



# Improving Clinical Reasoning in Interns through a Deliberate Practice and Kolb-Based Closed-Loop Training System: A Non-Randomized Controlled Study

Zhen Zhang<sup>1#</sup>, Lili Yang<sup>1#</sup>, Chujie Chen<sup>2</sup>, Dan Liu<sup>1</sup>, and Peng Yun<sup>1\*</sup>

Departments of Endocrinology<sup>1</sup> and Urology<sup>2</sup>, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen 518000, China.

# Zhen Zhang and Lili Yang contributed equally

**Corresponding author:** Peng Yun, MD/PhD, Department of Endocrinology, The Seventh Affiliated Hospital, Sun Yat-sen University, 628 Zhenyuan Road, Guangming District, Shenzhen, China. Tel: +86-0755-81206795; Email address: [yunpeng@sysush.com](mailto:yunpeng@sysush.com)

## ABSTRACT

**Purpose:** To evaluate a virtual patient-based closed-loop training system combining deliberate practice with Kolb's experiential learning cycle and assess its impact on clinical reasoning in medical interns.

**Methods:** In this quasi-experimental study, 235 fifth-year interns (control: n = 120; intervention: n = 115) received the same conventional clinical teaching. Additionally, the intervention group completed a four-phase closed-loop training program, orientation, hands-on training, structured debriefing, and targeted reinforcement, using the ZhiQu virtual patient platform, grounded in deliberate practice and Kolb's experiential learning cycle. The control group used the same platform for unsupervised self-practice. Outcomes included longitudinal clinical reasoning scores, Objective Structured Clinical Examination (OSCE) performance, graduation examination results, and teaching satisfaction.

**Results:** Compared with the control group, the intervention group showed greater improvement in systematicity, accuracy, and logic, with higher final scores in these dimensions. No significant differences were observed for agility or cost-effectiveness. The intervention group also performed better on OSCE stations for history taking, medical record writing, and case analysis, as well as on the graduation clinical skills examination. Theory examination scores did not differ significantly. Teaching satisfaction was significantly higher in the intervention group.

**Conclusions:** A closed-loop training system driven by deliberate practice and Kolb's experiential learning cycle effectively improves specific dimensions of clinical reasoning and cognitive-intensive clinical skills in medical interns. Virtual patient platforms should be embedded within structured, faculty-led debriefing rather than used as standalone electronic question banks.

## ARTICLE HISTORY

Received: May 9, 2026

Revised: May 26, 2026

Accepted: May 27, 2026

## KEYWORDS

clinical reasoning,  
deliberate practice,  
Kolb cycle, virtual patient

## Introduction

In current clinical internship training, traditional teaching methods, such as lectures, medical record writing, bedside teaching, and case discussions, remain the mainstay. While widely used, these approaches often lean toward one-way knowledge delivery and static assessment, with relatively little attention paid to fostering systematic and dynamic clinical reasoning. As a result, many interns struggle to connect theory with practice, especially in real-world settings where decision-making gaps frequently emerge [1]. Existing teaching models rarely provide structured feedback on diagnostic logic, treatment rationale, or cost-effectiveness, leaving the

development of clinical reasoning skills somewhat fragmented and superficial [2]. There is a clear need, therefore, to explore new instructional strategies that more effectively integrate theory, practice, and meaningful reflection.

Some studies suggest that high-fidelity virtual patient systems, such as the “ZhiQu Clinical Reasoning System”, may help fill this gap [3]. This platform evaluates clinical reasoning across five domains: accuracy, systematicity, logic, cost-effectiveness, and agility, while also capturing detailed process data. However, technology alone does not guarantee better learning outcomes. Simply giving students access to virtual patients without structured guidance may lead to

**Table 1.** Comparison of Demographic Data and Baseline Scores Between the Two Groups

	Control Group (n=120)	Intervention Group (n=115)	Statistical Value	p Value
Age	20 (20, 21)	20 (20, 21)	6781.50	0.79
Gender (M/F)	56/64	65/50	2.28	0.13
Theoretical Scores	78.62±7.31	78.72±6.71	-0.12	0.91
History Taking	5 (5, 6)	5 (4, 6)	6138.00	0.13
Physical Examination	6 (5, 6)	6 (5, 6)	6389.50	0.28
Procedural Skills	5 (4, 6)	5 (4, 6)	6788.50	0.82
Doctor-Patient Communication	5 (4, 6)	5 (4, 6)	6571.50	0.52
Medical Record Writing	5 (4, 6)	5 (4, 6)	6777.50	0.81
Case Analysis	5 (3, 6)	5 (4, 6)	6537.50	0.47

what some researchers call “naive practice,” where learners repeat their existing mistakes without real improvement [4].

Two complementary theoretical frameworks guided our intervention design. The first is deliberate practice theory, which argues that expert performance is built through focused practice with clear goals and immediate, specific feedback [5]. The second is Kolb’s experiential learning theory, which views learning as a four-stage cycle: concrete experience, reflective observation, abstract conceptualization, and active experimentation [6]. According to this view, learning requires completing the full cycle; experience alone, without structured reflection, is not enough.

Although there is growing interest in combining these two theories in medical education, research on their systematic application to virtual patient-based clinical reasoning curricula is still limited. To address this gap, we developed and implemented a four-phase closed-loop training system: theory orientation (abstract conceptualization), hands-on training (concrete experience), structured debriefing (reflective observation), and targeted reinforcement (active experimentation).

## Objects and Methods

### Participants and Grouping

This study used a quasi-experimental design with non-randomized historical controls, conducted at a university-affiliated teaching hospital between May 2023 and May 2025. Participants were fifth-year clinical medicine interns from a medical school in China. The control group comprised 120 interns from the Class of 2019 (internship period: May 2023 - April 2024). The intervention group comprised 115 interns

from the Class of 2020 (internship period: May 2024 - May 2025). Baseline demographic and academic characteristics were comparable between groups (Table 1). The study protocol was approved by the hospital’s ethics committee, and all participants gave written informed consent.

### Teaching Methods

Both groups received the same conventional clinical teaching, including lectures, bedside teaching, case discussions, and medical record writing and review.

### Control group

In addition to conventional teaching, students completed a standardized orientation for the ZhiQu virtual patient system at the start of their rotation. They were given system access and required to complete at least three virtual cases per subspecialty. No structured debriefing was provided; students mainly relied on the system’s automated scoring for self-assessment.

### Intervention group

Beyond conventional teaching, the intervention group took part in a theory-driven closed-loop training system using the ZhiQu platform, structured around Kolb’s experiential learning cycle.

### Phase 1: Theory orientation (abstract conceptualization)

In the first week of the rotation, students attended a two-hour introductory course on clinical reasoning, covering dual-process theory, illness script theory, common cognitive biases, and the five-dimensional scoring framework of the ZhiQu system.

*Phase 2: Hands-on training (concrete experience)*

For each sub-specialty, at least three virtual patient sessions (45-60 minutes each) were arranged. The “timeline” and “complication” functions were enabled to simulate time pressure and disease progression.

*Phase 3: Structured debriefing (reflective observation)*

After each session, the faculty organized small-group debriefings (2-4 students per group, 30-40 minutes). These sessions involved data-driven identification of common problems, individual learning path tracking, Socratic questioning to encourage metacognitive reflection, and immediate individualized feedback [7].

*Phase 4: Targeted reinforcement (active experimentation)*

Following debriefing, faculty selected variant cases (e.g., similar etiology but different presentations) from the ZhiQu case bank based on each student's weaknesses and assigned them for independent re-training. Students then applied the reasoning strategies developed during debriefing to new scenarios. This cycle of “practice → feedback → error correction → re-practice” could be repeated 2-3 times until reaching a mastery standard (comprehensive score ≥ 80), reflecting the iterative nature of deliberate practice and the essence of Kolb's active experimentation.

*Outcome Measures**Clinical reasoning process indicators*

Using data extracted from the ZhiQu platform, we assessed clinical reasoning across five dimensions: accuracy (match between diagnosis and key evidence), systematicity (completeness of history taking and differential diagnosis coverage), logic (causal coherence of reasoning), cost-effectiveness (rationality of test and treatment choices), and agility (accuracy of key decisions per unit time). Data were collected at three time points: baseline (pre-intervention), mid-intervention, and post-intervention (before rotation completion). Individual “clinical reasoning growth curves” were generated for each intern.

*Objective Structured Clinical Examination (OSCE)*

At the end of the rotation, all interns completed an OSCE consisting of six stations: history taking, physical examination, procedural skills, doctor-patient communication, medical record writing, and

case analysis. Each station was scored out of 10, for a total of 60 points. Two independent, blinded faculty members (at least intermediate professional titles) assessed each station, and the average score was used as the final station score. All assessors received standardized training before the examination to ensure scoring consistency.

*Far transfer to academic performance*

To assess longer-term transfer effects, we tracked students' scores on the university-level graduation theory examination and graduation clinical skills examination after the internship period.

*Affective evaluation*

After the intervention, we administered an anonymous, self-developed teaching satisfaction questionnaire (scored 0-10, with higher scores indicating greater satisfaction) to assess students' attitudes toward the new teaching model, thereby reducing social desirability bias.

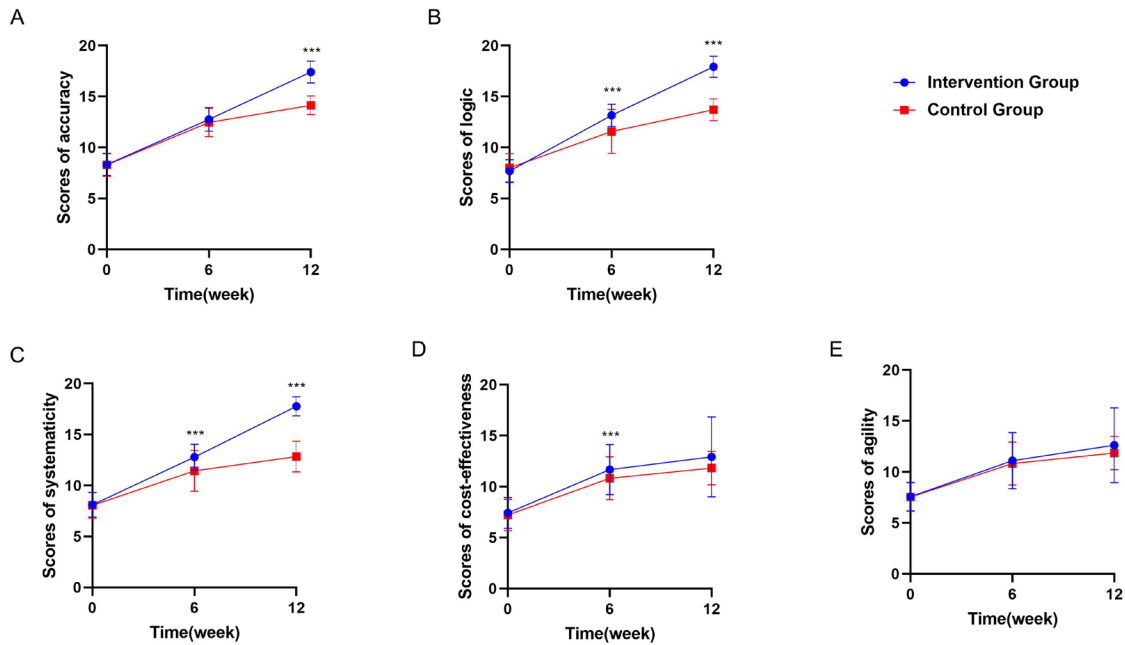
*Statistical Analysis*

All statistical analyses were performed using SPSS version 26.0. Continuous variables were tested for normality and homogeneity of variance. Data meeting these assumptions were compared using independent-samples *t*-tests; otherwise, the Mann-Whitney U test was applied. Categorical variables were compared using the chi-square test. A two-tailed  $p < 0.05$  was considered statistically significant.

**Results***Clinical reasoning process indicators*

As shown in Figure 1, the two groups had similar baseline scores across all five dimensions. During the internship, the control group showed slow improvement in systematicity, accuracy, and logic, eventually reaching a plateau. In contrast, the intervention group demonstrated continuous improvement in these three dimensions and achieved significantly higher final scores. For agility and cost-effectiveness, both groups improved over time. Regarding cost-effectiveness, the intervention group showed a significantly higher score than the control group in week 6 ( $p < 0.001$ ), but the difference was no longer statistically significant at week 12. For agility, no statistically significant differences were observed between the two groups at any time point.

**Figure 1:** Comparison of clinical reasoning growth curves over three time points between the intervention and control groups. (A) Accuracy; (B) Systematicity; (C) Logic; (D) Cost-effectiveness; (E) Agility.



#### OSCE assessment

The intervention group performed significantly better than the control group in history taking, medical record writing, and case analysis, while no significant differences were observed in physical examination, procedural skills, or doctor–patient communication (Table 2).

#### Far transfer to academic performance

The intervention group achieved significantly higher scores in the graduation clinical skills examination, whereas no significant difference was found between the two groups in the theory examination (Table 2).

#### Affective evaluation

Teaching satisfaction was significantly higher in the intervention group than in the control group (Table 2). Students in the intervention group particularly highlighted the value of structured debriefing, noting that it helped them identify specific errors in their clinical reasoning.

## Discussion

This study addressed a key challenge in clinical reasoning education by developing a theory-driven closed-loop training system that integrates deliber-

ate practice with Kolb’s experiential learning theory. Compared to unsupervised virtual patient practice, our four-phase model (orientation → training → debriefing → reinforcement) significantly improved interns’ performance on specific reasoning dimensions and OSCE stations. The intervention group also performed better on graduation clinical skills examinations and reported higher teaching satisfaction.

Among the five reasoning dimensions, the intervention group showed marked improvement in systematicity, accuracy, and logic. In contrast, unsupervised self-exploration with automated feedback was insufficient to achieve similar gains. These findings are consistent with deliberate practice theory, which emphasizes clear goals, immediate feedback, and iterative refinement, elements largely absent in the control condition [8]. They are also consistent with Kolb’s experiential learning cycle, which requires structured reflection to complete the learning process [9]. In our intervention, structured debriefing served as that critical “reflective observation” phase, a component the control group lacked.

This finding also has practical implications for the use of AI-assisted tools in medical education. Technology platforms can generate rich behavioural data, but their educational value depends on how they are used. Simply treating AI tools as an electronic

**Table 2.** Comparison of post-intervention outcomes between the intervention and control groups

		Control Group (n = 120)	Intervention Group (n = 115)	Statistical Value	p Value
<b>OSCE stations</b>	History Taking	6 (5, 6)	6 (5, 7)	4349.50	< 0.001
	Physical Examination	6 (5.25, 6)	6 (5, 6)	6538.50	0.43
	Procedural Skills	6 (6, 7)	6 (5, 7)	6218.50	0.17
	Doctor-Patient Communication	5 (4.25, 6)	5 (5, 6)	6850.50	0.92
	Medical Record Writing	5 (4, 6)	6 (5, 6)	4528.50	< 0.001
	Case Analysis	6 (5, 6)	7 (6, 8)	3853.50	< 0.001
<b>Graduation examination</b>	Theory Examination	86 (81, 89)	86 (80, 89)	6187.50	0.17
	Clinical Skills	83 (77, 86)	87 (82, 91)	3960.00	< 0.001
<b>Teaching satisfaction</b>		7 (6,8)	7 (6, 9)	5534.50	0.007

question bank for independent practice is unlikely to develop core reasoning or metacognitive reflection [10]. Backend data requires faculty interpretation, and variant case recommendations need to be integrated with debriefing discussions to effectively reshape reasoning pathways. The observed improvements in systematicity, accuracy, and logic can be interpreted through two theoretical mechanisms. First, faculty-led debriefing corresponds to reflective observation in Kolb's cycle. Second, deliberate practice with variant cases operationalizes the iterative loop of practice, feedback, error correction, and re-practice, a process central to deliberate practice theory.

The lack of improvement in agility and cost-effectiveness is a null finding. For agility, the "timeline" function may not have fully simulated real-time pressure, and the intervention duration may have been too short. For cost-effectiveness, debriefing may have placed insufficient emphasis on cost considerations, and interns lacked formal health economics training.

The intervention group performed significantly better in history taking, medical record writing, and case analysis, all cognitively intensive tasks requiring higher-order reasoning. No significant differences were found in physical examination, procedural skills, or doctor-patient communication. This pattern indicates that the intervention specifically enhanced clinical reasoning rather than motor or communication skills, ruling out a global Hawthorne effect. These findings can be interpreted through two theoretical mechanisms. Time-pressured virtual patient simulations may activate deeper cognitive processing, and structured debriefing with backend data may promote metacognitive reflection. This synergy aligns with Kolb's assertion that concrete experience alone is insufficient; it must be paired with reflective

observation to generate abstract conceptualization [11]. The deliberate practice framework further explains the iterative nature of our intervention [12]. Each simulation and debriefing cycle provided focused practice with specific performance targets and immediate feedback, hallmarks of deliberate practice. Together, these mechanisms may support the transfer of knowledge to new contexts.

Conversely, the lack of transfer to physical examination and procedural skills reflects the inability of virtual patient platforms to replicate tactile feedback. Similarly, the absence of improvement in communication suggests that current VP systems cannot capture empathy or nonverbal cues. These findings support the task-training alignment principle: VP-based simulation is effective for reasoning but cannot replace hands-on practice or standardized patient interactions for psychomotor and communication skills.

The graduation clinical skills examination, administered several months after the intervention ended, assessed far transfer. The intervention group's superior performance suggests that structured reasoning training generated durable cognitive schemas. From a Kolbian perspective, this indicates that students completed the full experiential learning cycle, from concrete experience (simulation) through reflective observation (debriefing) and abstract conceptualization (rule formation) to active experimentation (variant cases), thereby constructing transferable mental models. Deliberate practice theory attributes this durability to repeated, feedback-driven iterations that consolidated reasoning patterns into long-term cognitive structures [13].

In addition, the intervention group reported significantly higher teaching satisfaction. Students consistently attributed this to structured debriefing,

which helped them identify specific errors in their clinical reasoning. This finding suggests that the educational value of the closed-loop system lies not merely in repeated exposure to virtual cases but in the opportunity for guided reflection that makes reasoning gaps visible and correctable effects of hot and cold debriefing in simulation with case-based learning [14]. By transforming abstract feedback into actionable insight, structured debriefing serves as the key mechanism distinguishing deliberate practice from naive repetition, a distinction central to both theoretical frameworks.

### Limitations

This study has several limitations. First, the non-randomized design may have introduced confounding bias, although baseline characteristics were comparable. Future multicenter randomized controlled trials are needed to establish causal effects. Second, the single-site sample limits generalizability. Third, the null results for agility and cost-effectiveness suggest that the current intervention intensity was insufficient to improve all reasoning dimensions. Future studies could incorporate greater time pressure training and health economics cases. Finally, follow-up ended at graduation; longer-term outcomes during residency and career development remain to be examined.

### Conclusion

This study provides empirical evidence that a closed-loop training system driven by deliberate practice and Kolb's experiential learning theory effectively improves interns' clinical reasoning. The findings demonstrate that the educational value of virtual patient systems depends on the structured pedagogical framework in which they are embedded; such technologies are most effective when integrated with faculty-led debriefing rather than used alone. Accordingly, we recommend that medical schools implementing virtual patient systems simultaneously develop structured debriefing mechanisms. This approach may help lay the foundation for evidence-based, precision-oriented medical education.

### Authors' Contributions:

Zhen Zhang and Lili Yang (co-first authors) designed the study and drafted the manuscript. Chujie Chen collected and interpreted the clinical data and assisted in manuscript preparation. Dan Liu analyzed

the data. Peng Yun (corresponding author) conceived the study, supervised the project, and revised the manuscript.

### Conflicts of interest:

The authors declare no competing interests.

### Ethical Statement:

This study was examined and approved by the Medical Research Ethics Committee of The Seventh Affiliated Hospital, Sun Yat-sen University (KY-2026-066-01).

### Funding Support:

Teaching Quality Engineering Project of Sun Yat-sen University (#78000-12253001)

### References:

1. Fang X, Jin S, Sun Z: **Application of practical training scenario simulation teaching method oriented to strengthen clinical skills in the clinical internship stage of medical students.** *Med Educ Online* 2025, **30**(1):2500557; doi:10.1080/10872981.2025.2500557; PMC12054555.
2. Luo P, Shen J, Yu T, Zhang X, Zheng B, Yang J: **Formative objective structured clinical examination with immediate feedback improves surgical clerks' self-confidence and clinical competence.** *Med Teach* 2023, **45**(2):212-218; doi:10.1080/0142159X.2022.2126755.
3. Dale MacLaine T, Juengst C, Harris D, Fenn C, Gabathuler H, Davies S: **The (future) doctor will see you now: Piloting a longitudinal virtual patient in medical education, simulating general practice.** *Med Teach* 2021, **43**(4):472-474; doi:10.1080/0142159X.2020.1853689.
4. Plackett R, Kassianos AP, Timmis J, Sheringham J, Schartau P, Kambouri M: **Using Virtual Patients to Explore the Clinical Reasoning Skills of Medical Students: Mixed Methods Study.** *J Med Internet Res* 2021, **23**(6):e24723; doi:10.2196/24723; PMC8214179.
5. Perretta JS, Duval-Arnould J, Poling S, Sullivan N, Jeffers JM, Farrow L, Shilkofski NA, Brown KM, Hunt EA: **Best Practices and Theoretical Foundations for Simulation Instruction Using Rapid-Cycle Deliberate Practice.** *Simul Healthc* 2020, **15**(5):356-362; doi:10.1097/SIH.0000000000000433.

6. Butler K, Klein S, Mackenzie D, Zayed N, Ndiaye MC, Marsh M: **Kolb's Experiential Learning in Action: A Curriculum for Residents.** *Clin Teach* 2026, **23**(2):e70344; doi:10.1111/tct.70344.
7. Senette CL, Dingley C, Doolen J, Gordon H: **Fluctuating cohesion: A grounded theory study of nursing students engaged in a combined debriefing format.** *Nurse Educ Pract* 2024, **76**:103943; doi:10.1016/j.nepr.2024.103943.
8. Draaisma E, Maggio LA, Bekhof J, Jaarsma ADC, Brand PLP: **Impact of deliberate practice on evidence-based medicine attitudes and behaviours of health care professionals.** *Perspect Med Educ* 2021, **10**(2):118-124; doi:10.1007/s40037-020-00634-9; PMC7952477.
9. Kaushik R: **A Comprehensive Outpatient Pediatric Resident Complex Care Curriculum.** *MedEdPORTAL* 2023, **19**:11319; doi:10.15766/mep\_2374-8265.11319; PMC10326167.
10. Lee J, Wu AS, Li D, Kulasegaram KM: **Artificial Intelligence in Undergraduate Medical Education: A Scoping Review.** *Acad Med* 2021, **96**(11S):S62-S70; doi:10.1097/ACM.0000000000004291.
11. Astbury J, Ferguson J, Silverthorne J, Willis S, Schafheutle E: **High-fidelity simulation-based education in pre-registration healthcare programmes: a systematic review of reviews to inform collaborative and interprofessional best practice.** *J Interprof Care* 2021, **35**(4):622-632; doi:10.1080/13561820.2020.1762551.
12. Lundquist LL, Bilich LA, Jackson SC, Stevens KV, Tipton EJ: **Measurable reflection in simulation: A pilot study.** *J Dent Educ* 2021, **85**(5):606-614; doi:10.1002/jdd.12506.
13. Enoch LC, Abraham RM, Singaram VS: **A comparative analysis of the impact of online, blended, and face-to-face learning on medical students' clinical competency in the affective, cognitive, and psychomotor domains.** *BMC Med Educ* 2022, **22**(1):753; doi:10.1186/s12909-022-03777-x; PMC9628081.
14. Ha EH: **Effects of hot and cold debriefing in simulation with case-based learning.** *Jpn J Nurs Sci* 2021:e12410; doi:10.1111/jjns.12410.