.g., once every 2 years)
N	Female	40s	<5 years	Basic Medicine	Clinical students, Graduate students	Regular coursework (undergraduate/graduate) - At least once a year	Informal educational activities (seminars, lectures, workshops, etc.) - Less than once a year (e.g., once every 2 years)
O	Male	40s	5-10 years	Clinical Medicine	Pre-clinical students, Clinical students, Graduate students	Regular coursework (undergraduate/graduate) - At least once a year	Informal educational activities (seminars, lectures, workshops, etc.) - At least once a year
P	Male	30s	5-10 years	Basic Medicine	Pre-clinical students, Clinical students	Regular coursework (undergraduate/graduate) - At least once a year	Informal educational activities (seminars, lectures, workshops, etc.) - Less than once a year (e.g., once every 2 years)

ID	Q30-35	Responses
A	**Educational Contribution**	*I was surprised by how well it incorporated the concerns researchers would likely have. The sequence and slight differences are similar to the actual concerns I had. This is amazing!*
	Suggestions for Improvement & Limitations	*However, the diary style has a lot of emotion, making it a bit cringe-inducing. This might be a limitation of text as a medium, but it could be educationally effective since these 'emotions' might have been real. However, depending on the student, those just trying it out might see it as shallow based on its appearance before understanding its true value.*
	Ethical Considerations	*As with any AI, details differ from the paper (number of subjects, submitted journals, etc., which were simple and clearly stated). Still, this can be reasonably accommodated.*
	Evaluation of	*The 'View Instructor Questions' feature is novel and great. It's good that it points out important aspects that I would otherwise have to*

	Instructor Question Feature	*study on my own from the paper.*
	Suggestions for Additional Features	*How about allowing the research note generation results to be downloaded as files? I'm sure there were various considerations, but since it can be screenshotted anyway...*
	Additional Comments	*This seems like it will be a great help to both students and instructors. This is brilliant!*
B	**Educational Contribution**	*I'm confused; I don't know how to use it in class. I need to be able to match it with the content targeted in class, but if the system proceeds in its own way and doesn't match the direction of my class, it would be difficult to use.*
	Suggestions for Improvement & Limitations	*1. It seems only very standard papers can be used. If it deviates even slightly from the set direction, it writes a novel. 2. It invents meanings of important words or statistical methods from the paper's content. It's difficult to use for understanding the paper's content. 3. I don't know how to position this system (Is the system the main tool and I'm the assistant? Do I use it as a supplement to my class objectives? Do I use it to explain key concepts of the paper?). 4. What is its precise purpose: Is it for simulating the process of obtaining research data? For deciding on analysis methods? For understanding the results in detail?*
	Ethical Considerations	*I don't think there are many ethical aspects to consider. As for reliability... it seems like the AI is just making things up based on the paper's content for predefined items.*
	Evaluation of Instructor Question Feature	*They are disconnected from the paper's content. T_T*
	Suggestions for Additional Features	*I don't really understand the system.*
	Additional Comments	*A clear purpose is needed. If it's to replace instructors, it should be something usable for a 15-week course. If it's to assist instructors, specific (detailed) purposes are needed, and learning objectives for specific topics should come first.*
C	**Educational Contribution**	*Research process education generally requires direct research experience, but experiencing it in advance through research notes is quite novel.*
	Suggestions for Improvement & Limitations	*It would be good if tables and figures could also be generated and displayed in the research notes.*
	Ethical Considerations	*If the paper's authors claim there is a discrepancy between the content of the research notes and their actual research process, it could be a burden in terms of educational effectiveness. I'm curious if it has undergone actual usability testing.*
	Evaluation of Instructor Question Feature	*I think it's appropriate that you've created categories for questions.*
	Suggestions for Additional Features	*I'd like to see functions for generating tables and figures added. For content like schedules, organizational charts, and task execution stages.*
	Additional Comments	*It seems like a good idea.*
D	**Educational Contribution**	*Yes. I think it would be useful for students who have received basic education in research methodology. It would provide an educational effect like a rehearsal before conducting actual research. It helps in understanding the entire research process, so it would be good to*

		use during a course.
	Suggestions for Improvement & Limitations	*When the feedback on a question becomes long, readability decreases. It would be good to use symbols, paragraph breaks, etc., to improve readability.*
	Ethical Considerations	*I have concerns about accuracy and reliability issues.*
	Evaluation of Instructor Question Feature	*I think it could be used in class.*
	Suggestions for Additional Features	*None*
	Additional Comments	*None*
E	**Educational Contribution**	*The idea is very good. It feels like a simulation of what the process would be if one were to apply research methodology in practice to conduct their own research. I think it will be helpful for students who have studied basic research methodology and want to create their own research plans.*
	Suggestions for Improvement & Limitations	*For now, there don't seem to be any particular limitations or needs for improvement.*
	Ethical Considerations	*I don't think there would be a major issue as long as students are told in advance that since it uses generative AI, it may not be accurate and that there should be a double-check for realistic/legal/ethical aspects.*
	Evaluation of Instructor Question Feature	*The instructor's discussion questions are divided into bachelor's/master's/doctoral levels, but the difficulty of the questions themselves doesn't seem to differ much. Instead of bachelor's/master's/doctoral levels, you might consider generating questions by parts such as 'research questions,' 'major research methodology,' 'data presentation,' 'major results,' 'interpretation,' and 'conclusion drawing'.*
	Suggestions for Additional Features	*No response*
	Additional Comments	*You seem to be implementing a very excellent idea. When the final version is made, I would definitely like to use it ^^*
F	**Educational Contribution**	*It could be used in relevant classes to help students think about the general research process.*
	Suggestions for Improvement & Limitations	*The research note generation took longer than expected (seemed to take over 10 minutes). Also, for experimental research, specific experimental details and conditions should be written down for each day and time, but it's expressed vaguely, making it a bit detached from reality.*
	Ethical Considerations	*I don't think there will be a problem.*
	Evaluation of Instructor Question Feature	*It seems a bit difficult for undergraduate students to answer. I don't feel a big difference between the questions for bachelor's, master's, and doctoral levels.*

	Suggestions for Additional Features	*No response*
	Additional Comments	*No response*
G	**Educational Contribution**	*It seems students can identify the parts they missed while receiving education.*
	Suggestions for Improvement & Limitations	*The flow of thought presented by the program may not be what the student or instructor intended, and there might be the inconvenience of having to re-learn the generated content.*
	Ethical Considerations	*There seems to be no issue from an ethical standpoint.*
	Evaluation of Instructor Question Feature	*The fact that it generates questions is good, but it would be nice if example answers were also provided.*
	Suggestions for Additional Features	*While the flow of thought can be important, it would be good if it could briefly present how to analyze the paper or in what direction to approach related research.*
	Additional Comments	*It would be nice to have a function to summarize the content for a quick view. The 10-page volume is more than I thought, and it takes a lot of time to produce the results.*
H	**Educational Contribution**	*The overall flow can be understood.*
	Suggestions for Improvement & Limitations	*It might be understandable for someone who has experienced it or is an existing researcher, but the terminology can still be difficult for students who have never experienced it.*
	Ethical Considerations	*I think using it is a good idea. However, it's difficult to rule out the possibility that students' responses will also use generative AI.*
	Evaluation of Instructor Question Feature	*The difficulty was higher than expected. Answering requires a lot of thought and research, so the barrier to entry is high.*
	Suggestions for Additional Features	*It seems inevitable, but there are too many lines, making readability difficult.*
	Additional Comments	*The system itself is amazing, and the idea is very brilliant, but to move beyond novelty to actual use, I feel a high barrier to entry. It seems necessary to set the level for undergraduate students who know nothing about EBM and go in an easier direction.*
I	**Educational Contribution**	*It seems it could improve students' accessibility to medical papers, especially English ones.*
	Suggestions for Improvement & Limitations	*The current system is unstable. It generated completely different content from the uploaded paper.*
	Ethical	*We will need to collect feedback after use to address any future problems that arise.*

	Considerations	
	Evaluation of Instructor Question Feature	*It is a useful feature.*
	Suggestions for Additional Features	*The basic answers produced nonsensical content, so I think the system is unstable. Overall, sufficient testing and improvement are needed.*
	Additional Comments	*No response*
J	**Educational Contribution**	*As a result of uploading and evaluating three types of papers (RCT, experimental paper, SR), the research notes created by the current system do not properly understand the paper and write content different from the paper. It is distorting the results (recording the opposite results or stating that an SR paper marked as IRB-exempt needed to go through IRB preparation, etc.). When I use a commonly available LLM to upload the same paper and ask it to write 10 days' worth of research notes, there are not many result distortions or interpretation errors. This part seems to need improvement. If this is improved and the research process is presented correctly, it could be helpful by using quizzes to evaluate if the students have understood it well.*
	Suggestions for Improvement & Limitations	*I'm not sure how understanding a research process through research notes is more efficient compared to other methods, so I'm not confident in giving feedback. ^^;*
	Ethical Considerations	*It seems necessary to provide a warning about the risk of being unethical or the possibility of giving low-reliability answers. However, I believe the right direction is to actively utilize this.*
	Evaluation of Instructor Question Feature	*Separating questions for evaluating research understanding by level is a good idea, but I think its utility would increase if options were added to present questions in more diverse ways than at present.*
	Suggestions for Additional Features	*Equipping a tool to evaluate [the paper] according to the research design might be more efficient for understanding the research and the necessary methods and processes than the research notes.*
	Additional Comments	*No response*
K	**Educational Contribution**	*I look forward to its excellent use in class.*
	Suggestions for Improvement & Limitations	*No response*
	Ethical Considerations	*No response*
	Evaluation of Instructor Question Feature	*The usability is good, and it doesn't feel too difficult for a researcher to use.*
	Suggestions for Additional Features	*No response*

	Additional Comments	*No response*
L	**Educational Contribution**	*For those who haven't conducted research, it can function as a step-by-step guide and help reduce fear.*
	Suggestions for Improvement & Limitations	*The statistical software used alternates between R, SPSS, and STATA. Content about visualization appears for a file without figures. Analysis methods not in the research content appear (e.g., meta-analysis).*
	Ethical Considerations	*The educational use of generative AI is ethically encouraged, but if this AI is used for actual research note creation, not just for educational purposes, there could be concerns about authenticity, so preventive measures are needed.*
	Evaluation of Instructor Question Feature	*No response*
	Suggestions for Additional Features	*It would be good to set it up so that the names of the first author, corresponding author, and research assistants can be entered separately and used automatically in the research notes.*
	Additional Comments	*Thank you for developing a good program.*
M	**Educational Contribution**	*Compared to existing AIs like ChatGPT that mainly provide PICO-style summaries, this system has many questions about the practical aspects of research design, so I think it is suitable for advanced education. In particular, the questions presented in the diary format can arouse interest in undergraduate classes and enable deep thinking, making them suitable for discussion-type questions.*
	Suggestions for Improvement & Limitations	*I didn't find any particular problems with the system. I uploaded and used it on a PC, but it would be good if tablets work smoothly.*
	Ethical Considerations	*I think it's important that instructors do not input raw data or charts when uploading materials.*
	Evaluation of Instructor Question Feature	*For example, if you set it to generate 5 questions, it would be good to have 1-2 questions about the basic PICO or study design. (For reference, the materials I uploaded were an SR paper and a CRIS Protocol.)*
	Suggestions for Additional Features	*I'm curious about the basis for the score on the answers. If there's no clear basis, I think it might be better to use a relaxed ordinal scale like A, B, C rather than xx/100 points.*
	Additional Comments	*It seems very suitable not just for EBM but especially for clinical education (chart analysis, patient judgment, etc.).*
N	**Educational Contribution**	*If this system is used after sufficiently understanding the basic concepts of systematic reviews, it could be helpful for education.*
	Suggestions for Improvement & Limitations	*The content provided in the system-generated research notes is not very realistic when compared to the timeline of actually writing a paper. Also, the generated material is very different from actual research notes. However, if it's used for research methodology education rather than for actually writing a paper, it could be a great help in understanding the overall process of writing a systematic review.*
	Ethical	*Using published papers to generate research notes does not seem to pose a major ethical problem.*

	Considerations	
	Evaluation of Instructor Question Feature	*I think the process of answering and grading the instructor's discussion questions is helpful for research methodology education.*
	Suggestions for Additional Features	*It would be helpful for education if the research notes and the instructor's discussion questions/responses could be downloaded.*
	Additional Comments	*Since this is a research note generator that tracks the researcher's thought process, it would be good if the paper's authors were reflected in the generated notes. Currently, generic surnames are used (e.g., Professor Lee, Researcher Kim), but if it's difficult to reflect the paper's authors, it might be better to remove the surnames. Also, it seems some content in the generated notes does not reflect the guidelines for writing systematic reviews (e.g., the paper screening process after a literature search should be done by two or more authors, but the research note is written as if one researcher did it).*
O	**Educational Contribution**	*It seems to be a tool that allows one to quickly think about meaningful research scenarios and the resulting problems and solutions.*
	Suggestions for Improvement & Limitations	*At first, I thought I was supposed to upload several papers. I think there should be an explanation of this app's development purpose and how to use it.*
	Ethical Considerations	*I'm not sure about that part.*
	Evaluation of Instructor Question Feature	*I think this part requires gathering opinions from experts in the future.*
	Suggestions for Additional Features	*It would be nice if there were a feature to export this content.*
	Additional Comments	*No response*
P	**Educational Contribution**	*The composition of the UI itself is intuitive, and the fact that it can easily show example research notes can be seen as helpful for education. It analyzes papers relatively appropriately, and the content of the research notes can also be seen as generally adequate.*
	Suggestions for Improvement & Limitations	**The limitations of the AI itself are bound to be revealed.* *1. It generally analyzes papers well, but there is a slight tendency for the AI to arbitrarily make up some of the content.* *2. The answers are primarily descriptive, so I think if charts or tables were utilized, it would become more intuitive and increase understanding.* *3. When producing 10 pages, the creation speed is on the slow side, so I wish it would become faster.* **Feedback on instructor questions* *The feedback for the questions is thought to require much improvement. Especially in the case of the evaluation scores, a high score is measured even when there is no answer content (or when a simple, irrelevant answer like 'a' was given).*
	Ethical Considerations	*In some cases, the AI also fabricates false information, so from the instructor's perspective, education on how to judge this part well seems necessary.*
	Evaluation of	*The question generation seems to be generally well done as it is divided by difficulty level. I wish it were possible to generate more*

	Instructor Question Feature	*concise questions as well. The feedback needs improvement overall.*
	Suggestions for Additional Features	*Separate development for instructor-use and student-use is needed. (e.g., For instructor-use - allow only for generating questions, checking student answers, and providing feedback. For student-use - allow only for answering the created questions and checking the feedback).*
	Additional Comments	*No response*

Note. Q30. Educational Contribution: How do you think this system can contribute to EBM/research methodology education? Please describe specific applications or expected effects.; Q31. Suggestions for Improvement & Limitations: Please freely describe any limitations or areas for improvement you foresee when using the system.; Q32. Ethical Considerations: Please share your opinions on ethical aspects or reliability issues related to using generative AI in education.; Q33. Evaluation of Instructor Question Feature: Please evaluate the discussion questions for instructors generated by the system (e.g., usability, difficulty, improvement directions).; Q34. Suggestions for Additional Features: Please suggest any features or improvement ideas you would like to see added to the system.; Q35. Additional Comments (Optional)