Research Square

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

AI-Assisted BOPPPS Model in Nursing Teaching Rounds: A Quasi-Experimental Study

Xuemei Liu
Second Affiliated Hospital of Guangzhou Medical University
Haiping Chen
Second Affiliated Hospital of Guangzhou Medical University
Xuelian Wu
Second Affiliated Hospital of Guangzhou Medical University
Hongbo Yan
Second Affiliated Hospital of Guangzhou Medical University
Jian Song
Second Affiliated Hospital of Guangzhou Medical University
Chang Zhang
Second Affiliated Hospital of Guangzhou Medical University
Bing Wang
Second Affiliated Hospital of Guangzhou Medical University
Runfang Zhong
Second Affiliated Hospital of Guangzhou Medical University
Hongjing Yu
yuhongjing@163.com

Second Affiliated Hospital of Guangzhou Medical University

Research Article

Keywords: Artificial intelligence, BOPPPS model, Clinical judgment, Nursing education, Teaching rounds

Posted Date: September 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10653619/v1

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

Additional Declarations: No competing interests reported.

Abstract

Background

Traditional nursing teaching ward rounds rely heavily on instructors' manual preparation of case materials, leading to excessive preparation time and limited scaffolding for higher-order thinking. While the BOPPPS model provides a structured instructional framework, its clinical implementation remains time-consuming. Generative AI offers potential for automating routine teaching tasks; however, no empirical study has fully integrated AI across all six BOPPPS stages in clinical nursing ward rounds.

Methods

A total of 47 clinical nursing instructors and 63 nursing interns were enrolled. The control group (n = 23 instructors, 26 interns) received conventional BOPPPS-based ward rounds, while the intervention group (n = 24 instructors, 37 interns) received AI-assisted BOPPPS-based ward rounds, with AI support embedded across all six stages. Outcome measures included the Instructor Teaching Ward Round Evaluation Scale, the Student Teaching Ward Round Evaluation Scale, and the Lasater Clinical Judgment Rubric (LCJR). Data were analyzed using independent t-tests, χ^2 tests, and Kruskal-Wallis H tests.

Results

Intervention group instructors reported significantly higher scores on four evaluation items: clarity of objectives and reduced preparation time ($P = 0.010$), timeliness and depth of discussion guidance ($P = 0.008$), perceived clinical practice value ($P = 0.029$), and overall satisfaction ($P = 0.033$), with effect sizes ranging from medium to large (Cohen's d = 0.65–0.81). Interns in the intervention group scored significantly higher across all four LCJR dimensions compared with the control group (all $P < 0.001$; Cohen's d = 1.05–1.31), indicating large effects. Student satisfaction did not differ significantly between groups (all $P > 0.05$).

Conclusion

The AI-assisted BOPPPS model reduces instructors' perceived workload and improves interns' clinical judgment with large effect sizes. However, these benefits do not automatically transfer to students' satisfaction, underscoring the need for deliberate pedagogical translation. The model offers a feasible, low-cost approach for clinical teaching; yet successful implementation requires faculty development to transform AI-generated content into interactive, dialogic teaching.

Introduction

Nursing teaching ward rounds constitute a fundamental pedagogical activity in clinical nursing education, designed to cultivate nursing students' clinical reasoning through systematic discussion of patient cases [1]. In many clinical settings, however, instructors must reconcile heavy clinical workloads with the time-consuming manual preparation of teaching materials. This dual burden frequently results in excessive case preparation time, insufficient scaffolding for learner engagement, and an absence of structured mechanisms for promoting higher-order thinking [2].

The BOPPPS model—comprising Bridge-in, Objectives, Pre-assessment, Participatory Learning, Post-assessment, and Summary—provides a structured framework for instructional design in nursing education [3]. In teaching ward rounds, BOPPPS facilitates case discussion structuring and formative assessment integration [4]. Nevertheless, its implementation continues to depend on instructors' manual case preparation, question formulation, and performance tracking—processes that remain time-consuming and cognitively demanding within busy clinical environments [3, 4]. As a teaching hospital, our institution has already applied the BOPPPS model to its teaching ward rounds.

Recent advances in generative artificial intelligence offer new possibilities for automating routine instructional tasks. Emerging evidence suggests that AI-assisted teaching rounds can enhance nursing students' clinical competency and critical thinking [5]. However, this previous work employed ChatGPT as a supplementary tool without embedding it within a structured pedagogical framework, leaving the optimal integration model unclear. The present study addresses this gap by systematically integrating a generative AI assistant (DeepSeek) across all six stages of the BOPPPS model and evaluating its effects on instructors' teaching experience, interns' clinical judgment, and student satisfaction through a quasi-experimental design.

Methods

Study Design

This study employed a two-group parallel quasi-experimental controlled design. A randomized controlled trial was not feasible for two reasons. First, the hospital's internship rotation schedule required interns to remain within their assigned clinical departments throughout each rotation, precluding individual-level randomization without disrupting routine teaching operations. Second, cluster allocation by department was necessary to prevent cross-group contamination, as interns and instructors within the same department share clinical workspaces and could inadvertently exchange teaching content if assigned to different conditions. The study was reported in accordance with the STROBE guidelines for observational studies, with adaptations for the quasi-experimental design. Intervention and data collection occurred between November 2025 and April 2026.

Participants and Setting

The study was conducted at a tertiary teaching hospital in China with an established nursing internship program accommodating approximately 150 interns annually. A total of 47 clinical nursing instructors and 63 nursing interns were enrolled.

Inclusion and Exclusion Criteria

Instructors were eligible if they held a valid nursing license and hospital faculty certification and provided written consent; those unable to complete the 6-month intervention due to leave were excluded. Interns were eligible if enrolled in a full-time associate or baccalaureate nursing program and provided consent; those with incomplete rotations during the intervention were excluded.

Participant Allocation

Instructors were allocated by clinical department to prevent cross-contamination. Interns were assigned to corresponding groups according to their rotation schedules.

Blinding

Instructors could not be blinded. Interns were not explicitly informed of group assignment. Data analysts were blinded.

Sample Recruitment

Convenience sampling over the 6-month window enrolled all eligible participants (47 instructors, 63 interns). No a priori power calculation was performed. Post-hoc power analyses were conducted using G*Power (version 3.1.9.7; Heinrich Heine University Düsseldorf) for the primary outcome measure (LCJR) to evaluate the achieved statistical power given the observed effect sizes and sample sizes. For the LCJR Interpretation dimension (Cohen's $d = 1.31$, $n_1 = 26$, $n_2 = 37$, two-tailed $\alpha = 0.05$), the achieved power exceeded 0.95. For the other three LCJR dimensions (Observation, Response, and Reflection), achieved power values ranged from 0.92 to 0.98. These post-hoc estimates suggest that, despite the absence of a priori sample size planning, the current sample provided sufficient statistical power to detect the large effect sizes observed in the primary outcomes.

Intervention

Development of the AI-Assisted BOPPPS Process

The protocol was developed through four phases: (1) preliminary framework based on BOPPPS and literature; (2) expert review by seven specialists; (3) pilot testing with two instructors and six interns (not in the sample) to refine prompts and timing; and (4) finalization of a manual specifying AI functions, role

division, and quality standards. Intervention instructors received 30-minute training on prompt formulation and critical evaluation of AI outputs, followed by supervised simulation; competency was required before implementation.

Intervention Protocols

Control Group: Conventional BOPPPS-Based Teaching Ward Round

Instructors in the control group delivered teaching ward rounds following the standard BOPPPS framework without AI assistance. At the bridge-in stage, they selected a typical case and manually compiled case summaries. For objectives, they manually formulated two to three learning objectives and three to five discussion questions, distributing these materials electronically two days in advance. Pre-assessment consisted of a 10-item quiz administered via Wenjuanxing, with instructors manually calculating accuracy rates. During the participatory learning phase, instructors led group discussions, allocating 50% of session time to this activity. Immediately following the discussion, interns completed a 10-item post-session quiz via Wenjuanxing, which instructors manually reviewed and scored. In the summary phase, instructors orally recapped key learning points, highlighted common errors observed during discussion, and assigned take-home reflection questions.

Intervention Group: AI-Assisted BOPPPS-Based Teaching Ward Round

Instructors in the intervention group delivered teaching ward rounds following an AI-enhanced BOPPPS framework. At the bridge-in stage, they independently selected a typical case and structured the case information around core nursing problems, potential risks, and decision-making conflicts (no AI assistance at this stage). For objectives, instructors submitted the structured case materials to the AI assistant (DeepSeek) with a standardized prompt requesting three hierarchically progressive teaching questions and three measurable learning objectives based on Bloom's taxonomy. For pre-assessment, instructors administered a 10-item quiz via Wenjuanxing and submitted the aggregated results to the AI assistant, which generated dimension-specific accuracy rates (basic knowledge, clinical assessment, emergency response) and identified specific knowledge points with accuracy rates below 60% for targeted review. During the participatory learning phase, interns individually articulated their clinical reasoning and decision-making processes. The instructor de-identified key reasoning points and entered them into the AI assistant, which critiqued reasoning chains for logical gaps, evidence deficiencies, or cognitive biases (analysis output time: approximately 5 minutes). The instructor then facilitated a structured debate and verification session in which students compared their own reasoning with the AI-generated critique. For post-assessment, instructors administered the post-session quiz and submitted individual student results to the AI assistant, which generated one variant extension question of the same type but a different clinical context for each incorrect answer as a take-home consolidation task. In the summary phase, instructors submitted the aggregated post-test results and the session's learning objectives to the AI assistant, which synthesized a summary of overall knowledge mastery, frequently missed concepts, common clinical reasoning deficiencies, evaluated the extent to which session objectives were achieved, and provided structured recommendations for teaching optimization.

Fidelity and Quality Assurance

To ensure intervention fidelity, the research leader conducted monthly unannounced observations of at least one teaching ward round per instructor, using a standardized fidelity checklist based on the BOPPPS framework. The checklist comprised nine essential components: (1) case structure submission (Bridge-in), (2) AI-generated hierarchical objectives (Objectives), (3) pre-assessment data submitted to AI for weak-point identification (Pre-assessment), (4) AI-assisted reasoning critique during discussion (Participatory Learning), (5) AI-generated post-test variant questions (Post-assessment), (6) AI-synthesized summary and recommendations (Summary), (7) student participatory learning accounting for ≥ 50% of session time, (8) AI content validation by instructor before use, and (9) de-identification of patient and student data before AI input. Each checklist item was scored as "fully adhered," "partially adhered," or "not adhered."

Over the 6-month study period, 32 unannounced observations were completed across the 24 intervention instructors. The overall fidelity adherence rate was 93.75%. The only missed item was ensuring student participatory learning accounted for ≥ 50% of session time, which occurred on three occasions.

For AI content quality assurance, all AI-generated outputs were validated by instructors prior to classroom use. Four outputs (less than 1%) were rejected and regenerated due to clinical inaccuracies. This review process ensured that only clinically accurate and pedagogically appropriate AI content was delivered to students.

Outcome Measures

All outcomes were administered immediately after the teaching ward round session (Month 6).

Self-designed Evaluation of Nursing Teaching Ward Round Effectiveness (Instructor Version): 14 items (5-point Likert), two domains. Cronbach's α = 0.89; I-CVI ≥ 0.83.

Self-designed Evaluation of Nursing Teaching Ward Round Effectiveness (Student Version): 17 items (5-point Likert), four dimensions. Cronbach's α = 0.92; I-CVI ≥ 0.80.

Lasater Clinical Judgment Rubric (LCJR) [6]: 4 dimensions (Observation, Interpretation, Response, Reflection), each rated 1–4 (total 4–16). Instructors evaluated each intern.

Data Collection and Statistical Analysis

Questionnaires were administered electronically via Wenjuanxing, a Chinese online survey platform. Responses with a completion time of less than one minute or showing identical answers across all items were predefined as invalid and excluded from analysis; no such responses were identified.

Statistical analyses were performed using SPSSAU version 24.0 (Qingxi Technology, China). Continuous variables that followed a normal distribution, as assessed by the Shapiro–Wilk test, were summarized as

means and standard deviations (SD) and compared between groups using independent-samples t-tests. Categorical variables were compared using Pearson's chi-square (χ^2) tests. For subgroup analyses stratified by teaching experience (three levels: 1–5 years, 6–10 years, and > 10 years), the Kruskal–Wallis H test was applied due to the ordinal nature of the grouping variable and non-normal distributions within subgroups.

Satisfaction data exhibited ceiling effects (medians at the scale maximum in both groups) and non-normal distributions. To ensure robustness, sensitivity analyses were conducted using both independent t-tests and Mann–Whitney U tests; the results were consistent across both approaches, and t-test results are reported for ease of interpretation.

For the LCJR, internal consistency was assessed using Cronbach's α (α = 0.89 for the instructor scale, α = 0.92 for the student scale). Content validity was evaluated using the item-level content validity index (I-CVI), with values ≥ 0.80 considered acceptable. Inter-rater reliability for the thematic analysis of open-ended responses was assessed using Cohen's κ, yielding a value of 0.86, indicating almost perfect agreement.

Effect sizes for group comparisons were calculated using Cohen's d, with values of 0.20, 0.50, and 0.80 interpreted as small, medium, and large effects, respectively. All statistical tests were two-tailed, with a significance level set at α = 0.05. Exact P values are reported to three decimal places, except where P < 0.001, which is reported as such.

Results

All 47 instructor questionnaires and 63 student questionnaires were valid, yielding a 100% response rate with no missing data.

Baseline Characteristics

No statistically significant differences were observed between the control and intervention groups at baseline. Among instructors, groups were comparable in professional title (χ^2 = 0.179, P = 0.672), teaching experience (χ^2 = 1.113, P = 0.573), and pre-intervention AI literacy (4.30 ± 0.74 vs. 4.00 ± 0.91, P = 0.231). Among interns, no significant between-group differences were found in education level (χ^2 = 1.063, P = 0.303), gender (χ^2 = 0.524, P = 0.469), or institution type (χ^2 = 3.801, P = 0.284).

Instructor Evaluation

Intervention instructors scored significantly higher on four items: clarity of objectives and reduced preparation time (t = 2.691, P < 0.01), timeliness and depth of discussion guidance (t = 2.760, P < 0.01), perceived clinical practice value (t = 2.296, P < 0.05), and overall satisfaction (t = 2.225, P < 0.05). Cohen's d values ranged from 0.65 to 0.79. No differences were found for the other five items (all P > 0.05; Table 1).

Table 1
Instructor evaluation of teaching ward rounds [mean ± SD]

Item	Control (n = 23)	Intervention (n = 24)	*t*	*P*
Clear objectives, reduced preparation time	3.67 ± 0.96	4.26 ± 0.45	2.691	0.010**
Timely and in-depth discussion guidance	3.67 ± 0.96	4.30 ± 0.56	2.760	0.008**
Precise professional feedback, enhanced student thinking	3.71 ± 0.86	4.04 ± 0.47	1.665	0.105
Students adequately prepared, good case knowledge	3.58 ± 0.83	4.00 ± 0.74	1.816	0.076
Active participation, constructive contributions	3.58 ± 0.97	3.74 ± 0.69	0.631	0.532
Collaborative team atmosphere, mutual respect	3.88 ± 1.12	4.30 ± 0.70	1.571	0.123
Clinical practice value of the ward round	3.67 ± 1.01	4.17 ± 0.39	2.296	0.029*
Appropriate time arrangement and pace	3.63 ± 1.10	3.74 ± 0.92	0.387	0.701
Overall satisfaction	3.63 ± 0.92	4.09 ± 0.42	2.225	0.033*
* p < 0.05 ** P < 0.01				

Subgroup Analysis by Teaching Experience

Kruskal–Wallis tests showed significant differences in "student preparedness" (H = 8.111, P < 0.05) and "overall satisfaction" (H = 6.349, P < 0.05), with instructors having > 10 years of experience reporting lower median scores (3.000) than the 1–5-year and 6–10-year groups (4.000). No other items differed significantly across groups (Table 2).

Table 2
Instructor evaluation by years of teaching experience [M (P25, P75)]

Item	1–5 years (n = 8)	6–10 years (*n* = 30)	>10years (*n* = 9)	H	P
Clear objectives, reduced preparation time	4.000(4.0,4.0)	4.000(4.0,4.0)	4.000(3.0,4.0)	4.404	0.111
Timely and in-depth discussion guidance	4.000(4.0,5.0)	4.000(4.0,4.0)	4.000(3.0,4.0)	5.426	0.066
Precise professional feedback, enhanced student thinking	4.000(4.0,4.8)	4.000(4.0,4.0)	4.000(3.0,4.0)	4.912	0.086
Students adequately prepared, good case knowledge	4.000(4.0,4.8)	4.000(3.0,4.0)	3.000(3.0,4.0)	8.111	0.017*
Active participation, constructive contributions	4.000(3.3,4.8)	4.000(3.0,4.0)	3.000(2.5,4.0)	2.606	0.272
Collaborative team atmosphere, mutual respect	5.000(4.3,5.0)	4.000(4.0,5.0)	4.000(3.5,4.0)	5.875	0.053
Clinical practice value of the ward round	4.000(4.0,4.8)	4.000(4.0,4.0)	4.000(3.0,4.0)	3.656	0.161
Appropriate time arrangement and pace	4.000(3.3,5.0)	4.000(3.0,4.0)	3.000(2.5,4.0)	2.820	0.244
Satisfaction	4.000(4.0,4.0)	4.000(4.0,4.0)	3.000(3.0,4.0)	6.349	0.042*

* $p < 0.05$ ** $p < 0.01$

LCJR Scores

Intervention interns scored significantly higher across all four LCJR dimensions (Table 3): Observation (t = 3.849, P < 0.001), Interpretation (t = 4.394, P < 0.001), Response (t = 4.406, P < 0.001), and Reflection (t = 3.538, P < 0.001). Cohen's d values ranged from 1.05 to 1.31, indicating large effects.

Table 3
Instructor evaluation of interns' clinical judgment using LCJR [mean ± SD]

Dimension	Control(*n* = 23)	Intervention(*n* = 24)	*t*	*P*
Noticing	6.38 ± 1.91	8.35 ± 1.58	3.849	<0.001
Interpreting	4.13 ± 1.39	5.65 ± 0.93	4.394	
Responding	9.33 ± 2.50	12.00 ± 1.57	4.406	
Reflecting	4.42 ± 1.28	5.52 ± 0.79	3.538	
Note. LCJR total score ranges from 4 to 16. Cohen's d ≥ 0.80 indicates large effects.				

Student Satisfaction

No significant differences were found between groups on any satisfaction dimension (Mann–Whitney U, all P > 0.05; Table 4).

Table 4
Student evaluation of teaching ward rounds [M (P25, P75)]

Item	Control (n = 26)	Intervention (n = 37)	*U*	*z*	*P*
Preparation	15.000(14.0,15.0)	15.000(13.0,15.0)	452.500	-0.469	0.639
Implementation	15.000(14.0,15.0)	15.000(13.5,15.0)	456.000	-0.407	0.684
Teaching interaction	20.000(20.0,20.0)	20.000(20.0,20.0)	447.000	-0.746	0.455
Learning gains	25.000(23.8,25.0)	25.000(24.0,25.0)	476.500	-0.074	0.941
Satisfaction	5.000(5.0,5.0)	5.000(5.0,5.0)	474.000	-0.161	0.872
* $p < 0.05$ ** $p < 0.01$					

Open-Ended Feedback

Among instructors, 80.85% provided suggestions; thematic categories included classroom interaction and clinical reasoning guidance (57.89%), personalized teaching optimization (21.05%), and case resource development (13.16%). Among students, only 9.52% provided suggestions, focusing on nurse–patient communication practice and case variety.

Discussion

The present findings align with those of Yin et al. [5] in demonstrating that AI-assisted teaching rounds improve nursing students' clinical reasoning outcomes compared with conventional approaches. However, the current study extends this evidence in three substantive ways. First, while Yin et al. reported improvements in general "clinical competency" and "critical thinking" using instruments without reported psychometric validation, the present study employed the LCJR—a well-validated instrument

with established psychometric properties—to assess four distinct dimensions of clinical judgment (Observation, Interpretation, Response, and Reflection), each of which showed significant improvement with large effect sizes (Cohen's d = 1.05–1.31). Second, whereas Yin et al. focused exclusively on student outcomes, the present study incorporated both instructor and student measures, revealing an asymmetry in benefit perception: instructors perceived substantial efficiency gains, while students' subjective satisfaction remained unchanged. Third, this study is the first to systematically embed AI across all six stages of the BOPPPS model, demonstrating that a structured pedagogical framework is essential for translating AI-generated efficiencies into measurable learning outcomes. These contributions position AI not merely as a supplementary tool but as an integral component of instructional design in clinical nursing education.

Summary of Principal Findings

This study developed and evaluated an AI-assisted nursing teaching ward round process grounded in the BOPPPS model. The AI-assisted approach significantly reduced instructors' perceived transactional workload, enhanced their capacity for guided discussion, and produced substantial improvements in interns' clinical judgment, with large effect sizes across all four LCJR dimensions (Cohen's d = 1.05–1.31). Student satisfaction did not differ between groups, revealing a disconnect between instructor-perceived benefits and student-reported experiences. Subgroup analyses further indicated that instructors with over 10 years of experience reported lower satisfaction than their less experienced counterparts, suggesting that implementation strategies must be sensitive to instructors' career stages.

Workload Reduction and Pedagogical Capacity Enhancement

Instructors in the intervention group reported significantly reduced preparation time and enhanced capacity for in-depth discussion guidance. This finding aligns with prior systematic evidence that generative AI tools effectively automate case material organization, quiz generation, and preliminary performance analysis [7]. The present study extends this evidence to the clinical ward round setting—an environment characterized by high time pressure, variable case complexity, and competing clinical responsibilities. The intervention group scored approximately 0.6–0.8 points higher than the control group on relevant items (on a 5-point scale), representing a 15–20% relative improvement—a practically meaningful gain given the demanding clinical teaching context.

The absence of significant differences in "precise professional feedback" and "active participation" suggests that workflow efficiency alone does not guarantee enhanced pedagogical quality. AI tools are only as effective as the instructional strategies within which they are embedded. This finding resonates with prior research cautioning that technology-mediated efficiency gains must be deliberately channeled into pedagogical improvement rather than simply content coverage [8]. Faculty development programs should therefore equip instructors with strategies to translate AI-saved time into higher-order teaching behaviors, such as Socratic questioning, structured debate facilitation, and adaptive scaffolding.

A Mechanism for Effective Participatory Learning

The substantial improvement in interns' clinical judgment may be attributable to a mechanism termed the Cognitive Anchor Drive: personalized pre-assessment feedback focuses learners' attention on specific knowledge gaps, thereby enhancing the efficiency of subsequent participatory learning.

Both groups completed a pre-assessment quiz before group discussion. In the intervention group, instructors submitted aggregated results to the AI assistant, which generated dimension-specific accuracy rates and identified weak knowledge points for each intern. This created clearly defined learning gaps that each intern could actively address during participatory learning. In contrast, control group instructors manually calculated overall accuracy rates, offering only group-level estimates that lacked the individualization needed for learners to orient their subsequent engagement. This distinction likely explains why both groups, despite allocating 50% of session time to participatory learning, produced substantially different outcomes. The effectiveness of participatory learning thus appears to depend less on discussion time quantity than on the quality of learners' preparation and their awareness of specific learning needs prior to discussion—a finding consistent with cognitive load theory, which posits that pre-assessment feedback serves as an advance organizer that reduces extraneous cognitive load and enables more efficient use of germane load during discussion [9]. The AI's role extended beyond feedback generation to offering precise, individualized information—a level of personalization that is logistically challenging to accomplish manually in group teaching settings.

The Disconnect Between Instructor Gains and Student Experience

Despite significant instructor and clinical judgment benefits, student satisfaction did not differ between groups. This "instructor gains, student indifference" pattern warrants consideration of several complementary explanations.

First, ceiling effects constrained differentiation: satisfaction scores approached the scale maximum in both groups (median = 5.00/5.00 for overall satisfaction), leaving minimal room for the intervention to demonstrate incremental improvement. Second, the transmission of efficiency gains from instructors to students was incomplete—instructors experienced immediate benefits, but these savings did not automatically translate into more interactive or responsive teaching from the students' perspective. Without active instructor interpretation or transformation into dialogic teaching, AI-generated materials may function as "thinking substitutes" rather than "cognitive scaffolds." Third, cognitive offloading may have attenuated deep engagement: when AI-generated analyses were presented directly to students, some may have delegated their thinking to the AI rather than actively constructing their own reasoning, reducing the "desirable difficulty" required for durable learning [8]. Fourth, students may lack metacognitive awareness of their own learning gains—satisfaction measures capture affective responses, whereas the LCJR assesses objective competencies. Students' clinical judgment improved

substantially, yet they may not consciously perceive these cognitive gains, a pattern further supported by the low student open-ended response rate (9.52%).

These findings suggest that AI benefits are not automatically perceived by learners. Effective implementation therefore requires explicit attention to learner awareness—helping students recognize the cognitive gains they have achieved through the AI-assisted approach. This may involve debriefing sessions that highlight specific improvements in clinical reasoning or self-assessment exercises that make learning progress salient to students.

Differential Responses Across Instructor Experience Levels

Instructors with over 10 years of experience rated student preparedness and overall satisfaction significantly lower than less experienced counterparts, although no differences were observed in ratings of discussion quality. This pattern may reflect distinct cognitive and professional orientations toward AI integration. Early-career instructors may be more receptive to AI as a "cognitive aid" that compensates for limited experience in case formulation and question design. Senior educators, by contrast, may hold higher standards, exhibit greater skepticism toward algorithmic recommendations, or perceive AI as undermining professional autonomy and diluting relational aspects of clinical teaching [7].

These findings suggest that AI teaching tools should not be promoted uniformly. For early-career instructors, AI may serve as an accelerant for teaching competency development. For senior instructors, implementation strategies should respect experiential expertise, positioning AI as a supplement rather than a replacement for professional judgment. Customized training that addresses instructors' specific concerns across career stages may enhance acceptance and satisfaction and ultimately improve implementation fidelity.

Theoretical Contributions

This study proposes two concepts with direct implications for both theory and practice in AI-mediated clinical teaching.

First, Technology Beneficiary Asymmetry describes the conditional transfer of AI benefits from direct users (instructors) to indirect beneficiaries (students), contingent on instructors' pedagogical transformation capacity. This concept extends the Technology Acceptance Model (Davis, 1989) by specifying that user acceptance of AI tools is necessary but insufficient for educational benefit; actual learning gains depend on the extent to which instructors transform AI-generated efficiencies into higher-order teaching practices. In practical terms, institutions must invest in faculty development that helps instructors convert AI-saved time into dialogic teaching strategies rather than simply covering more content.

Second, the Cognitive Anchor Drive articulates how personalized pre-assessment feedback focuses learners' attention on specific gaps, thereby enhancing participatory learning efficiency. This concept

extends cognitive load theory (Sweller, 1988) by specifying a mechanism through which AI-generated analytics can reduce extraneous cognitive load and optimize germane load during group discussions. Its practical implication is straightforward: pre-assessment should be treated as a pedagogical intervention that shapes learners' cognitive orientation, not as a routine administrative task. Free, web-based AI tools such as the one used in this study are sufficient to generate these analytics in most clinical settings, making this approach accessible even in resource-constrained environments.

Practical Implications for Clinical Implementation

The AI-assisted model evaluated in this study offers a feasible, low-cost approach for clinical education, particularly in resource-constrained settings. Unlike proprietary AI systems that require institutional licensing or infrastructure upgrades, the AI assistant used in this study is accessible via a free, web-based interface without additional hardware requirements, making the model potentially scalable across hospitals with varying resource levels.

For clinical educators seeking to implement this model, the following actionable recommendations are offered. First, start with pre-assessment analytics. This requires minimal AI proficiency and provides immediate, visible benefits for both instructors and learners. Instructors can begin by using AI to identify learners' specific weakness points before group discussions—a practice that can be gradually integrated into existing ward-round routines. Second, develop stage-specific faculty training, not generic digital literacy courses. Training should focus on: (a) formulating effective prompts for AI at each BOPPPS stage; (b) critically evaluating and validating AI-generated content for clinical accuracy; and (c) transforming AI-generated analytics into dialogic teaching strategies. Training should be differentiated by instructors' career stage, with early-career instructors receiving foundational support and senior instructors receiving implementation support that respects their experiential expertise. Third, monitor not only student outcomes but also teacher acceptance and fidelity. The subgroup analysis revealed that senior instructors reported lower satisfaction, suggesting that one-size-fits-all implementation may be counterproductive. Implementation leaders should identify early adopters among less experienced instructors to build local evidence and success stories before scaling to the entire teaching team, while addressing senior faculty concerns about professional autonomy and the relational aspects of teaching.

Integration with Existing Theoretical Frameworks

The two concepts proposed in this study—Technology Beneficiary Asymmetry and Cognitive Anchor Drive—extend established theoretical frameworks to the distinct context of AI-mediated clinical education. The Technology Acceptance Model (TAM) has been widely used to predict user acceptance of educational technologies; however, the present findings suggest that acceptance is only the first step. The translation of AI efficiency gains into measurable student outcomes is not automatic but depends on instructors' pedagogical capacity. This insight refines TAM by introducing a conditional transmission mechanism between instructor acceptance and student benefit.

Similarly, cognitive load theory has been extensively applied to instructional design, but its application to AI-mediated clinical teaching has been limited. The Cognitive Anchor Drive concept specifies how AI-generated pre-assessment analytics can serve as advance organizers that reduce extraneous cognitive load and enhance germane load during participatory learning. This extends cognitive load theory by articulating a concrete mechanism—personalized gap identification—through which AI can optimize cognitive resource allocation during group learning activities.

Limitations and Future Directions

This study has several limitations. First, the single-center design and convenience sampling limit generalizability to institutions with different teaching cultures or resources. Second, the LCJR assessments were completed by the same instructors who delivered the intervention, introducing potential assessor bias that may have inflated the observed effect sizes. Third, the sample was determined by availability rather than prospective power analysis; although post-hoc power exceeded 0.90 for the primary outcomes, the study may be underpowered to detect small or moderate effects. Fourth, the absence of long-term follow-up leaves the durability of clinical judgment gains unclear.

Despite these limitations, the large effect sizes across all LCJR dimensions support the intervention's educational potential. Future multicenter randomized controlled trials with blinded assessments, prospective power calculations, and extended follow-up are needed to strengthen causal inference and evaluate long-term retention.

Conclusion

This study demonstrates that embedding a free, web-based AI assistant across all six stages of the BOPPPS model significantly reduces instructors' perceived preparation workload and substantially improves interns' clinical judgment, with large effect sizes across all LCJR dimensions. However, these benefits do not automatically extend to students' subjective satisfaction, underscoring that AI-mediated efficiency gains require deliberate pedagogical translation to become educationally meaningful. The AI-assisted model offers a feasible, low-cost approach for clinical teaching in resource-constrained settings; yet, successful implementation demands faculty development that equips instructors to transform AI-generated materials into interactive, dialogic teaching. Future research should employ multicenter designs with blinded assessments and long-term follow-up to evaluate the durability and generalizability of these findings.

Abbreviations

AI
Artificial Intelligence
BOPPPS
Bridge-in, Objectives, Pre-assessment, Participatory Learning, Post-assessment, Summary

LCJR
Lasater Clinical Judgment Rubric
STROBE
Strengthening the Reporting of Observational Studies in Epidemiology.

Declarations

Ethics approval and consent to participate

The study was approved by the Clinical Research and Application Ethics Committee of the Second Affiliated Hospital of Guangzhou Medical University. Written informed consent was obtained from all participants. All methods were performed in accordance with the relevant guidelines and regulations of the Declaration of Helsinki.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated or analyzed during the current study are not publicly available due to institutional data protection policies but are available from the corresponding author on reasonable request.

Competing Interests

The authors declare that they have no competing interests.

Funding

This study was supported by provincial and municipal nursing research grants (Grant Nos. GDHLYJYM202418 and L2025030). The funding sources had no role in the study design, data collection, analysis, interpretation, or manuscript preparation.

Authors' contributions

Xuemei Liu contributed to conceptualization, methodology, investigation, formal analysis, writing – original draft, project administration, and funding acquisition. Haiping Chen, Xuelian Wu, and Hongbo Yan contributed to investigation and data curation. Jian Song and Chang Zhang contributed to validation and supervision. Bing Wang and Runfang Zhong contributed to project administration. Hongjing Yu contributed to methodology, supervision, and writing – review and editing. All authors reviewed and approved the final manuscript.

Acknowledgements

The authors thank all participating instructors and interns for their time and contributions to this study.

References

1. Haddon Mullins C, Roderick A, Deaver J, Willig J. Contemporary practice of standardised bedside teaching rounds. Clin Teach. 2020;17(5):483–8.
2. Thomas C, Plumblee L, Dieffenbaugher S, Talley C. Teaching on rounds and in small groups. Surg Clin North Am. 2021;101(4):555–63.
3. Xiao X, Xie Y, Zhao X, Dai X, Liang F, Wang W, et al. The application effectiveness of the One Case to the End teaching model based on BOPPPS in pediatric nursing: a quasi-experimental study. BMC Med Educ. 2025;25(1):556.
4. Zhu J, Xiao H, Zhou R, Gan X, Gou Q, Tie H. The efficacy of the BOPPPS teaching model in clinical and health education: a systematic review and meta-analysis. BMC Med Educ. 2025;25(1):997.
5. Yin J, Hao X, Xing G, Xu M. The effects of ChatGPT-driven blended teaching model on nursing rounds: a quasi-experimental study. Nurse Educ Pract. 2025;88:104545.
6. Lasater K, Nielsen A. The Lasater Clinical Judgment Rubric: 17 years later. J Nurs Educ. 2024;63(3):149–55.
7. Alrazeeni DM, Alharrasi M, Khan Rony MK, Biswas RK, Tama IJ, Halder CR, et al. Transforming nursing education with artificial intelligence: a systematic review (2010–2025). SAGE Open Nurs. 2026;12:23779608261424597.
8. Chan MMK, Wan AWH, Cheung DSK, et al. Integration of artificial intelligence in nursing simulation education: a scoping review. Nurse Educ. 2025;50(4):195–200.
9. Kaylor SK. Preventing information overload: cognitive load theory as an instructional framework for teaching pharmacology. J Nurs Educ. 2014;53(2):108–11.

Supplementary Files

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

- SupplementaryFile1.doc