ORIGINAL RESEARCH

Evaluating a zero-shot GenAI assistant for clinical record writing in nursing education

Asahiko Higashitsuji,* Tomoko Otsuka

Department of Nursing, Chiba Prefectural University of Health Sciences, Chiba, Japan

Received: November 12, 2025 **Accepted:** December 1, 2025 **Online Published:** December 12, 2025
DOI: 10.63564/jnep.v15n12p29 **URL:** https://doi.org/10.63564/jnep.v15n12p29

ABSTRACT

Objective: This study pilot-tested a zero-shot prompting approach for generating pedagogically appropriate AI feedback to support nursing students' reflective record writing. A generative AI system was developed using GPT-4o, with prompts carefully designed to align with ethical, instructional, and contextual principles relevant to clinical practicum education. Following the ADDIE framework, this evaluation examined the feasibility and instructional applicability of the AI-generated feedback rather than aiming to establish generalizable effects.

Methods: Three nursing faculty members independently reviewed 126 AI-generated responses using a 5-point scale based on clarity, relevance, and educational appropriateness. Items scoring less than 5 were revised, and a second round of evaluation was conducted.

Results: The proportion of responses rated 5 increased from 69% to 96% after prompt refinement, confirming that even minor adjustments—such as clarifying vague instructions or reinforcing ethical boundaries—had a measurable impact. Inter-rater reliability was moderate, reflecting diverse faculty perspectives—a feature aligned with the contextual complexity of nursing education. The conservative scoring approach, in which the lowest score was adopted per item, ensured that potential pedagogical risks were not overlooked.

Conclusions: These findings suggest that zero-shot prompting offers a practical and scalable method for aligning generative AI systems with educational goals, even without training data or programming expertise. Rather than positioning AI as a replacement for instructors, this study frames GenAI as a formative support tool shaped by educator input. The study is limited by its use of simulated student inputs and a small, single-institution faculty sample; therefore, future work should assess implementation with real students to examine usability, personalization, and effectiveness in authentic practicum environments.

Key Words: Artificial Intelligence, Clinical competence, Computer-assisted instruction, Education, Nursing, Students

1. INTRODUCTION

Clinical practicums are an essential component of nursing education, providing students with opportunities to integrate theoretical knowledge with real-world experiences and cultivate their critical thinking through direct patient care.[1,2] In this context, a "practicum" refers to supervised clinical training in a healthcare setting, and a "reflective record" refers to structured student documentation used to reflect on patient care experiences and clinical reasoning. During practicums, the reflective record acts as a vital part of professional development that helps deepen students' understanding of care recipients and their own clinical reasoning.[3] Despite its importance, many students encounter difficulties in producing high-quality reflective records, particularly in the absence of

*Correspondence: Asahiko Higashitsuji; Email: asahiko.higashitsuji@cpuhs.ac.jp; Address: Department of Nursing, Chiba Prefectural University of Sciences, 2-10-1 Wakaba, Mihama-ku, Chiba-shi, Chiba 261-0014, Japan.

structured guidance and continuous supervision from faculty members.[4] As educational demands continue to grow and faculty resources remain limited, there has been increasing interest in the use of generative artificial intelligence (GenAI) to offer scalable, consistent pedagogical support.[5,6]

Although interest in the use of GenAI in education is rapidly increasing, concerns remain about students using such tools without faculty supervision and the absence of models specifically tailored for nursing education.[7,8] Given that GenAI models generate responses probabilistically based on input prompts and prior training data, their outputs cannot be deterministically controlled or guaranteed to align with pedagogical standards.[9] This limitation highlights the need for faculty-led evaluation methods to ensure educational validity and ethical alignment prior to practical implementation in nursing education. Indeed, methodologies for systematically incorporating GenAI into educational practice have not yet been clearly defined,[10] and the deployment of AI in real-world educational contexts remains at an early stage.[11] Furthermore, both faculty and students recognize that current GenAI tools are not yet sufficiently advanced to replace or fully replicate expert instruction.[6,11] Nursing faculty across institutions are therefore actively exploring how to responsibly adapt GenAI for educational purposes.[8] Sharing concrete cases of such adaptation—including the time, effort, and resources required for implementation—can contribute to the advancement of nursing education and promote informed, practice-oriented innovation within the field. Recent reviews in healthcare AI ethics also emphasize that safe and responsible implementation of AI necessitates attention to technical accuracy and ethical risks, including bias, opacity, accountability, and data governance, as well as the broader social and institutional factors influencing adoption.[12,13] These findings highlight the importance of ensuring that AI-assisted educational tools are aligned with ethical standards and supported by appropriate governance and oversight mechanisms.

A further challenge lies in developing AI systems genuinely optimized for nursing education. This study addressed this issue by employing a zero-shot learning approach to construct an AI model and a student-accessible application system, drawing exclusively on information from classroom instructional materials.

Zero-shot prompting offers a practical method for adapting GenAI to nursing education. In this approach, a language model is guided by standalone instructions without requiring any example responses.[14] Through such prompting, even a general-purpose AI model not originally designed for nursing education can be directed to generate responses that align with specific learning goals—or even with the unique needs of particular domains within nursing education, such as clinical practicum documentation or institution-specific instructional priorities. This approach also offers significant advantages: it requires no training data, programming expertise, or high-performance computing resources. These features make zero-shot learning especially suitable for clinical practicum settings, where time and technical resources are often limited. In contrast, "few-shot learning" embeds several input–output examples into the prompt, requiring additional preparation and potentially limiting generalizability.[15] "Fine-tuning" goes further by retraining the model on domain-specific datasets, improving accuracy but demanding substantial resources, technical expertise, and annotated data—and it may also introduce biases.[15] Given these constraints, zero-shot prompting provides an efficient, scalable, and educator-friendly strategy for implementing GenAI in nursing education.

Given the emerging nature of AI-assisted instruction in nursing education, this study was deliberately designed as a pilot instructional method. Rather than generating generalizable findings, this pilot study aimed to explore the feasibility, instructional alignment, and practical considerations associated with implementing a zero-shot prompting approach in clinical practicum education.

To examine the pedagogical viability of this approach, the present study evaluated an AI-based support system designed to assist nursing students with reflective documentation during clinical practicums. The system was developed using classroom-based instructional materials and implemented via a web-based interface. Faculty members with expertise in practicum education systematically reviewed AI-generated responses to simulated student inputs, assessing their clarity, relevance, and instructional appropriateness. Through this evaluation, the study aimed to determine whether a carefully designed zero-shot prompting approach could serve as a practical and ethically sound tool for supporting individualized learning in resource-constrained clinical settings.

2. METHODS

2.1 Study design

This study employed a pilot instructional evaluation, using a cross-sectional design with repeated evaluation points to assess and refine AI-generated feedback using a zero-shot prompting approach in nursing education. The development of the generative AI (GenAI) system was guided by the ADDIE framework[16] —Analysis, Design, Development, Implementation, and Evaluation. The present study specifically focused on the Development, Implementation, and Evaluation phases, examining the pedagogical quality of AI-

generated responses and refining prompt design based on faculty feedback (see Figure 1).

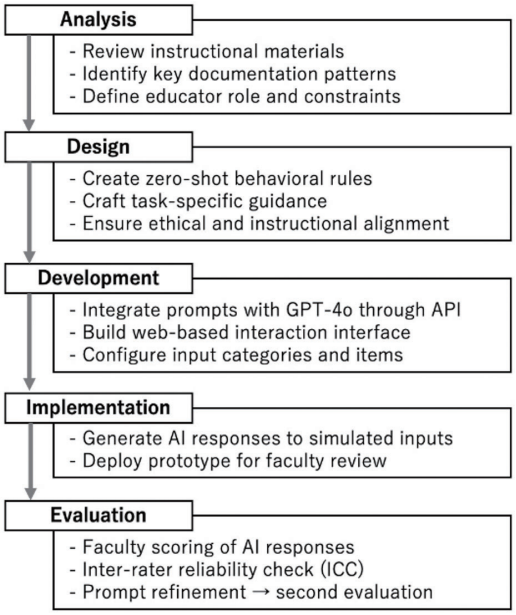

Figure 1. ADDIE-based development and evaluation workflow for zero-shot GenAI assistant

Note. This figure summarizes the instructional design process employed in this study. The Analysis and Design phases involved reviewing instructional materials and constructing zero-shot behavioral rules. The Development encompassed integrating prompts with GPT-4o and building a web interface. During Implementation, AI responses to simulated inputs were generated. Additionally, the Evaluation phase included faculty scoring, inter-rater reliability assessment, and iterative prompt refinement.

2.2 Participants

Participants were nursing faculty members who agreed to participate in the study and were involved in practicum education within the institution where the AI system was intended for implementation. There were no specific inclusion or exclusion criteria regarding teaching experience, academic rank, or clinical background. Faculty were recruited based on their familiarity with the educational context and their willingness to participate in the evaluation of AI-generated feedback.

2.3 Sampling procedures

A purposive sampling strategy was employed to recruit nursing faculty members directly involved in the development and utilization of the Zero-Shot GenAI Assistant for Reflective Record Writing. As the purpose of the study was to evaluate a specific AI system designed with a single standardized prompt, participants were selected from the educational context in which this particular AI tool had been implemented.

2.4 Development of the zero-shot GenAI assistant

An AI-based support system was developed to assist nursing students with reflective clinical record writing during their practicums. This development process consisted of three main stages:

(1) Analysis

Instructional materials routinely used to teach reflective documentation, such as rubrics, classroom worksheets, and exemplar records, were reviewed to identify common student challenges and key learning objectives.

(2) Design

A set of instructional rules was developed to guide AI responses, defining:

- Expected tone and pedagogical stance
- Ethical constraints (e.g., no sample answers, no speculation)
- Structural requirements for practicum documentation
- Reflective prompts to support student reasoning

These instructional constraints provided the foundation for the zero-shot prompting approach. Table 1 summarizes the general behavioral rules applied across all documentation items, whereas Table 2 presents additional section-specific rules developed for the nursing problem statement item, as an illustrative example.

(3) Development

The instructional rules were integrated into a single prompt and implemented using GPT-4o via an API. The system was deployed through a web-based interface, enabling students to receive feedback in a chat-style format (see Figure 2).

Given the length and technical detail of the complete prompt, the full version is provided in Supplementary Appendix A to ensure transparency and reproducibility. A conceptual overview is presented in the main manuscript to maintain readability.

2.5 Data sources

This study utilized two primary sources of data. First, pedagogical evaluation scores were collected from three nursing faculty members who reviewed a total of 126 AI-generated responses. Each response was independently rated on a five-point scale for clarity, relevance, and educational appropriateness. Second, background information on each evaluator, including years of teaching experience, was gathered to support descriptive analysis and explore potential variations in pedagogical judgment. These datasets formed the basis for assessing the instructional quality and consistency of the AI-generated feedback in the subsequent phase.

Table 1. Basic rules on AI behavior

Rules	Description
Positive reinforcement	Begin feedback by acknowledging the student's effort with positive encouragement.
Explanation of patient's characteristics	Provide guidance that reflects the specific characteristics and needs of the relevant patient population in the clinical domain.
Gap identification	Indicate what is missing in the student's entry and how it can be improved.
Character limit	Provide comments within 500 characters (in Japanese) to maintain conciseness and encourage focused revision.
No sample answers	Avoid generating sample answers, rephrasing student input, or offering direct solutions under any circumstances.
Reflective questions	Ask reflective questions that help students reconsider the data they have documented.
No speculation	Avoid prompting students to speculate or fabricate missing information.
Context-bound	Refrain from referencing content outside the designated practicum period (e.g., earlier clinical settings).
Privacy protection	Avoid collecting or encouraging the inclusion of personally identifiable patient information.

Table 2. Zero-shot instruction for "nursing problem statement" section

Instruction area	Details
Expected format	Students were instructed to write nursing problem statements in the format "[causal factor] related to [problem]."
One problem per line	The AI verified that each line contained only one problem. If multiple problems were listed in a single sentence, the AI advised separating them.
Causal factor specificity	If the causal factor was missing, vague, or not found in prior assessment data, the AI prompted students to review their clinical notes instead of making assumptions.
No feedback on rationale	The AI was restricted from commenting on the rationale or clinical importance of the problem, as these were evaluated separately.
Role of AI	The AI's role was to support structural clarity, ensure logical consistency, and preserve the student's authorship and reasoning.

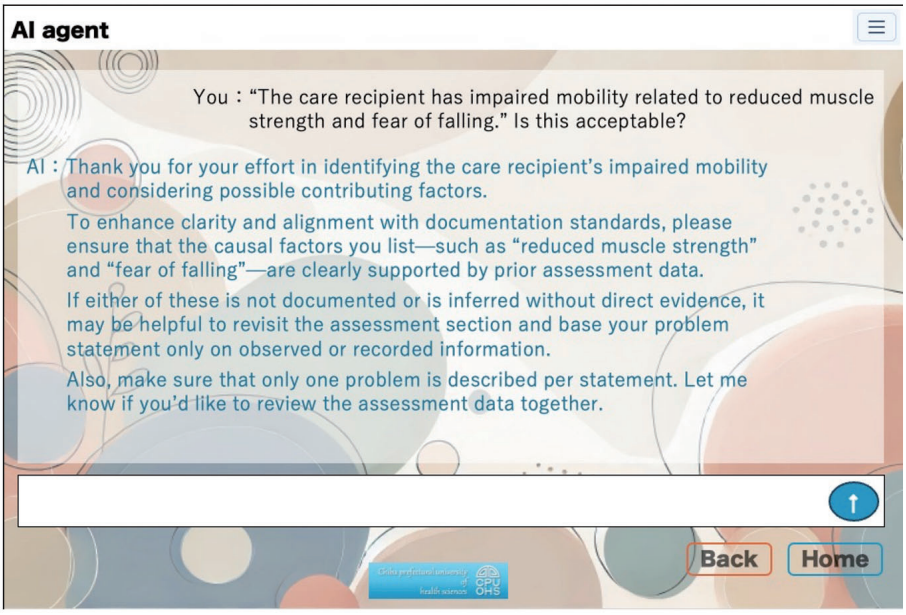

Figure 2. Screenshot of the AI interface for clinical practicum support

Note. This web-based interface allows students to interact with an AI system via a chat-style format that is accessible on both desktop and mobile devices. The system responds to student input with feedback aligned with documentation standards and pedagogical guidance.

2.6 Evaluation procedures and data analysis

To evaluate the educational appropriateness of the AI-generated feedback, a structured review process was conducted by the participating faculty members. This section outlines the design and procedures used for assessing the system's output, including the categorization of input items, the rating methodology, and the iterative refinement process based on faculty review.

The system's output was then independently reviewed by participants. The AI was designed to support six major categories of input, which corresponded to the five structured clinical practicum documentation forms (e.g., assessment sheets, nursing problem lists, care plans) and one general category for questions about practicum flow. These six categories contained a total of 18 individual items (e.g., "long-term goals," "short-term goals," "nursing interventions"), each of which was targeted by three types of simulated student questions. These items reflected three distinct categories:
(1) items that followed the expected documentation rules (e.g., the correct format for nursing problem statements);
(2) items that violated structural or content rules (e.g., missing causal factors or multiple problems listed in one sentence);
(3) items that vaguely asked for guidance (e.g., "What should I write here?") without referencing prior assessment data.

Initially, a total of 54 simulated student inputs were constructed by the research team, corresponding to three types of representative questions for each of the 18 documentation items. These examples were carefully developed based on common issues and recurring patterns observed during practicum instruction over the past three years. As the AI system occasionally responded with follow-up questions (e.g., asking the student to clarify their reasoning, requesting additional context), some interactions evolved into multi-turn dialogues. Each resulting student–AI exchange (consisting of student input followed by a complete AI response) was counted as one unit for evaluation. The 126 units fell into the three distinct categories discussed above: 57 followed the expected documentation rules, 37 violated structural or content conventions, and 31 vaguely requested guidance without reference to prior assessment data.

Each student–AI interaction unit was evaluated individually rather than as a full conversation thread. This approach was taken to account for the possibility that AI responses may vary in consistency across separate turns. Evaluating each interaction independently also allowed reviewers to detect issues within a single exchange, even if the rest of the dialogue appeared appropriate. Each AI-generated response was evaluated on a five-point scale (1 = lowest pedagogical quality, 5 = highest pedagogical quality). Faculty members were instructed to assign one overall score per response based on holistic judgment. The scale represented subjective impressions of pedagogical appropriateness, with 1 indicating "not appropriate" and 5 indicating "highly appropriate." For each response, the lowest score among the three faculty evaluations was adopted as the final rating. This conservative approach was employed to ensure that any potential concerns regarding pedagogical validity were not overlooked. In the context of educational applications, the identification of a pedagogical flaw by even a single evaluator was deemed sufficient grounds for prompt revision, reflecting the priority placed on minimizing instructional risk in formative feedback environments. Evaluations were conducted using the following three pedagogical criteria as reference points:
(1) Clarity: Was the language understandable to undergraduate nursing students?
(2) Relevance: Did the response appropriately address the student's original input and align with the intended learning objective?
(3) Educational appropriateness: Did the feedback encourage reflection and independent reasoning without overguiding or giving answers?

These criteria were used holistically to determine a single score per response, reflecting the overall pedagogical quality. In cases where evaluators assigned different scores to the same response, the lower score was adopted to maintain a conservative and rigorous standard.

Based on the faculty evaluations, the prompts were systematically revised to better align with educational goals and ethical expectations. Specifically, ambiguous wording was rewritten, and new behavioral constraints were added to prevent overguidance. This study adopted a threshold consistent with the commonly accepted standard for content validation—an average Content Validity Index (CVI) score of 0.80 or higher.[17] Accordingly, all AI-generated responses that did not receive the highest rating (a score of 5) were subject to revision. This decision also aimed to prevent overfitting in educational applications, where maintaining both prompt flexibility and instructional integrity is essential. Prompt revision was considered complete when more than 80% of responses were rated 5 and none received a rating below 3. Following these refinements, a second round of reviews was conducted using the same criteria, confirming improved clarity, consistency, and alignment of the AI output with student learning needs. To evaluate inter-rater reliability among faculty scorers, the intraclass correlation coefficient (ICC) was calculated using a two-way random-effects model with absolute agreement, implemented in SPSS version 29.

2.7 Ethical considerations

This study involved the voluntary participation of three nursing faculty members with expertise in clinical education. They were recruited to evaluate the pedagogical appropriateness of AI-generated feedback and provided informed consent prior to participation. No personal or identifying information was included in any of the simulated student inputs or AI responses. The study did not involve any risks or conflicts of interest for participants. The study protocol was reviewed and approved by the institutional ethics committee (Approval No. 2025–04).

3. Results

A total of 126 AI-generated responses were evaluated by the three nursing faculty members using the five-point scale based on the criteria of clarity, relevance, and educational appropriateness. The participating faculty had an average of 7 years of teaching experience (SD = 6.3).

In the initial review, 87 responses (69.0%) received a score of 5, indicating the highest level of pedagogical quality. Twenty-eight responses (22.2%) were rated 4, and 11 responses (8.7%) received a score of 3. Notably, no responses were rated 1 or 2, suggesting that none of the AI-generated feedback was perceived by faculty as clearly inappropriate or pedagogically harmful. Items rated 3 were typically characterized by vague phrasing, limited contextual relevance, or insufficient guidance to facilitate student reflection. While not harmful, these responses were considered pedagogically suboptimal. In accordance with the predefined evaluation threshold—requiring at least 80% of responses rated 5 and none below 3—prompt revision was undertaken. All responses scoring below 5 were systematically reviewed and revised based on faculty feedback, which included refining ambiguous phrasing, enhancing pedagogical alignment, and reinforcing ethical constraints.

Following prompt refinement, a second evaluation was conducted using the same criteria. In this review, 121 responses (96.0%) were rated 5, demonstrating a substantial improvement in overall quality. The remaining responses included three rated at 4 (3.2%) and one rated at 3 (0.8%). This notable enhancement confirmed that even minor adjustments to prompt wording or structure had measurable effects on the quality of AI-generated feedback. The continued absence of low-rated responses (scores of 2 or below) across both evaluations further suggests that the prompt design maintained a consistent baseline of educational adequacy. A visual summary of these results is provided in Table 3, which shows the distribution of evaluation scores before and after the prompt revision.

To assess the consistency of faculty evaluations, the ICC was calculated. The ICC for the overall ratings across the three faculty members was 0.58 (95% CI: 0.45–0.70), indicating moderate inter-rater reliability.

Table 3. Improvement in highest evaluation score

Practicum documentation form	Initial evaluation (%)	Second evaluation (%)
Practicum flow & timing of documentation	55.6	77.8
Form 1: Assessment sheet	61.1	94.4
Form 2: Patient overview	89.7	100.0
Form 3: Nursing problem list	57.1	100.0
Form 4: Nursing care plan	78.6	96.4
Form 5: Daily records	52.4	95.2

Note. This table shows the proportion of the highest evaluation scores (5) for each practicum documentation form. Each form consists of several sub-items, and the results shown here represent the aggregated outcomes.

4. Discussion

This study should be interpreted as a pilot instructional evaluation examining whether a zero-shot prompting approach could feasibly generate pedagogically appropriate feedback within clinical practicum education. Rather than demonstrating generalizable effectiveness, the findings provide preliminary insight into how carefully designed zero-shot instructions can guide a generative AI system to produce outputs aligned with faculty expectations. These exploratory results highlight the potential of zero-shot AI tools to contribute to instructional support in nursing education while underscoring the need for further validation with authentic student data and more diverse faculty reviewers.

The results offer early evidence that carefully engineered zero-shot prompts can generate output that aligns with course goals and ethical standards, even without domain-specific training data. This suggests that educators—regardless of programming expertise—can construct prompts that provide individualized, pedagogically consistent guidance while preserving student autonomy and promoting reflective thinking.

These preliminary findings also resonate with recent literature demonstrating the scalability and flexibility of zero-shot prompting in educational contexts.[12] Importantly, all prompts in this study were designed by educators rather than AI specialists, reinforcing the accessibility and transferability of this approach to various teaching contexts.

Moreover, the notable improvements observed following iterative prompt refinement demonstrate the value of a feedback-driven design cycle. Minor adjustments, such as clarifying ambiguous phrasing or reinforcing ethical boundaries, resulted in substantial improvements in clarity, relevance, and instructional value. The decision to conclude refinement after a single iteration was guided by predefined content-validation criteria ($\geq$ 80% of responses rated 5; none rated below 3), which is in line with established validation methodologies.[17] This process underscores the critical role of faculty expertise in shaping AI-assisted instruction and aligns with previous research highlighting the need for domain-expert involvement to ensure pedagogical integrity.[18]

The advantages of the zero-shot approach, particularly its scalability, adaptability, and minimal resource requirements, were evident in this pilot implementation. Unlike few-shot prompting, which requires curated examples, or fine-tuning, which requires substantial technical infrastructure and annotated datasets, zero-shot prompting can be quickly deployed and easily modified by educators.[19] Although these findings are promising, they should be interpreted cautiously because this study represents a pilot evaluation rather than a comprehensive assessment across diverse instructional contexts.

A key limitation of this study is that all student inputs were simulated rather than derived from real learners. While this approach provided consistency for the initial evaluation, it limits the ecological validity of the findings. Authentic student documentation often exhibits greater ambiguity, contextual complexity, and variability than simulated inputs can capture. Consequently, a logical next step is to examine the system's responsiveness, clarity, and ethical robustness when interacting with genuine student submissions in real practicum settings.

Another important consideration is the ethical structure embedded within the zero-shot prompts. Constraints prohibiting direct answers, speculative advice, or breaches of confidentiality demonstrate how pedagogical and ethical principles can be operationalized within AI outputs. These pilot findings suggest that carefully designed prompts may help ensure AI systems function as responsible collaborators, supporting (rather than replacing) human instruction and promoting critical thinking, professional reasoning, and individualized learning experiences.

Furthermore, the moderate inter-rater reliability observed in this study (ICC = 0.58) should be interpreted within the context of this pilot study. In this regard, the purpose was not to establish standardized scoring but to explore how faculty interpret AI-generated feedback in relation to instructional goals. Meanwhile, variation among evaluators is expected in nursing education, where pedagogical judgments are shaped by professional background, teaching philosophy, and clinical expertise.[20] In this context, the observed variability may reflect the richness of faculty perspectives rather than a methodological limitation.

Looking ahead, the next phase of research should examine the system's use with actual nursing students, with attention to usability, clarity, and the impact on reflective practice and clinical reasoning. Further work should also expand evaluation across multiple institutions, incorporate diverse faculty perspectives, and refine prompt design based on real-world implementation. Additional enhancements, such as personalization mechanisms or integration with longitudinal learning analytics, could provide deeper insights into how AI tools support student growth over time.

Challenges and limitations

This study has several limitations. First, all student inputs were simulated by faculty members rather than actual nursing students. Although these inputs were constructed to approximate common patterns of student documentation, they may introduce implicit bias and do not fully capture authentic student reasoning processes. As a result, the findings should be interpreted as preliminary until they are validated with real student interactions.

Second, the evaluation involved only three faculty members from a single institution. This small, institution-specific sample limits the generalizability of the results. Future research should incorporate a larger and more diverse group of educators across multiple nursing programs to strengthen external validity.

Third, the current system was applied only to practicum documentation forms, which served as a proof-of-concept domain. Its applicability to other areas of nursing education—such as clinical reasoning assignments, simulation debriefing, or classroom-based learning activities—has yet to be examined.

Finally, although the zero-shot strategy offers scalability and reduces the need for technical customization, it may face challenges when responding to highly individualized or complex inputs. Additional design supports, such as iterative prompt refinement or educator oversight mechanisms, may be necessary to improve reliability in these contexts.

To guide future development, a structured research roadmap

is essential. The next phase should involve evaluation with authentic student submissions to assess the system's responsiveness, clarity, and pedagogical safety in real-world conditions. Multi-institutional implementation across diverse nursing programs will be critical for assessing external validity and contextual adaptability. Furthermore, future studies should also evaluate learning outcomes, such as improvements in reflective reasoning, documentation quality, and practicum performance, to determine the educational effectiveness of AI-supported feedback. As the system progresses toward real-world deployment, ongoing refinement of safety and ethical guardrails will be essential to ensure that AI-generated guidance remains aligned with institutional policies, professional standards, and evolving norms for responsible AI use.

5. Conclusions

This pilot instructional evaluation demonstrated that a zero-shot prompting approach can generate AI-based feedback that is educationally appropriate, reflective, and aligned with clinical learning objectives. Although the results offer preliminary insight into how generative AI may emulate the instructional stance of expert educators, the findings should be interpreted as exploratory rather than generalizable. Iterative refinement informed by faculty input produced substantial improvements in pedagogical quality, underscoring the value of educator involvement in shaping AI behavior. Even minor prompt adjustments—such as clarifying ambiguous phrasing or reinforcing ethical constraints—yielded measurable gains in clarity and relevance, highlighting the importance of a feedback-driven design process. These early results suggest that zero-shot prompting may serve as a practical and accessible preparatory phase for the responsible integration of AI into nursing education, enabling educators to customize AI guidance without programming expertise or large datasets. Future work should examine system performance with real nursing students in authentic practicum environments to evaluate its usability, effectiveness, and adaptability to individual learning contexts because this is essential for determining how AI-supported feedback can evolve into a sustainable, ethically aligned component of clinical education.

Acknowledgements

Not applicable.

Authors contributions

Dr. Asahiko Higashitsuji conceptualized the study, designed the AI-assisted educational model, conducted the data collection and analysis, and drafted the manuscript. Dr. Tomoko Otsuka contributed to theoretical refinement, data collection, and critically revised the manuscript. All authors read and approved the final version.

Funding

Not applicable.

Conflicts of Interest Disclosure

The authors declare that they have no competing financial or personal interests that could have influenced the work reported in this article.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Association for Health Sciences and Education. The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

References

[1] Chen L, Kong D, Zhang S, et al. A quasi-experimental study of specialized training on the clinical decision-making skills and social problem-solving abilities of nursing students. Contemp Nurse. 2021; 57(1-2): 4-12. PMid:33820488 https://doi.org/10.1080/10376178.2021.1912616

[2] Zhang J, Shields L, Ma B, et al. The clinical learning environment, supervision and future intention to work as a nurse in nursing students: a cross-sectional and descriptive study. BMC Med Educ. 2022; 22: 548. PMid:35841091 https://doi.org/10.1186/s12909-022-03609-y

[3] Bjerkvik LK, Hilli Y. Reflective writing in undergraduate clinical nursing education: a literature review. Nurse Educ Pract 2019; 35: 32-41. PMid:30660960 https://doi.org/10.1016/j.nepr.2018.11.013

[4] Reierson IÅ, Mofossbakke RG, Solli H. Nursing students' perspectives on learning electronic health record documentation in an academic setting and during clinical placement: a qualitative study. BMC Nurs. 2025; 24: 764. PMid:40597200 https://doi.org/10.1186/s12912-025-03320-5

[5] Shen M, Shen Y, Liu F, et al. Prompts, privacy, and personalized learning: integrating AI into nursing education—a qualitative study. BMC Nurs. 2025; 24: 470. PMid:40301862 https://doi.org/10.1186/s12912-025-03115-8

[6] Khlaif ZN, Salameh N, Ajouz M, et al. Using Generative AI in nursing education: students' perceptions. BMC Med Educ. 2025; 25: 926. PMid:40598139 https://doi.org/10.1186/s12909-025-07416-z

[7] Saban M, Dubovi I. A comparative vignette study: evaluating the potential role of a generative AI model in enhancing clinical decision-making in nursing. J Adv Nurs. 2024; 81(11): 7489-99. PMid:38366690 https://doi.org/10.1111/jan.16101

[8] Rony MKK, Ahmad S, Tanha SM, et al. Nursing educators' perspectives on the integration of artificial intelligence into academic settings. SAGE Open Nurs. 2025; 11. PMid:40386172 https://doi.org/10.1177/23779608251342931

[9] Yampolskiy RV. On the controllability of artificial intelligence: an analysis of limitations. J Cyber Secur Mob. 2022; 11(3): 321-403. https://doi.org/10.13052/jcsm2245-1439.1132

[10] Ogunleye B, Zakariyyah KI, Ajao O, et al. A systematic review of generative AI for teaching and learning practice. Educ Sci. 2024; 14(6): 636. https://doi.org/10.3390/educsci14060636

[11] Chan CKY, Tsi LHY. Will generative AI replace teachers in higher education? A study of teacher and student perceptions. Stud Educ Eval. 2024; 83: 101395. https://doi.org/10.1016/j.stueduc.2024.101395

[12] Hussein R, Yahya N, Chew HSJ, et al. Ethical and social considerations of applying artificial intelligence in healthcare: a two-pronged scoping review. BMC Med Ethics. 2025; 26: 68. PMid:40420080 https://doi.org/10.1186/s12910-025-01198-1

[13] Abdelwanis M, Simsekler MCE, Gabor AF, et al. Artificial intelligence adoption challenges from healthcare providers' perspectives: a comprehensive review and future directions. Saf Sci. 2026; 193: 107028. https://doi.org/10.1016/j.ssci.2025.107028

[14] Ramesh G, Sahil M, Palan SA, et al. A review on NLP zero-shot and few-shot learning: methods and applications. Discov Appl Sci. 2025; 7(9): 2025. https://doi.org/10.1007/s42452-025-07225-5

[15] Pan R, García-Díaz JA, Valencia-García R. comparing fine-tuning, zero and few-shot strategies with large language models in hate speech detection in English. CMES 2024; 140(3): 2849-68. https://doi.org/10.32604/cmes.2024.049631

[16] Branch RM. Instructional design: the ADDIE approach. New York: Springer; 2009. https://doi.org/10.1007/978-0-387-09506-6

[17] Yusoff MSB. ABC of content validation and content validity index calculation. Educ Med J. 2019; 11(2): 49-54. https://doi.org/10.21315/eimj2019.11.2.6

[18] Holmes W, Bialik M, Fadel C. Artificial intelligence in education: promises and implications for teaching and learning. Boston, MA, USA: Center for Curriculum Redesign; 2019.

[19] Kojima T, Gu SS, Reid M, et al. Large language models are zero-shot reasoners. arXiv preprint arXiv:2205.11916. 2022. Available from: https://arxiv.org/abs/2205.11916

[20] Okechukwu C. Exploring the concepts of diversity, equity and inclusion in nursing curricula. J Nurs Educ Pract. 2024; 14(3): 1-7.