

An innovative Socratic method-based artificial intelligence platform for healthcare education: A quasi-experimental study

Yan Li^{a,b,c}, Janelle Yorke^a, Jiaying Li^{a,*}, Mengting He^a, Yushen Dai^a, Ivy Yan Zhao^a, Jing Qin^a, Xiangen Hu^d

^a School of Nursing, The Hong Kong Polytechnic University, Hong Kong, China

^b Mental Health Research Centre, The Hong Kong Polytechnic University, Hong Kong, China

^c Research Institute for Intelligent Wearable Systems, The Hong Kong Polytechnic University, Hong Kong, China

^d Department of Applied Social Sciences, The Hong Kong Polytechnic University, Hong Kong, China

ARTICLE INFO

Keywords:

The Socratic teaching method
Artificial intelligence
Healthcare education
Self-efficacy
Critical Thinking

ABSTRACT

Aim: To investigate the impact of the Socratic Playground for Learning system on self-efficacy, critical thinking skills and learning experiences among healthcare professional students.

Background: The Socratic teaching method fosters critical thinking and self-reflection but depends on instructor expertise. Generative AI enables personalized learning, though it may introduce biases and inaccuracies. The Socratic Playground for Learning system integrates both approaches to provide personalized learning experiences.

Design: This study employed a quasi-experimental, single-group pretest-posttest design.

Methods: This study collected data from a convenience sample of 31 students. Over the study period, participants were asked to explore the system alongside their scheduled teaching program. Self-efficacy and critical thinking ability were assessed before and after the exploration of the system. Additionally, participants' subjective learning experiences were evaluated after using the system. Independent sample t-tests were conducted to determine the significance of difference.

Results: The study included participants from six healthcare disciplines, with a mean professional experience of 5.71 ± 3.36 years. The results indicated a significant increase in self-efficacy scores at post-test, with an effect size of 0.57 (95% CI [0.02, 1.12], $p = 0.041$), though critical thinking skills showed no statistically significant change. At post-test, participants reported a stronger preference for the system compared with traditional teaching methods, reflecting overall positive perceptions of its usability.

Conclusions: Integrating generative artificial intelligence with the Socratic teaching method represents a promising advancement in healthcare education. Future research could continue to explore this approach to transform traditional educational paradigms.

1. Introduction

The Socratic teaching method, rooted in the philosophy of the ancient Greek thinker Socrates (Dalim et al., 2022), emphasizes knowledge generation through individual inquiry (Oyler and Romanelli, 2014). This approach encourages learners to ask questions, explore ideas, challenge assumptions and construct knowledge through dialogue (Ho et al., 2023). Its effectiveness is well-documented, particularly in healthcare settings, where it enhances critical thinking (Makhene, 2019), self-reflection (Dinkins and Cangelosi, 2019) and the ability to

justify beliefs and examine assumptions (Ho et al., 2023). Additionally, it facilitates knowledge acquisition through conversational learning (Acim, 2018). However, the success of Socratic questioning relies heavily on the teacher's expertise in applying these techniques and a lack of understanding of the authentic Socratic method can lead to suboptimal learning outcomes (Dinkins and Cangelosi, 2019).

Generative Artificial Intelligence (AI) is a category of AI designed to simulate human-like creativity and decision-making by generating new text, images, video, audio, or code based on its training data (Pescapé, 2024). In recent years, generative AI has been increasingly used in

* Correspondence to: School of Nursing, The Hong Kong Polytechnic University, 11 Yuk Choi Road, Hung Hom, Hong Kong, China.

E-mail address: jia98ying.li@connect.polyu.hk (J. Li).

<https://doi.org/10.1016/j.nepr.2026.104770>

Received 12 March 2025; Received in revised form 15 December 2025; Accepted 9 February 2026

Available online 10 February 2026

1471-5953/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

educational settings, ranging from kindergarten (Zhang and Tur, 2024) to higher education (Boscardin et al., 2024). Research indicates that generative AI has substantial potential to facilitate personalized instructional strategies tailored to individual student needs and aspirations (Boscardin et al., 2024). Additionally, it enhances learning experiences by providing feedback and facilitating interactive simulations (Al Murshidi et al., 2024). However, its use in education raises concerns, such as generating biased or misleading information, factual inaccuracies and insufficient contextual understanding (Boscardin et al., 2024; Elbanna and Armstrong, 2023). Furthermore, generative AI may challenge academic integrity and raise ethical issues (Nguyen et al., 2023).

The Socratic Playground for Learning (SPL) system integrates the Socratic method with advanced generative AI technology, employing GPT-4-based prompt strategies to create personalized learning experiences (Zhang et al., 2024). By using sophisticated natural language processing capabilities, the system delivers seamless multi-turn dialogue tutoring, reducing reliance on human effort and overcoming geographical limitations (Litman, 2016). The system guides learners toward self-discovery through thought-provoking questions rather than direct answers (Dalim et al., 2022), minimizing the risk of generating inaccurate information. Through dynamic and interactive experiences, SPL enhances the contextual relevance of learning inputs (Zhang et al., 2024). Consequently, this integrated system shows promise in encouraging learners to develop their own ideas, thereby mitigating biases in information and addressing academic integrity issues.

Healthcare professionals need a mix of knowledge, skills and decision-making abilities along with a commitment to continuous learning. Self-efficacy is crucial for students in healthcare fields as it has an impact on their drive, resilience and capability to handle the demands of tough training programs. A study indicated that during the COVID-19 crisis, professionals who trust their capacity to learn new treatments can swiftly adapt, leading to better patient outcomes (Baluszek et al., 2023). Critical thinking is another vital skill in healthcare. It helps students tackle complex situations, organize patient information and use evidence to make decisions (Richards et al., 2020). By encouraging logical reasoning, critical thinking enables healthcare professionals to reach informed conclusions when faced with tough choices based on careful analysis of facts and observations (Papathanasiou et al., 2014).

By combining the Socratic method with generative AI, the SPL system advances dialogue-driven, personalized learning while addressing limitations of traditional pedagogy and conventional AI tools. Through structured, multi-turn questioning sequences that transition from broad exploration to targeted inquiry, SPL empowers learners to construct knowledge through guided self-discovery. This iterative process strengthens self-efficacy by enabling incremental mastery and cultivates critical thinking by challenging learners to articulate reasoning, confront assumptions and refine solutions. Pilot experimental results from essay-based tasks suggest that SPL enhances tutoring experiences, with participants valuing its use of AI-driven Socratic dialogues to promote critical thinking and independent learning (Zhang et al., 2024). Notably, the integration of Socratic teaching with AI technology remains underexplored, highlighting the novelty and potential of the SPL system.

1.1. Aim of the study

This study aims to bridge the gap by implementing and evaluating the SPL system, an innovative Socratic method-based AI system, in the healthcare education context.

1.2. Research hypotheses

Integrating the SPL system into healthcare education will significantly enhance students' learning outcomes, as demonstrated by increased self-efficacy and improved critical thinking skills.

Implementing the SPL system will lead to positive and enriching

learning experiences for students in healthcare professions.

2. Methods

2.1. Design and participants

A single-group pre-post design was employed, with the study population comprising 31 master's students enrolled in the Concepts and Technological Applications of Smart Hospitals teaching program at The Hong Kong Polytechnic University. A convenience sampling method was used to include all students enrolled in the program to investigate the effects of applying the SPL system in the educational environment. According to previous research recommendations (Braun et al., 2020; Magnani et al., 2017), a sample size of 20–40 participants is generally considered appropriate for single-group pre-post studies, making the sample size of 31 suitable for achieving the study's objectives.

2.2. Intervention

The SPL system was applied as an intervention in this study. Powered by an advanced large language model (LLM), SPL is dialogue-based and designed to support in-context learning through the Socratic teaching method (Zhang et al., 2024). Fig. 1 illustrates various components of the system's interface. As shown in the login interface (A1), users begin by selecting their preferred system language, either using English or choosing their native language by selecting the appropriate region and sub-region. After logging in, users can generate their Socratic playground by describing the problem they wish to solve or the topic they want to explore, as seen in the A2 Generate the Socratic Playground interface. Once the user selects a learning theme, such as "Welcome to Smart Hospitals", the learning journey begins. The system then guides users through interactive Socratic dialogue, as shown in the theme Socratic dialogue interface (B), while generating related sub-themes in the left panel for deeper exploration. As users progress, they can explore detailed sub-topics under each sub-theme through the Socratic dialogue interface (C), engaging in multi-turn dialogue sessions.

Before the program began, an expert panel comprising two LLM experts and two healthcare disciplinary experts, conducted an internal test of the SPL system. The panel identified issues such as difficulty selecting certain languages, slow loading times when generating the Socratic Playground and occasional failures due to limited system concurrency. These issues were reported to the technical team, who implemented SPL system debugging and enhancements to improve the system's robustness and performance. The system was optimized for better concurrency, ensuring stable performance during peak usage times, with the ability to handle up to 100 simultaneous users from multiple devices.

Participants began engaging with the SPL system in Week 3 of the teaching program, following the completion of baseline questionnaires. Prior to this, two PhD teaching assistants from healthcare disciplines conducted a comprehensive orientation session, introducing the SPL system's background, objectives and usage instructions. Structured interactions with the system were scheduled during Weeks 3, 5, 7, 9 and 12 of the program in-class, while students were also encouraged to use the system independently during their after-class study time. Throughout the program, the teacher and teaching assistants provided ongoing technical support to ensure a seamless learning experience.

2.3. Measurements

2.3.1. General self-efficacy scale (GSES)

General self-efficacy scale of the English version was used to assess an individual's overall sense of perceived self-efficacy. The scale aims to predict a person's ability to cope with daily challenges and adapt to various stressful life events. It consists of 10 items rated on a four-point Likert scale (1 = Not at all true, 4 = Exactly true). The total score is

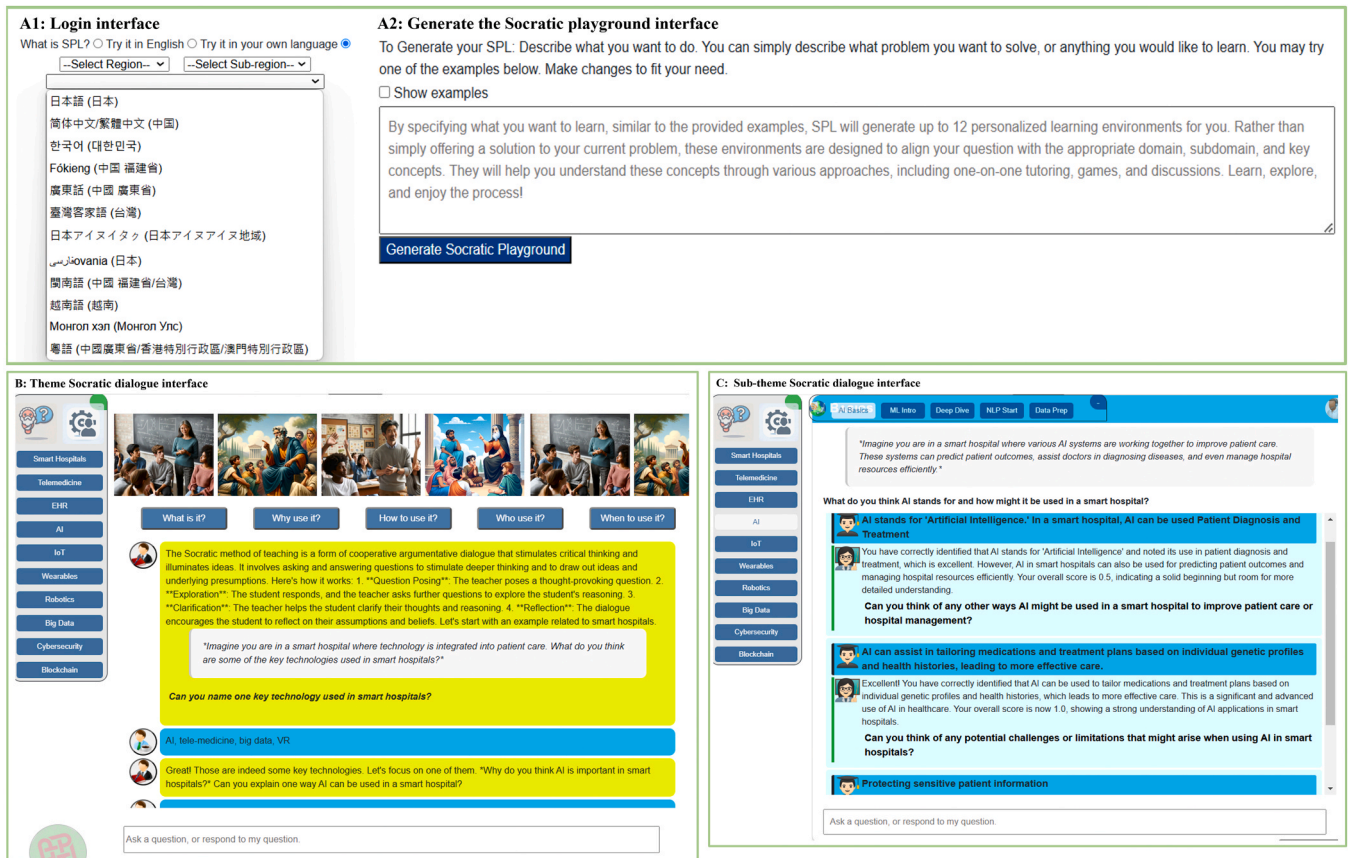


Fig. 1. The interfaces of the Socratic Playground for Learning system

obtained by summing all items, resulting in a range from 10 to 40, with higher scores reflecting greater self-efficacy (Schwarzer et al., 1995). Cronbach’s α ranged from 0.82 to 0.93 in the original study and was 0.82 in the present sample.

2.3.2. Critical thinking self-assessment (CTSA)

The Critical Thinking Self-Assessment, a widely used instrument in educational settings, was employed to evaluate students’ critical thinking abilities (UCEN Manchester, 2024). This tool consists of seven items, each rated on a three-point Likert scale (1 = Never, 2 = Sometimes, 3 = Always). Total scores range from 7 to 21, with higher scores reflecting stronger critical thinking skills. In the present study, the instrument demonstrated a Cronbach’s α of 0.76, indicating acceptable internal consistency.

2.3.3. Subjective learning experience

Subjective learning experience was assessed using four single-item statements and one open-ended question. Students were asked to provide feedback on the SPL system by rating four statements on a five-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree) (Hsiao et al., 2016): (1). The Socratic Playground for Learning (SPL) on “Concepts and Technological Applications of Smart Hospitals” was helpful to my learning; (2). I like the Socratic Playground for Learning (SPL) on “Concepts and Technological Applications of Smart Hospitals” more than conventional teaching methods; (3). The Socratic Playground for Learning (SPL) on “Concepts and Technological Applications of Smart Hospitals” helped me achieve the subject learning outcomes; and (4). The Socratic Playground for Learning (SPL) on “Concepts and Technological Applications of Smart Hospitals” helped me become a better learner. Higher scores on these items reflected a more positive attitude toward the SPL system and its role in enhancing the learning experience. Additionally, students were encouraged to provide qualitative feedback

through an open-ended question: What are your comments about the Socratic Learning Playground (SPL)?

2.4. Data collection

Data collection was conducted by two trained PhD teaching assistants specializing in healthcare disciplines. They distributed and collected paper-based questionnaires during teaching sessions at two time points: Week 3 (baseline, prior to SPL implementation) and Week 12 (post-intervention). This dual-phase approach enabled a comparison of students’ perceptions and learning outcomes, capturing potential shifts in attitudes or skill development over time. The data collection procedure is presented in Fig. 2.

2.5. Ethical consideration

The study received approval from the Human Subjects Ethics Committee at the Hong Kong Polytechnic University (No. HSEARS20240906003). In Week 3, all participants were given a written informed consent form and an information sheet outlining the study’s purpose, procedures and their rights. They were assured that their completed questionnaires would remain anonymous and that participating in the study would not affect their evaluation results for the subject. Participation was completely voluntary and participants could withdraw from the study at any time during the study period.

2.6. Data analysis

Data were analyzed using the IBM SPSS version 26.0. Statistical significance was set at $p < 0.05$. Descriptive statistics were used to summarize the data based on the types and distributions of the variables, including means and standard deviations, as well as frequencies and

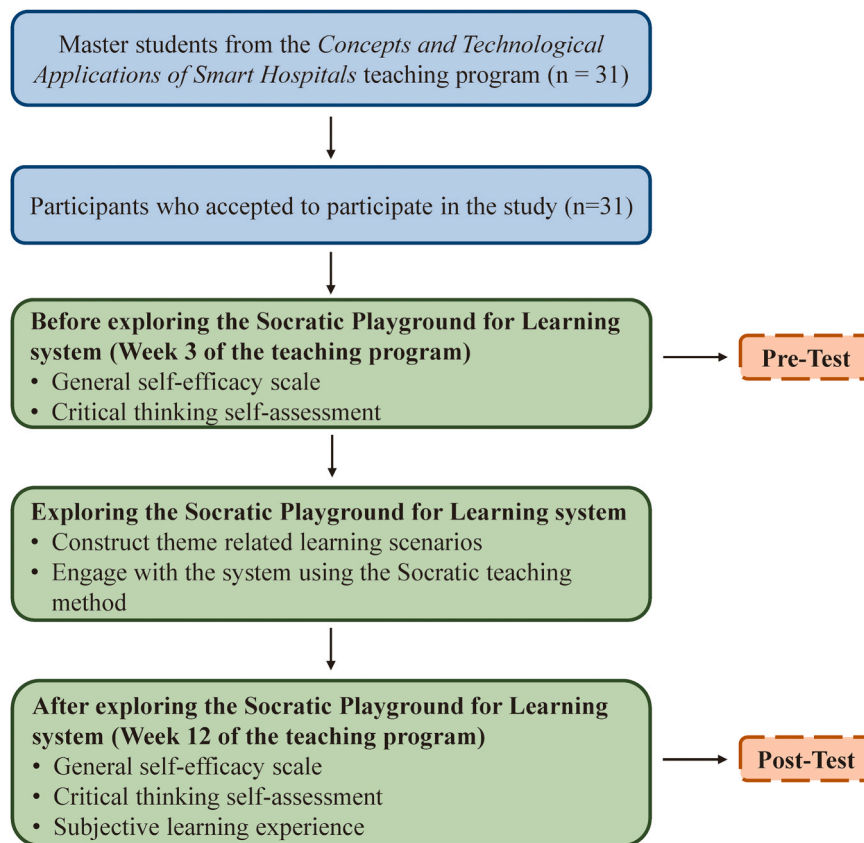


Fig. 2. Data collection procedure

percentages. An independent sample *t*-test was conducted to compare self-efficacy and critical thinking abilities before and after using SPL. The effect size was calculated using Cohen’s *d*, with *d* = 0.20 considered a small effect size, *d* = 0.50 a medium effect size and *d* = 0.80 a large effect size (Cohen, 1988).

3. Results

Most students (46.7 %) came from the radiography health discipline. On average, participants completed their bachelor’s degree 6.6 years ago and have accumulated 5.7 years of working experience. The detailed characteristics of the student sample are summarized in Table 1.

A comparison of pre-test and post-test scores for self-efficacy and critical thinking is presented in Table 2. The results demonstrated a significant improvement in self-efficacy, with a medium effect size of 0.57 (95 % CI [0.02, 1.12], *p* = 0.041). While the improvement in critical thinking was not statistically significant (*p* = 0.639), there was a slight positive change in score from a pre-test mean of 17.00 to a post-test mean of 17.29.

Table 1 Characteristics of the student sample.

Characteristics	n (%) / Mean (SD)
Health discipline (N = 30)	
Radiography	14 (46.67 %)
Nursing	10 (33.33 %)
Health Technology and Informatics	3 (10.00 %)
Occupational Therapist	1 (3.33 %)
Health Studies	1 (3.33 %)
Medicine	1 (3.33 %)
Years of achieving the bachelor’s degree (N = 29)	6.62 (3.98)
Years of working experience (N = 29)	5.71 (3.36)

Note: Sample size varied due to missing data.

Table 2 Comparison of mean scores on the GSES and CTSA for pre- and post-test.

Variables	T0 (pre, N = 23)	T1 (post, N = 31)	<i>t</i> value	<i>p</i> value	Cohen’s <i>d</i> (95 % CI)
GSES, mean (SD)	27.70 (4.27)	29.77 (3.01)	-2.101	0.041	0.57 (0.02, 1.12)
CTSA, mean (SD)	17.00 (2.17)	17.29 (2.28)	-0.471	0.639	0.13 (-0.41, 0.67)

Notes: GSES = General Self-Efficacy Scale; CTSA = Critical Thinking Self-Assessment; Sample size varied due to missing data.

Table 3 presents the subjective learning experience with the SPL system, based on four single-item statements collected at post-test. All items scored above the neutral midpoint (3.0), indicating a generally positive perception of the system. Specifically, students found the SPL system helpful for their learning, preferred it over conventional teaching

Table 3 Subjective learning experience at post-test (N = 31).

Items	Mean (SD)
1. The Socratic Playground for Learning (SPL) on “Concepts and Technological Applications of Smart Hospitals” was helpful to my learning.	3.68 (0.65)
2. I like the Socratic Playground for Learning (SPL) on “Concepts and Technological Applications of Smart Hospitals” more than conventional teaching methods.	3.81 (0.75)
3. The Socratic Playground for Learning (SPL) on “Concepts and Technological Applications of Smart Hospitals” helped me achieve the subject learning outcomes.	3.55 (0.57)
4. The Socratic Playground for Learning (SPL) on “Concepts and Technological Applications of Smart Hospitals” helped me become a better learner.	3.68 (0.70)

methods and believed it contributed to achieving subject learning objectives and enhancing their learning abilities.

Students also provided qualitative feedback through an open-ended question about the system. Most responses were positive, highlighting the system's engaging and useful nature and its potential to improve critical thinking, as reflected in one comment: *"The system is quite fun and very useful, improved our critical thinking."* Suggestions for improvement focused on usability, with a student noting, *"The user interface can be simplified to facilitate more efficient use of the SPL."*

4. Discussion

The SPL system introduces a structured, dialogue-driven application of GPT-4 tailored for healthcare education, addressing gaps in personalized learning. Unlike existing GPT systems that generate tutoring scripts without pedagogical guidance (Schmucker et al., 2023), SPL employs iterative, context-sensitive questioning to scaffold self-reflection and critical analysis. This approach is consistent with recent findings that integrating the Socratic method into large language models for teaching fosters deeper student thinking and self-discovery through structured dialogue (Ding et al., 2024), positioning such systems as uniquely pedagogical tools.

Preliminary results demonstrate SPL's success in improving learning self-efficacy - a critical competency for healthcare professionals facing complex scenarios (Klassen and Klassen, 2018; Yu et al., 2021). This finding aligns with previous research that highlights the effectiveness of an educational chatbot in a nursing training program, which enhances self-efficacy by enabling interactive questioning to acquire professional knowledge (Chang et al., 2021). The SPL system operationalizes Bandura's self-efficacy theory through adaptive prompts that break tasks into manageable steps, coupled with formative feedback (Bandura, 1991). Additionally, the SPL system's success can be attributed to its interactive conversational process, which deepens understanding, corrects misconceptions and guides students toward their learning goals. This approach aligns with the principles of expectation misconception tailoring (Graesser, 2016; Graesser et al., 2001), designed to effectively address and rectify students' misconceptions. By consistently offering targeted prompts and feedback, the SPL system has the potential to reinforce knowledge, improve problem-solving skills and boost confidence among students. Although a moderate effect size was found, the wide 95 % confidence interval (0.02–1.12) indicates substantial uncertainty, likely due to the small sample size, response variability and single-group pre-post design. Thus, causal conclusions about SPL's impact on self-efficacy should be made cautiously. These preliminary findings highlight the need for larger, controlled studies to obtain more reliable estimates.

Unlike findings from studies in nursing education that reported significant gains in critical thinking using AI-based dialogue systems, our results did not show a statistically significant improvement after using the SPL system (Chang et al., 2022; Makhlouf et al., 2024). Although a slight increase in critical thinking scores was observed, several factors may account for the lack of significant change. First, the relatively small sample size likely limited statistical power, making it difficult to detect subtle effects. Second, the intervention consisted of only five in-class sessions with voluntary self-practice, which may not have provided sufficient duration or intensity to support students' adaptive learning and reflection, thereby limiting measurable improvements in critical thinking. Third, the assessment tool used in this study may have lacked sufficient sensitivity and validity to detect subtle changes in critical thinking over a short intervention period. Future research could address this limitation by employing well-validated instruments, such as the California Critical Thinking Disposition Inventory or the Watson-Glaser Critical Thinking Appraisal (Facione, 1990; Watson, 1980). Another key distinction is the system design. The SPL system relies on the GPT-4 API and a generalized, potentially outdated knowledge base, which may limit its effectiveness in delivering accurate, current and context-specific

information (OpenAI et al., 2023). This limitation is particularly relevant in healthcare education, where knowledge evolves rapidly and the need for up-to-date, evidence-based content is critical. In contrast, teaching systems in prior studies were built on regularly updated, domain-specific knowledge bases, allowing them to provide more relevant and current educational resources tailored to their discipline (Chang et al., 2022; Makhlouf et al., 2024). To address these challenges, future iterations of the SPL system could integrate discipline-specific knowledge bases or employ retrieval-augmented generation techniques. These approaches combine real-time retrieval from specialized databases with generative AI models to deliver high-quality, up-to-date teaching materials that better support the development of critical thinking (Lozano et al., 2024; Soong et al., 2024).

The SPL system, powered by a LLM and employing standard prompt strategies, is designed to offer personalized, adaptive, flexible and versatile dialogue to enhance educational interactions (Zhang et al., 2024). This capability is further validated by post-intervention survey results, which reveal that students view the SPL system as an easy-to-use and valuable resource for achieving learning outcomes and improving their overall educational experience. These findings are consistent with previous research that underscores the potential benefits of LLM-based chatbot systems in intelligent education in the healthcare sector (Abd-Alrazaq et al., 2023). Such positive feedback highlights the SPL system's potential as a valuable tool in educational practice and suggests that its integration into conventional healthcare education could significantly transform medical curriculum development.

4.1. Strengths and limitations

The SPL system is a groundbreaking application of generative AI in healthcare education. The results showed that it effectively enhances self-efficacy and received positive user feedback, underscoring its potential for scalable implementation across various educational settings. However, this study has several limitations. First, the small sample size of healthcare professional students may have reduced statistical power. Second, the absence of longitudinal outcome measurements restricts the generalizability of findings over time. Third, the quasi-experimental single-group pre-post design limits the ability to control for confounding factors and restricts causal inferences regarding the true effectiveness of the system. Lastly, the absence of a well-established critical thinking assessment, along with the lack of standardized qualitative data collection and analysis procedures, may have limited the depth and rigor of the findings.

5. Conclusion

Our preliminary evaluation indicates that the SPL system has a positive impact on students' self-efficacy and learning experiences. The integration of generative AI with the Socratic teaching pedagogy represents a promising step toward advancing intelligent healthcare education. Future research should continue to explore and refine this combination to unlock its potential for transforming traditional educational paradigms.

5.1. Recommendations

Future research should prioritize evaluating the SPL system's effectiveness through large-scale and longitudinal studies to strengthen the generalizability and durability of findings. We recommend conducting randomized controlled trials with adequate sample sizes to enhance statistical power and more effectively detect the effects of the SPL system on students' learning self-efficacy and critical thinking abilities in educational settings. Employing validated quantitative and qualitative approaches will provide a robust evaluation of the system. Refining user experience through multimodal enhancements, such as explanatory videos and intuitive interface redesigns, could address clarity gaps and

foster greater engagement. To establish SPL as a scalable learning platform in healthcare education, technical improvements such as optimized API integration and retrieval-augmented generation are necessary to ensure response accuracy.

CRedit authorship contribution statement

Xiangen Hu: Writing – review & editing, Supervision, Software. **Jiaying Li:** Writing – original draft, Project administration, Formal analysis, Data curation. **Mengting He:** Writing – review & editing, Project administration, Conceptualization. **Yan Li:** Writing – original draft, Project administration, Methodology, Funding acquisition, Data curation, Conceptualization. **Janelle Yorke:** Writing – review & editing, Supervision, Conceptualization. **Jing Qin:** Writing – review & editing, Supervision. **Yushen Dai:** Writing – review & editing, Project administration, Data curation. **Ivy Yan Zhao:** Writing – review & editing, Project administration, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Acknowledgement is given to the University Grants Committee-funded project "Socratic Playground for Learning in Health Informatics" and partial support from The Hong Kong Polytechnic University Strategic Hiring Scheme (Ref No. P0044765).

References

- Abd-Alrazaq, A., AlSaad, R., Alhuwail, D., Ahmed, A., Healy, P.M., Latifi, S., Aziz, S., Damsheh, R., Alabed Alrazak, S., Sheikh, J., 2023. Large language models in medical education: opportunities, challenges and future directions. *JMIR Med. Educ.* 9, e48291. <https://doi.org/10.2196/48291>.
- Acim, R., 2018. The Socratic method of instruction: an experience with a reading comprehension course. *J. Educ. Res.* Pr. 8 (1), 41–53. <https://doi.org/10.5590/JERAP.2018.08.1.04>.
- Al Murshidi, G., Shulgina, G., Kapuza, A., Costley, J., 2024. How understanding the limitations and risks of using ChatGPT can contribute to willingness to use. *Smart Learn. Environ.* 11 (1), 36. <https://doi.org/10.1186/s40561-024-00322-9>.
- Baluszek, J.B., Brønnick, K.K., Wiig, S., 2023. The relations between resilience and self-efficacy among healthcare practitioners in context of the COVID-19 pandemic – a rapid review. *Int. J. Health Gov.* 28 (2), 152–164. <https://doi.org/10.1108/ijhg-11-2022-0098>.
- Bandura, A., 1991. Social cognitive theory of self-regulation. *Organ. Behav. Hum. Decis. Process* 50 (2), 248–287. [https://doi.org/10.1016/0749-5978\(91\)90022-L](https://doi.org/10.1016/0749-5978(91)90022-L).
- Boscardin, C.K., Gin, B., Golde, P.B., Hauer, K.E., 2024. ChatGPT and generative artificial intelligence for medical education: potential impact and opportunity. *Acad. Med.* 99 (1), 22–27. <https://doi.org/10.1097/acm.0000000000005439>.
- Braun, S.S., Roeser, R.W., Mashburn, A.J., 2020. Results from a pre-post, uncontrolled pilot study of a mindfulness-based program for early elementary school teachers. *Pilot Feasibility Stud.* 6 (1), 178. <https://doi.org/10.1186/s40814-020-00718-7>.
- Chang, C.-Y., Hwang, G.-j., Gau, M.-l., 2021. Promoting students' learning achievement and self-efficacy: A mobile chatbot approach for nursing training. *Br. J. Educ. Technol.* 53, 171–188.
- Chang, C.-Y., Kuo, S.-y., Hwang, G.-H., 2022. Chatbot-facilitated nursing education: incorporating a knowledge-based chatbot system into a nursing training program. *Educ. Technol. Soc.*
- Cohen, J., 1988. *Statistical Power Analysis for the Behavioral Sciences*, second ed. Routledge. <https://doi.org/10.4324/9780203771587>.
- Dalim, S.F., Ishak, A.S., Hamzah, L.M., 2022. Promoting students' critical thinking through socratic method: Views and challenges. *Asian J. Univ. Educ.* 18 (4), 1034–1047.
- Ding, Y., Hu, H., Zhou, J., Chen, Q., Jiang, B., & He, L. (2024). Boosting Large Language Models with Socratic Method for Conversational Mathematics Teaching Proceedings of the 33rd ACM International Conference on Information and Knowledge Management, Boise, ID, USA. <https://doi.org/10.1145/3627673.3679881>.
- Dinkins, C.S., Cangelosi, P.R., 2019. Putting Socrates back in Socratic method: theory-based debriefing in the nursing classroom. *Nurs. Philos.* 20 (2), e12240. <https://doi.org/10.1111/nup.12240>.
- Elbanna, S., Armstrong, L., 2023. Exploring the integration of ChatGPT in education: adapting for the future. *Manag. Sustain. Arab Rev.* 3 (1), 16–29. <https://doi.org/10.1108/MSAR-03-2023-0016>.
- Facione, P. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi Report)*. The California Academic Press.
- Graesser, A.C., 2016. Conversations with AutoTutor help students learn. *Int. J. Artif. Intell. Educ.* 26, 124–132.
- Graesser, A.C., VanLehn, K., Rosé, C.P., Jordan, P.W., Harter, D., 2001. Intelligent tutoring systems with conversational dialogue. *AI Mag.* 22 (4), 39, 39.
- Ho, Y.R., Chen, B.Y., Li, C.M., 2023. Thinking more wisely: using the Socratic method to develop critical thinking skills amongst healthcare students. *BMC Med. Educ.* 23 (1), 173. <https://doi.org/10.1186/s12909-023-04134-2>.
- Hsiao, C.C., Tiao, M.M., Chen, C.C., 2016. Using interactive multimedia e-Books for learning blood cell morphology in pediatric hematology. *BMC Med. Educ.* 16 (1), 290. <https://doi.org/10.1186/s12909-016-0816-9>.
- Klassen, R.M., Klassen, J.R.L., 2018. Self-efficacy beliefs of medical students: a critical review. *Perspect. Med. Educ.* 7 (2), 76–82. <https://doi.org/10.1007/s40037-018-0411-3>.
- Litman, D. (2016). *Natural language processing for enhancing teaching and learning Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, Phoenix, Arizona.
- Lozano, A., Fleming, S.L., Chiang, C.C., Shah, N., 2024. Clinfo.ai: an open-source retrieval-augmented large language model system for answering medical questions using scientific literature. *Pac. Symp. Biocomput* 29, 8–23.
- Magnani, D., Lenoci, G., Balduzzi, S., Artioli, G., Ferri, P., 2017. Effectiveness of support groups to improve the quality of life of people with idiopathic pulmonary fibrosis a pre-post test pilot study. *Acta Biomed.* 88 (5s), 5–12. <https://doi.org/10.23750/abm.v88i5-S.6870>.
- Makhene, A., 2019. The use of the Socratic inquiry to facilitate critical thinking in nursing education. *Health SA* 24, 1224. <https://doi.org/10.4102/hsag.v24i0.1224>.
- Makhlouf, E., Alenezi, A., Shokr, E.A., 2024. Effectiveness of designing a knowledge-based artificial intelligence chatbot system into a nursing training program: A quasi-experimental design. *Nurse Educ. Today* 137, 106159. <https://doi.org/10.1016/j.nedt.2024.106159>.
- Nguyen, A., Ngo, H.N., Hong, Y., Dang, B., Nguyen, B.T., 2023. Ethical principles for artificial intelligence in education. *Educ. Inf. Technol.* 28 (4), 4221–4241. <https://doi.org/10.1007/s10639-022-11316-w>.
- OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., Zoph, B., 2023. GPT4 Tech. Rep. <https://doi.org/10.48550/arXiv.2303.08774>.
- Oyler, D.R., Romanelli, F., 2014. The fact of ignorance: revisiting the Socratic method as a tool for teaching critical thinking. *Am. J. Pharm. Educ.* 78 (7), 144. <https://doi.org/10.5688/ajpe787144>.
- Papathanasiou, I.V., Kleisiaris, C.F., Fradelos, E.C., Kakou, K., Kourkouta, L., 2014. Critical thinking: the development of an essential skill for nursing students. *Acta Inform. Med.* 22 (4), 283–286. <https://doi.org/10.5455/aim.2014.22.283-286>.
- Pescapé, A., 2024. *Exploring the Current State and Future Potential of Generative Artificial Intelligence Using a Generative Artificial Intelligence*. *Mind, Body and Digital Brains*. Springer, pp. 37–56.
- Richards, J.B., Hayes, M.M., Schwartzstein, R.M., 2020. Teaching Clinical Reasoning and Critical Thinking: From Cognitive Theory to Practical Application. *Chest* 158 (4), 1617–1628. <https://doi.org/10.1016/j.chest.2020.05.525>.
- Schmucker, R., Xia, M., Azaria, A., & Mitchell, T. (2023). Ruffle&riley: Towards the automated induction of conversational tutoring systems. *arXiv Prepr.*
- Schwarzer, R., Jerusalem, M., Weinman, J., Wright, S., Johnston, M., 1995. *Measures in health psychology: A user's portfolio. Causal and control beliefs*. Nfer-Nelson Windsor.
- Soong, D., Sridhar, S., Si, H., Wagner, J.S., Sá, A.C.C., Yu, C.Y., Karagoz, K., Guan, M., Kumar, S., Hamadeh, H., Higgs, B.W., 2024. Improving accuracy of GPT-3/4 results on biomedical data using a retrieval-augmented language model. *PLOS Digit. Health* 3 (8), e0000568. <https://doi.org/10.1371/journal.pdig.0000568>.
- UCEN Manchester. *Critical Thinking Self-Assessment*. Retrieved 14 September 2024 from https://www.ucenmanchester.ac.uk/media/filer_public/5e/0e/5e0ebb4e-9c1b-4e08-bc2f-1aacac4fb757/selfevaluation-critical-thinking.pdf.
- Watson, G., 1980. *Watson-Glaser critical thinking appraisal*, 3. Psychological Corporation, San Antonio, TX.
- Yu, M., Tong, H., Li, S., Wu, X.V., Hong, J., Wang, W., 2021. Clinical competence and its association with self-efficacy and clinical learning environments among Chinese undergraduate nursing students. *Nurse Educ. Pr.* 53, 103055. <https://doi.org/10.1016/j.nepr.2021.103055>.
- Zhang, L., Lin, J., Kuang, Z., Xu, S., Yeasin, M., & Hu, X. (2024). SPL: A Socratic Playground for Learning Powered by Large Language Mode. *arXiv preprint arXiv:2406.13919*.
- Zhang, P., Tur, G., 2024. A systematic review of ChatGPT use in K-12 education. *Eur. J. Educ.* 59 (2), e12599. <https://doi.org/10.1111/ejed.12599>.